

Robust grasping under object pose uncertainty

Kaijen Hsiao, Leslie Pack Kaelbling & Tomás Lozano-Pérez

Autonomous Robots

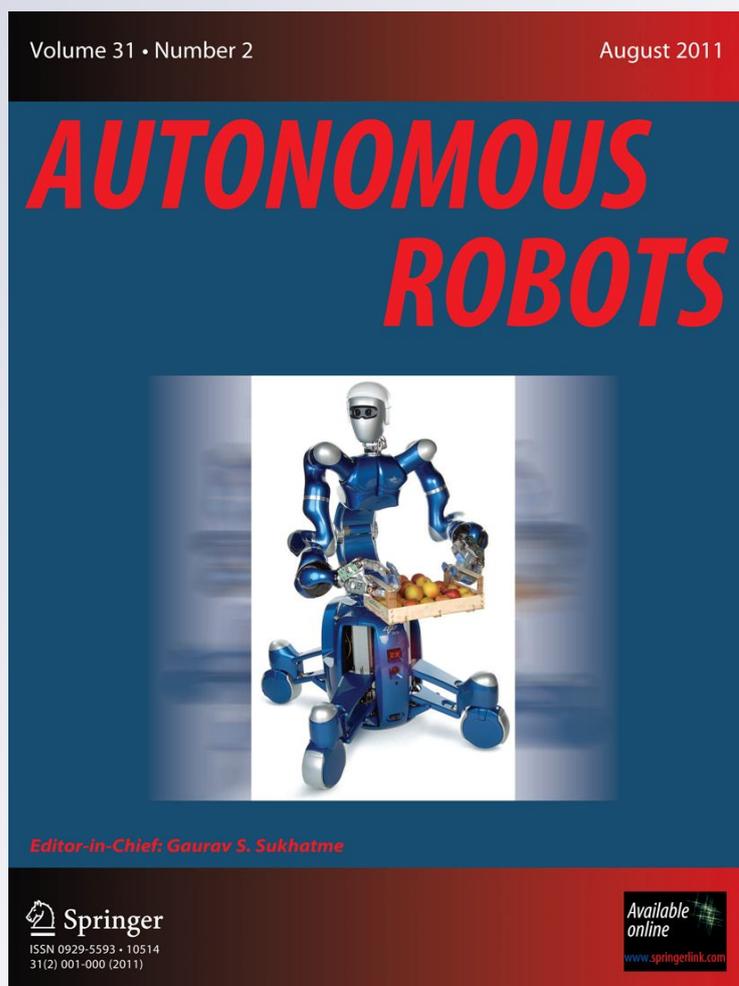
ISSN 0929-5593

Volume 31

Combined 2-3

Auton Robot (2011) 31:253-268

DOI 10.1007/s10514-011-9243-2



Your article is protected by copyright and all rights are held exclusively by Springer Science+Business Media, LLC. This e-offprint is for personal use only and shall not be self-archived in electronic repositories. If you wish to self-archive your work, please use the accepted author's version for posting to your own website or your institution's repository. You may further deposit the accepted author's version on a funder's repository at a funder's request, provided it is not made publicly available until 12 months after publication.

Robust grasping under object pose uncertainty

Kaijen Hsiao · Leslie Pack Kaelbling ·
Tomás Lozano-Pérez

Received: 8 November 2010 / Accepted: 28 June 2011 / Published online: 29 July 2011
© Springer Science+Business Media, LLC 2011

Abstract This paper presents a decision-theoretic approach to problems that require accurate placement of a robot relative to an object of known shape, such as grasping for assembly or tool use. The decision process is applied to a robot hand with tactile sensors, to localize the object on a table and ultimately achieve a target placement by selecting among a parameterized set of grasping and information-gathering trajectories. The process is demonstrated in simulation and on a real robot. This work has been previously presented in Hsiao et al. (Workshop on Algorithmic Foundations of Robotics (WAFR), 2008; Robotics Science and Systems (RSS), 2010) and Hsiao (Relatively robust grasping, Ph.D. thesis, Massachusetts Institute of Technology, 2009).

Keywords Grasping · Planning under uncertainty · POMDPs · Manipulation · Robustness

1 Introduction

Our goal is to develop a general-purpose strategy for task-driven manipulation of objects when there is uncertainty about the relative pose of the robot and the objects. This strategy applies to *relative placement* problems, which require the robot to achieve accurate placement with respect

to a target object whose position is not accurately known. Placement problems include grasping (placing the robot relative to an object to be grasped), insertions (placing an object the robot is holding relative to another object), and other fine-motion tasks. In this paper, we focus on grasping at a specified pose on a known object.

Vision and range sensors can estimate the pose of an object, but there is still residual uncertainty, especially when important features of the object are partially occluded. Tactile sensing, combined with proprioception, can give highly reliable information about object position. However, it is expensive to map out an entire object with tactile sensing, so we want to use the information requirements of the task to drive the sensing.

Decision theory frames problems of action selection when the true world state is unknown, providing a principled way to trade off the cost of performing information-gathering actions against the costs of performing inappropriate actions in the world. A decision-theoretic controller is constructed from two components: state estimation and action selection. The *state estimator* maintains a *belief state*, which is a probability distribution over the underlying, but not directly observable, states of the world. Each time an observation (such as a contact sensor reading) is made, the belief state is updated to incorporate the new information; each time an action is taken, the belief state is updated to reflect possible changes in the world state due to the action. The *action selection* component considers the current belief state and decides whether the state has been estimated sufficiently accurately to execute a final goal-achieving action and terminate, or whether additional observations should be made. If additional observations are to be made it chooses an action based on its potential for increasing the likelihood of success.

K. Hsiao (✉)
Willow Garage, 68 Willow Rd., Menlo Park, CA 94043, USA
e-mail: hsiao@willowgarage.com

L.P. Kaelbling · T. Lozano-Pérez
MIT CSAIL, 32 Vassar St., Cambridge, MA 02139, USA

L.P. Kaelbling
e-mail: lpk@csail.mit.edu

T. Lozano-Pérez
e-mail: tlp@csail.mit.edu

In the general case, optimal action selection is computationally intractable. In some cases, sensing actions can be chosen effectively by “greedily” choosing the action that will most reduce uncertainty in one step. Our problems do not necessarily exhibit that trait, but we want to structure the problem so that only a small amount of lookahead is necessary to choose appropriate actions.

It is typical, in lower-dimensional control problems such as mobile-robot navigation, to use a uniform discretization of the primitive action space. Such a fine-grained discretization of the space presents two problems: first, there is a large branching factor in the choice of actions; second, the horizon (number of “steps” that must be made before the goal is reached) is quite long, requiring significant lookahead in planning to select an appropriate action.

Our strategy will be to generate, off-line, a relatively small set of *world-relative trajectories* (WRTs). A WRT is a parameterized trajectory: it consists of a sequence of Cartesian poses for the robot’s end-effector, expressed relative to an estimated pose of the object to be grasped. Also off-line, we characterize each WRT’s effectiveness in terms of achieving the goal and gaining information. Then, during the on-line execution phase, we will use this information, together with the continually updated belief state, to select and execute appropriate trajectories. One way to think of these trajectories is as temporally extended “macro actions.” This approach has a relatively small branching factor, and results in effective goal-directed action even with only one step lookahead.

We consider several approximate decision procedures, based on WRTs. In the simplest case, we have a single WRT, which would succeed as a terminal action if it were parameterized with the correct object pose. On every step, we execute that WRT, parameterized by the object pose that is most likely in the current belief state. The procedure is terminated when the estimated likelihood of success is high enough, and the WRT is executed one last time (if the goal condition is not already satisfied). A single WRT is not always enough to guarantee that the uncertainty will be reduced sufficiently, so we augment the set of WRTs with trajectories designed expressly with the goal of gathering information and/or re-orienting the object so it will be easier to interact with. Finally, we consider an extension to lookahead search, allowing the selection of an initial WRT because of its ability to gain information that will enable a subsequent WRT to be more effective, even though the initial WRT does not substantially reduce uncertainty in the dimensions that are relevant to actual achievement of the goal. We show that these procedures are expected to terminate with a correct answer, under assumptions about the informativeness of observations and the degree to which actions affect the state of the world.

The use of WRT’s in these decision procedures allows us to limit the set of actions that need to be considered to a

small set. However, during lookahead search, we also need to consider all possible observations that result from each action. To keep the computational burden manageable, we developed a method for clustering observations into a very small, but useful, set.

We demonstrate the effectiveness of the resulting decision procedures both in simulation and in a real robotic grasping application.

2 Related work

Our work fits within the general paradigm of motion planning under both sensing and control uncertainty. This problem has been addressed in non-probabilistic formulations (for example, Lozano-Perez et al. 1984; Taylor et al. 1988; Latombe 1991) and in probabilistic formulations (for example, LaValle 2006). Several previous approaches have used probabilistic state estimation to represent uncertainty and integrate observational information in manipulation problems (Petrovskaya and Ng 2007; Gadeyne et al. 2005). Hsiao et al. (2007) frame the decision-making problem as a partially observable Markov decision process and solve it near-optimally, but can only address small problems. Cameron and Durrant-Whyte (1990) take a hypothesis-testing approach, applied to simple probes of a two-dimensional object.

Using tactile sensing to recognize and/or locate objects has a long history (Okada and Tsuchiya 1977; Allen and Michelman 1990; Erdmann 1998; Akella and Mason 1999), yet tactile sensing is used less often in robot manipulation than vision or range sensing. One possible reason is efficiency. Most work on tactile sensing has focused on recognizing/localizing objects in a task-independent manner and can be unnecessarily slow. Our goal is to integrate tactile sensing with the manipulation task, both in that the sensing arises from task-oriented motions and that the goal is to sense just enough to enable success on the task.

There are two paradigms for tactile recognition/localization. One obtains dense data, for example by surface scanning (Allen and Bajcsy 1985; Okamura and Cutkosky 2001); the other uses sparse data directly via “contact probes” (Gaston and Lozano-Perez 1984; Skiena 1989). Within the probe paradigm, there has been substantial work on “active” probing, choosing motions that best disambiguate among possible objects or poses (Ellis 1992; Skiena 1989). However, these probing motions have not typically been integrated into the goals of an overall manipulation task. In this paper, we use task-directed motions as primary probes, resorting to explicit information-gathering motions only when necessary.

Our work is related to the idea of “active localization” of Erickson et al. (2008). Their goal is a plan to localize a

robot in a known map from the expected contacts that result from “move-until-contact” commands. They maintain a belief state and use entropy as a heuristic for picking among actions. One key difference is that we employ world-relative trajectories and on-line belief updates to adjust to on-line outcomes, rather than planning off-line for a fixed action sequence.

A related class of problems is one where there is uncertainty in the outcomes of actions, but the uncertainty is immediately resolved through observations. Alterovitz et al. (2007) construct and solve such an MDP model for guiding non-holonomic needles. There has been a great deal of recent work on generalizations of the motion planning problem that take positional uncertainty and the potential for reducing it via observations into account and plan trajectories through the space that will maximize the probability of success or related other objectives (e.g. Prentice and Roy 2007; Gonzalez and Stentz 2005; Censi et al. 2008; Melchior and Simmons 2007). Burns and Brock (2007) have a different model, in which the system selectively makes observations to reduce uncertainty during planning, but the resulting plan is executed open-loop.

3 Action selection

We will start by framing the problem of placement under uncertainty decision-theoretically, describe strategies for selecting sensing actions and for terminating the sensing phase, and then characterize some theoretical properties of these strategies.

For concreteness, the rest of the paper discusses the problem of grasping an object using a pre-specified grasp (as shown in Fig. 1), when there is uncertainty about the position of the object with respect to the robot, but it is important



Fig. 1 Goal grasps for all experimental objects except the power drill

to keep in mind that the basic formulation is more broadly applicable.

3.1 Problem formulation

A grasp specification, G , consists of a set of relative poses for the hand and object, any of which is a successful grasp. The object is modeled as a 3D polygonal mesh; it is assumed to be positioned on a horizontal table with known z coordinate, resting on a known stable face. There are three degrees of pose uncertainty: x , y , θ . We call this space of object poses W .

A non-contact system (such as vision) generates an initial probability distribution over W ; the sensing and estimation process will refine this distribution over time. The distribution will, in general, be multi-modal; it could be represented with various non-parametric or mixture distributions, but for simplicity, we use a multinomial distribution over a uniform discretization of W . This is our belief state, b ; $b(s)$ is the probability of state s in distribution b .

The robot is a Barrett Arm and Hand, and the space of possible actions is enormous, if viewed as a space of trajectories or velocity commands. Thus, we use actions (a) drawn from a small set of WRTs, as introduced above and described below. The pre-determined set of WRTs will, in general, include a trajectory that is intended to carry out the target grasp, as well as additional trajectories that are designed to be useful for gaining information and for re-positioning the object.

When the robot has executed a WRT from the starting pose through termination, an observation (o) is composed of: the path the robot took through joint-angle space, estimated position and normal of the contact for each finger, obtained from 6-axis force-torque sensors on each of the three finger tips, and readings from binary contact sensors on the palm and hand.

3.2 State estimation

After taking a new action a in belief state b , with underlying states s , and making observation o , the new belief state $b' = SE(b, a, o)$ with underlying states s' is defined by

$$SE(b, a, o)(s') = \frac{\Pr(o|s', a) \sum_s \Pr(s'|s, a)b(s)}{\Pr(o|b, a)}. \quad (1)$$

The first factor in the numerator of (1) is an element of the observation model, $P(o|s', a)$, that specifies the probability of making an observation o after arriving in state s' by using action a , and the second factor is an element of the state transition model, $P(s'|s, a)$, that specifies a probability distribution over the resulting state s' , given an initial state s and action a . The denominator is determined by the constraint that the elements of b' must sum to 1.

3.3 World-relative trajectories

A *world-relative trajectory* (WRT) is a function that maps a world configuration $w \in \mathcal{W}$ into a sequence of Cartesian poses for the robot's end effector. In the simple case in which w is the pose of an object, a world-relative trajectory can just be a sequence of end-effector poses in the object's frame. Given a WRT τ and a world configuration w , the sequence of hand poses $\tau(w)$ can be converted via inverse kinematics (including redundancy resolution) into a sequence of waypoints for the arm in joint-angle space. So, if we knew w exactly and had a valid (collision-free and reachable) WRT for it, we could move the robot through the hand poses in $\tau(w)$ and reach the desired end-point configuration of the arm with respect to the object for that WRT. The first point on every trajectory will be the same "home" pose, in a fixed robot-relative frame, and the robot will begin each trajectory execution by moving back to the home pose ϕ_h .

In general, we won't know w exactly, but we will have to choose a single w to use in calculating $\tau(w)$. Let $\hat{w}(b)$ be the world state w for which $b(w)$ is maximized; it is the most likely state. We can execute $\tau(\hat{w}(b))$, and have the highest probability of reaching the desired terminal configuration according to our current belief state. We command the robot to follow the trajectory by executing *guarded move* commands to each waypoint in the sequence, terminating early if a contact is sensed. An early contact (or reaching the end of the trajectory with no contact) results in an observation that can be used to update the belief state. In addition to the collision point, we obtain further contact observations by carefully closing each finger until contact when any collision is sensed.

3.4 Decision procedures

An optimal behavior in this problem is a decision procedure that specifies which action to take in reaction to every possible belief state. We can assign a cost to taking actions, and then seek a decision procedure that minimizes the expected cost of the robot's behavior, taking into account both the cost of executing actions for the purpose of gaining information and the cost of grasping the object incorrectly (including failing to grasp it at all). This problem is a POMDP (Kaelbling et al. 1998), and can be computationally very difficult to solve in the general case. It requires time and space exponential in the number of actions and observations which, even if discretized, are enormous.

We present two approximate, but efficient approaches to this problem. The resulting behaviors are sub-optimal, in general. However, we seek to understand conditions that guarantee finite convergence to a desired grasp with high probability.

3.4.1 Single WRT

The simplest decision procedure assumes that we have a single WRT, τ^* , which, when executed with respect to the true object pose w^* , results in a grasp in G , the set of goal grasps. Letting b_t be the *belief state* at time t , that is, $\Pr(W = w | o_1, \dots, o_t, a_1, \dots, a_t)$, the distribution over possible object poses given the history of observations and actions, the decision procedure is to:

Step 1. Find the maximum *a posteriori* probability (MAP) pose $\hat{w}(b) = \arg \max_w b(w)$ and execute $\tau^*(\hat{w}(b))$. That is, to execute the grasping trajectory as if the object were at its most likely location.

Step 2. Obtain observation o_{t+1} and update the belief state.

Step 3. Terminate when a criterion on b is met, grasp the object using $\tau^*(\hat{w}(b))$ (if the hand is not already grasping at the new $\hat{w}(b)$), and pick it up.

The termination criterion depends on the expected loss of attempting to grasp based on the current belief state. Our loss function for executing a final grasp $\tau^*(w)$ and attempting to pick up the object is 0 if $\tau^*(w)$ results in a goal grasp and 1 otherwise. The expected loss, or *risk*, of executing $\tau^*(w)$, written $\rho(\tau^*(w), b)$, is an expectation of the loss taken with respect to belief distribution b .

We should select the w that minimizes $\rho(\tau^*(w), b)$ to parametrize the final grasp; that is the action that is optimal in the expected-loss sense. In practice, it can be expensive to evaluate ρ over the whole space W , so we commit to executing action $\tau^*(\hat{w}(b))$. Our termination condition is that the risk of this action be less than some risk threshold δ . In Sect. 3.5, we describe conditions under which this process requires a finite number of samples, in expectation, to terminate. It must be the case that each new grasping attempt yields information that ultimately decreases the risk.

3.4.2 Multiple WRTs

It may be that repeatedly executing τ^* will not give sufficient information to achieve the goal criterion. If the goal is to pick up a long object in the middle, repeated grasping will not give information about the object's displacement along its long axis. Thus, it is necessary to touch additional surfaces with the explicit purpose of gaining information. It may also be that the goal WRT is not executable: the object's handle, for example, may be out of reach of the robot, in which case the robot must first grasp the object using a different grasp and re-orient it. Therefore, we will generally have available a set of WRTs.

Given a set of possible WRTs, τ_i , each of which has execution cost c_i , the decision procedure uses *finite-lookahead search* in an attempt to minimize the total cost of information gathering actions required to achieve the goal criterion.

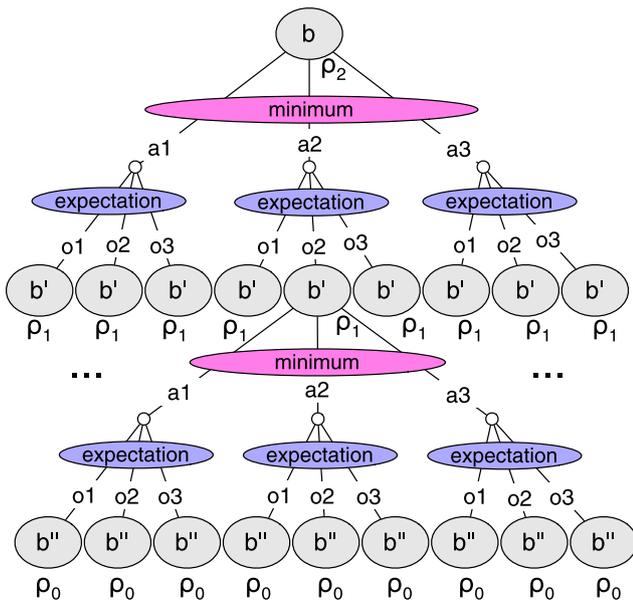


Fig. 2 Parts of a depth-2 search tree and associated value computation

Step 1. Construct a search tree whose layers alternate between action choices of the robot and stochastic outcome “choices” of the environment. The root node is the current belief state, b . It branches on the choice of action a , and then on the possible observations o . The node reached from b via action a and observation o is a new belief state $b' = SE(b, a, o)$. This process can be carried out to any depth; Fig. 2 shows parts of a depth-2 tree.

For computational simplicity, the only actions considered in belief state b are $\tau_j(\hat{w}(b))$, for each $\tau_j \in \mathcal{T}$. We consider each possible WRT only with respect to the world state that is the most likely in b . The observation space is actually continuous; Sect. 4.4.1 discusses the discretization process.

We are interested in $V(b)$, which is the expected cost of executing actions, starting in belief state b , until reaching a belief state for which $\rho_0(b)$, the probability of having a wrong grasp (one not within the desired goal region) in belief state b , is less than δ . We use the tree to compute a finite-horizon approximation of $V(b)$ at a given b , based on backward induction. The value at the leaves is approximated as the risk of terminating and using the goal grasp in that belief state: $V_0(b) = \rho_0(b)$. Then, we compute

$$V_n(b) = \min_j \left[\sum_o \Pr(o|b, \tau_j(b)) U_n(SE(b, \tau_j, o), \tau_j) \right],$$

where

$$U_n(b, \tau) = \begin{cases} \rho_0(b) & \text{if } \tau = \tau^* \text{ and } \rho_0(b) < \delta, \\ V_{n-1}(b) + c & \text{otherwise.} \end{cases}$$

The function U captures the fact that if the belief state meets the termination criteria, then the procedure terminates and

there is no remaining action cost. If it does not, then the expected cost is as defined by V for the belief resulting from the action taken, plus the cost of the additional action.

Step 2. Select the WRT τ_j that minimizes the expected value of V at the next step and execute it with respect to $\hat{w}(b)$.

Step 3. Obtain observation o_{t+1} , and update the belief state.

Step 4. Terminate when $\rho_0(b) < \delta$ after having grasped the object using τ^* , and pick it up. Otherwise, go to step 1.

In most situations, when re-positioning the object is not necessary, it is sufficient to set search depth $k = 1$, that is, to select the WRT that will, in expectation, lead to a belief state that has the least risk with respect to executing the nominal grasp trajectory. However, as we illustrate experimentally in Sect. 5, looking deeper can improve the rate at which the termination criterion is reached. It is sometimes the case that an initial action, although it does not itself reduce the risk significantly (because it reduces uncertainty in a dimension that is not important for the ultimate grasp), makes it possible to execute a subsequent action in a way that significantly improves its information-gathering effectiveness. As k goes to infinity, this process will choose optimal actions (DeGroot 1970). However, we find experimentally that in our application, increasing k beyond 2 yields no advantage.

In many information-gathering domains, the objective function is *submodular* in the set of observations. This means that subsequent observations do not offer as much incremental value as early observations. Krause and Guestrin (2007) have shown that when this property holds, a greedy observation strategy, which only looks ahead one step in picking observations, is within a constant factor of optimal. However, our domain does not have that property: it can happen that, even though the risk is nearly independent of uncertainty in the x dimension, that first localizing an object in x makes it much more efficient to localize in y ; thus, the initial observation may have less risk-reduction than a subsequent observation.

3.5 Termination and correctness

We would like to understand how these decision procedures are likely to perform, depending on properties of the domain to which we apply them. There are two important questions: Will the procedure terminate in finite time? When it terminates, what is the likelihood that it will have selected a final action that meets the goal criterion? The answer depends on the informativeness of the observations and on the degree to which the actions change the pose of the target object.

3.5.1 Single WRT with object fixed

We begin by assuming that the object’s pose does not change during the sensing process, that the loss is zero if and only if

the goal WRT is executed with respect to the true underlying state, and that we have only a single WRT.

In the standard statistics problem of hypothesis testing with two hypotheses and a single observation action, the sequential decision process depends on the ratio of the probabilities of the two hypotheses in the posterior distribution. When the probability ratio goes outside of a fixed interval, then the sampling procedure is terminated and a hypothesis is selected. This procedure was shown by Wald (1945) to terminate with a guaranteed risk after a finite number of trials, as long as the observation distribution is almost surely (with probability one) different, conditioned on the hypothesis.

This result is extended to the case of multiple discrete hypotheses in the M-ary sequential probability ratio test (MSPRT), which is also guaranteed to terminate after finitely many trials whenever the observation distributions are almost surely different for each hypothesis (Baum and Veeravalli 1994). These results hold whether the observation space is discrete or continuous. However, our decision problem deviates from the MSPRT setting, in that different “experiments” are being chosen on each step, because each time the belief state is updated, the MAP world state is likely to change, and so the WRT is executed with respect to a different hypothesis about the object’s pose.

Define the observation probability distribution for action $a = \tau^*(w_l)$ (the goal grasp applied to hypothesized world state w_l) when the true world state is w_k to be

$$f_{jk}^l(o) = \Pr(O = o | W = w_k, A = \tau^*(w_l))$$

and the *expected informativeness* of an action $\tau^*(w_l)$ to the distinction between states j and k to be:

$$I_{jk}^l = E_{f_k^l} \left[\sqrt{\frac{f_j^l(o)}{f_k^l(o)}} \right] = \int_o \sqrt{f_j^l(o)} \sqrt{f_k^l(o)} do.$$

A clearly sufficient condition for termination is that, for all pairs of world states w_j, w_k , for all w_l , $I_{jk}^l \neq 1$; that is, that no matter how we parametrize our WRT, the observation distribution that it generates will be different across each pair of possible world states w_j, w_k . This satisfies the conditions for termination of the MSPRT. Of course, this won’t be true in general. If w_l is a pose that is spatially non-overlapping with both w_i and w_j , then no matter whether w_i or w_j is true, the probability of observing *no contact* is 1 (or high) when the robot attempts to grasp the object as if it were at w_l .

If w_k is the true state of the world, we can plausibly assume that, for all world states $j \neq k$:

$$I_{jk}^j < 1 \quad \text{and} \quad I_{jk}^k < 1. \tag{2}$$

What this means is that grasping as if the object were at pose j provides information that differentiates j from the

true hypothesis k ; and that grasping as if the object were at the true pose k provides information that differentiates pose k from all other possible poses. This is, of course, not true for many objects; if it is not, then the single-WRT process will not terminate and we need to use multiple WRTs.

It is important to see that taking uninformative actions is never destructive to the estimation process: it is simply a no-op. So, when the information ratios in (2) are less than 1, we can apply the MSPRT theorem as follows: In any sequence $a = a_0, \dots, a_N$ where $n_j(a)$ of the trials in a are of action $\tau^*(w_j)$, the probability that the decision procedure will not have terminated is a constant factor times $\max_j (I_{jk}^j)^{n_j(a)} (I_{jk}^k)^{n_k(a)}$. It will be sufficient to focus on the number of trials, $n_k(a)$, of $\tau^*(w_k)$ (which we will abbreviate as action k).

By a similar argument to the termination of the MSPRT, we can argue that any action other than action k will be exponentially unlikely to continue to be selected during an execution. So, very quickly, action k will predominate. The argument is as follows: any hypothesis j that is currently more likely than hypothesis k will be selected, but because $I_{jk}^j < 1$, it will drive down the likelihood of j with respect to k exponentially quickly, and therefore eventually not be selected. This will happen for each hypothesis $j \neq k$, until hypothesis k is selected. In case the likelihood of some j rises up above k , then action j will be selected, and it will drive the likelihood of j back down exponentially quickly.

The statistics literature provides arguments that probability ratio tests can be configured (by choosing termination criteria appropriately) to minimize total risk (when a cost is assessed for each sensing action). In our case, it is guaranteed that the risk of the final action is less than the maximum tolerable risk δ whenever the procedure terminates.

3.5.2 Goal sets

The goal set G may be such that executing $\tau^*(w)$ has 0 risk in a whole set of world states, not just the true state; for example, we may be indifferent about the orientation of a round can when it is grasped. In such a case, the requirements from formula (2) can be weakened. Let W_G be the set of $w \in W$ such that the grasp resulting from executing $\tau^*(w)$ is in G . Then, it is sufficient for termination that for all $j \in W \setminus W_G$ and for all $k \in W_G$, $I_{jk}^j < 1$ and $I_{jk}^k < 1$; that is, that the actions are discriminative between goal and non-goal w , but need not be discriminative within those sets.

3.5.3 Observations can move the object

It is more generally the case that the observation actions can change the state of the world, by moving the object as it is being sensed. If the information gained by each observation update compensates sufficiently for any additional entropy

in the belief state introduced by the transition update, such that $\Pr(w_j)/\Pr(w_k)$ does not increase, in expectation, then the decision procedure will have a finite expected duration despite the object movement.

3.5.4 Multiple WRTs

We must increase the set of WRTs when the requirements from (2) are not satisfied for τ^* . Given a set of WRTs \mathcal{T} , it must be that, for any pair of world states $j \in W \setminus W_G$ and $k \in W_G$, there exists some $\tau \in \mathcal{T}$ such that $I_{jk}^{j,\tau} < 1$ and some (not necessarily the same or different) $\tau \in \mathcal{T}$ such that $I_{jk}^{k,\tau} < 1$, where we have extended the definition of I to depend on τ in the obvious way.

We need to perform lookahead search when the execution of any single WRT is insufficient to yield an immediate decrease in risk. We can treat length- k sequences of the original actions as a new set of “macro” actions. If these macro actions, in combination, satisfy the requirements on informativeness, then the procedure will terminate and generate correct answers with high probability.

4 Implementation

We implemented this method with a 7-DOF Barrett Arm and Hand, both in simulation (using ODE to simulate the physics of the world) and on an actual robot. The hand has ATI Nano17 6-axis force/torque sensors at the fingertips, and simple binary contact pads covering the inside (and some outside) surfaces of the fingers and the palm.

To apply the described framework, we needed to find an appropriate set of WRTs, and to define observation and transition models. In addition, because of the size of the set of possible raw observations and the fact that the lookahead search branches on all possible observations, we developed a method for clustering observations into a very small, but useful, set.

4.1 WRTs

We used WRTs of three different types, for each target:

- the goal WRT, τ^* , that grasps the object correctly if executed with respect to the correct world state,
- information WRTs, that attempt to contact non-goal surfaces of the object and that sweep through the space to make an initial contact when uncertainty is high, and
- re-orientation WRTs that use a grasp from above to rotate the object about its center of mass to make the goal WRT kinematically feasible.

The goal WRTs were generated through demonstration (by moving the robot and recording object-relative waypoints).

Some of the information WRTs were also generated this way; others were constructed automatically.

Automatic generation of WRTs was done by finding hand positions that place the fingers on nearly parallel pairs of object surfaces, using the OpenRAVE motion planning system (Diankov and Kuffner 2008) to find a collision-free trajectory from a starting pose to that hand position, and then expressing the trajectory in object-relative Cartesian coordinates. The resulting trajectories become candidate WRTs, which were then evaluated on the basis of their kinematic robustness, their potential to gain information, and their tendency to fail by hitting the object with a sensorless part of the hand.

We also added an additional information-gathering WRT that simply sweeps horizontally across the workspace, which is useful for initially locating the object when there is high uncertainty. The implicit goal in creating a set of WRTs is to satisfy the termination requirements of the sequential decision procedure by ensuring that, for any two possible poses of the object, there is a WRT (or a short sequence of them) whose observation distribution distinguishes between them. In the case where the workspace is expected to be cluttered, we might wish, in addition, to generate multiple WRTs for gaining the same information, which would allow the process to succeed even when some WRTs are infeasible due to possible collisions.

We can also execute trajectories that try to change the actual state of the world, rather than merely trying to reduce uncertainty in the belief state. If all of our goal-achieving trajectories are kinematically infeasible in $\hat{w}(b)$, for instance, we may wish to reorient the object so that at least one of them becomes feasible. To do so, we add a WRT that attempts to grasp the object (using a grasp that does not necessarily satisfy the goal conditions) and then rotates the object after successfully running to completion. In our implementation, all reorientation WRTs simply grasp the object from the top, about the object center, and rotate by the desired angle.

4.2 Belief state representation and update

We represent the belief state as a multinomial distribution over a three-dimensional grid of cells, with the x , y , and θ coordinates discretized into 31, 31, and 25 cells, respectively. A cell w comprises a set of actual poses. To handle this correctly in the transition and observation models, we should integrate over poses within w , which is computationally difficult. We instead treat cell w as if it were the pose in the center, which we call the *canonical pose* and write as \bar{w} .

The belief-state update operation is, in the worst case, quadratic in the size of the state space; but in our case the transition distribution is very sparse, so there are a bounded number of states s' such that $\Pr(s'|s, a)$ is non-zero, which

makes the complexity ultimately linear in the size of the state space instead. Furthermore, we are assuming that the belief state is also quite sparse (due to initial sensor information), so the complexity is further reduced considerably, making this operation straightforward to compute online.

4.3 Observations

In general, a WRT is executed in two phases: the arm is moved through the specified trajectory until a contact is felt anywhere on the hand or until the trajectory completes; then the three fingers are closed, each one terminating when it feels a contact or when it is fully closed. Each of these four motions (arm and three fingers) is treated as generating an observation tuple: $\langle \phi, c, \psi \rangle$, where ϕ is the observed pose of the robot at termination of the WRT based on the robot's proprioceptive sensors, c is a vector of readings from the contact sensors, and ψ is a representation of the 'swept path', that is, the volume of space through which the robot thinks it moved (based on proprioception) during the course of executing the WRT.

4.3.1 Observation model

For the purpose of belief-state update, we must specify an observation model, which is a probability distribution

$$\Pr(O = \langle \phi, c, \psi \rangle | W = w, A = \tau_i(w_j)).$$

The size and complexity of the underlying state and observation spaces makes the modeling quite difficult. For tractability we make several assumptions:

- There are no actual contacts that are not noticed or false triggering of the contact sensors.
- The information gained from each part of the robot (arm, fingers) is independent given the world state and action.
- The swept path information is independent from the position and contact information.
- The contact information at different contact points is independent.

In order to connect the observations to an underlying world state, it is necessary to reason about the (unobserved) true trajectory, ψ^* that the robot took. Letting τ_c stand for the commanded trajectory, we can write the observation probabilities $\Pr(\phi, c, \psi | w, \tau_c)$ as

$$\int_{\psi^*} \Pr(\psi^* | w, \tau_c) \Pr(\phi, c | \psi^*, w, \tau_c) \Pr(\psi | \psi^*, w, \tau_c, \phi). \tag{3}$$

The robot is driven by a servo control loop that causes ψ , the observed trajectory, to track τ_c , the commanded trajectory, quite closely, so we can assume that, in the last term,

$\psi = \tau_c$ (or, a prefix thereof, terminated at ϕ) and that it has probability 1. This integral is too difficult to evaluate, and so we approximate it by the maximum:

$$\Pr(\phi, c, \psi | w, \tau_c) \approx \max_{\psi^*} \Pr(\psi^* | w, \tau_c) \Pr(\phi, c | \psi^*, w, \tau_c).$$

The scale of this value will be considerably different from the actual probability, but it can be normalized in the belief update.

No contact: When the robot observes no contact with the object, we assume that there was, in fact, no contact; so we consider only trajectories ψ^* in which no contact would occur. The value of ψ^* that is most likely, then, is the non-colliding trajectory that is as close to the observed ψ as possible.

Letting $d^*(\psi, w)$ be the depth of the deepest point of collision between ψ and the object at pose \bar{w} , assuming that the nearest collision-free trajectory is at distance $d^*(\psi, w)$ from ψ , and assuming that the likelihood of an observed trajectory is described by a Gaussian on its distance from the actual trajectory, we have

$$\Pr(\phi, None, \psi | w, \tau_c) \approx \mathcal{G}(d^*(\psi, w); 0, \sigma_p^2) = P_f(\psi, w),$$

where \mathcal{G} is the Gaussian density function and σ_p^2 is a variance parameter. Although this approximation is efficient to compute, it can be inaccurate: there are situations in which the collision depth is small, but the distance between the sensed trajectory and the nearest non-colliding trajectory is quite large.

Contact: The observation probability, in the case of an observed contact, is the probability that as the robot executes commanded trajectory τ_c , that it will sense no contact up until ϕ , and then that it will sense the contacts c . We approximate the maximum of a product as a product of the maxima:

$$\begin{aligned} \Pr(\phi, c, \psi | w, \tau_c) &\approx \max_{\psi^*} \Pr(\psi^* | w, \tau_c) \Pr(\phi, c | \psi^*, w, \tau_c) \\ &\approx P_f([\psi], w) \max_{\phi^*} \Pr(\phi, c | \phi^*, w) \end{aligned}$$

where $[\psi]$ is the swept path, minus a short segment at the end, and ϕ^* is the final pose of ψ^* .

The fingertips have 6-axis force/torque sensors that are used to estimate the position and orientation of contacts, so each contact can be written as the pair $l_i(\phi, c), n_i(\phi, c)$, representing the location and normal of contact i . To do a careful job of estimating the probability of the contact, we would have to consider each pose ϕ^* , or possibly each face of the object, to find the most likely contact. Instead, as a fast approximation, we assume that the sensor reading was caused by contact with the closest object face, f^* , to $l_i(\phi)$, assuming the object is at pose \bar{w} . This choice maximizes the probability of the location, but not necessarily the normal. The

final model is

$$\Pr(\phi, c, \psi | w, \tau_c) \approx P_f(\lfloor \psi \rfloor, w) \mathcal{G}(d_l(l_i(\phi), f^*); 0, \sigma_l^2) \times \mathcal{G}(d_n(n_i(\phi), f^*); 0, \sigma_n^2)$$

where $d_l(p, f)$ is the Euclidean distance from point p to face f and $d_n(n, f)$ is the angle between the vector n and the normal to face f , and σ_n^2 and σ_l^2 are variance parameters.

We model the fingertip position error with a standard deviation of 0.5 cm, and the fingertip normals with a standard deviation of 30 degrees. Each contact pad on the rest of the hand is a thin rectangle, approximately 2–3 cm on a side, consisting of two raised metal plates separated by small patches of very light foam (placed in notches cut in the raised metal plates), that sense when the metal plates touch. While they are very sensitive to light contacts, they do not provide an estimate of contact location on the pad. Thus, we model the location of a pad contact as the center of the contact pad, with a standard deviation of 1 cm, and we model the normal as the pad's surface normal, with a standard deviation of 90 degrees.

4.4 Computing the observation model

Computing the observation model requires predicting the sensory conditions (e.g., finger contacts) that can result from executing a given WRT in a given state. In an off-line process, for each WRT τ , we construct a representation of the observation function, $\Omega_\tau(w, e)$, which specified the nominal observation tuple $\langle \phi, c, \psi \rangle$ described earlier. Observation functions are indexed by an actual world configuration w and an estimated world configuration e , specifying what would happen if $\tau(e)$ were executed in world w ; that is, if the robot acted as if the world were in configuration e , when in fact it was in configuration w . In the case of a single object with a canonical support surface on a table, the space of w and e is characterized by the (x, y, θ) coordinates of the object (although the approach can also be applied more generally).

Figure 3 shows the $\Omega_\tau(w, e)$ function for a WRT τ and a space of 3 world configurations, and how it is determined. Each row corresponds to a different true pose (x, y, θ) of the object in the world (w), which is drawn in blue. Each column corresponds to a different estimated pose of the object (e), which is drawn in red. On the diagonals, the true and estimated poses are the same, so the figures lie on top of one another. The estimated pose e determines the trajectory $\tau(e)$ that the robot will follow (in this case, our robot is a point robot in x, y). The trajectories are shown in black. Each one starts from the same home pose, shown in green, and then moves to a sequence of waypoints that are defined relative to the estimated pose of the object. Yellow circles indicate situations in which the robot will make contact with the object. It happens on each of the diagonal elements, because

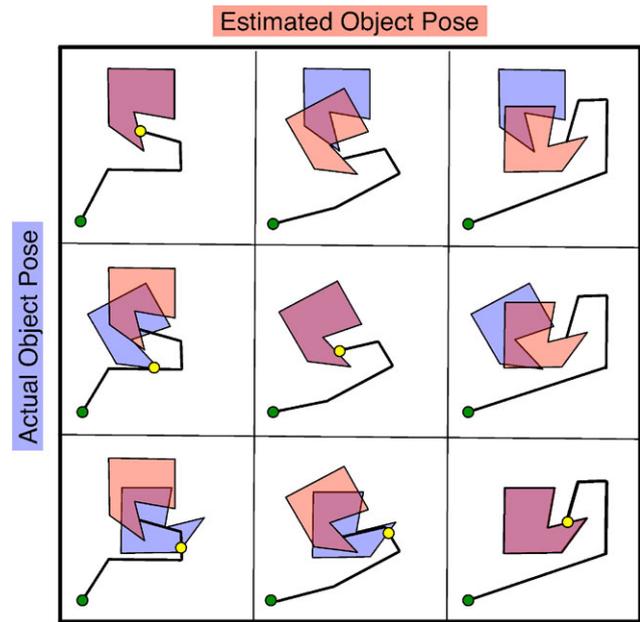


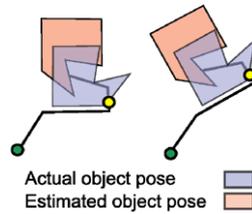
Fig. 3 (Color online) The $\Omega_\tau(w, e)$ matrix for a WRT τ

the nominal trajectory makes contact with the object. In the elements in the bottom-left part of the figure, there is a contact between the robot and the actual object during the execution of the trajectory, before it would have been expected if the estimated pose had been the true one. In the elements in the upper right part of the figure, the trajectory terminates with no contact. In all cases, the observation gives information about the object's true location, which is used to update the estimated pose.

Computing an entry of these matrices requires simulating a trajectory forward from a starting robot pose, and calculating if and when it contacts objects in the world, and, if it does, what the nominal sensory readings will be in that situation. This is a geometric computation that can be done entirely off-line, relieving the on-line system of performing simulations. Having computed the nominal observations in advance means that the observation probabilities required for the on-line belief-state update can be calculated with little additional work.

This computation may seem prohibitive, since for a x, y, θ grid of just $31 \times 31 \times 25 = 24,025$ points, having to simulate all combinations of w and e in pairs would require $24,025^2 = 207,792,225$ simulations. However, the crucial insight here is that if trajectory $\tau(e)$ is kinematically feasible and there are no other objects nearby, then the observation depends only on the relative transformation between w and e , as shown in Fig. 4. For two sets of w and e with the same relative transformation, as with the examples in the figure, w and $\tau(e)$ may differ, but $\Omega_\tau(w, e)$, which is expressed relative to e , is the same. Thus, when calculating the full $\Omega_\tau(w, e)$ matrix for a WRT τ , we can pick a

Fig. 4 (Color online) Predicted observations depend only on the relative transformation between the actual and estimated object pose



single e (for instance, the initial $w^*(b)$), compute $\tau(e)$, and simulate just that sequence of robot poses while varying w . The number of simulations required to compute $\Omega_\tau(w, e)$ is therefore merely the number of points in the belief grid that have nontrivial probability, and running them takes just a few seconds. Once the simulations are completed, the results can be stored for fast re-use when selecting actions on-line.

4.4.1 Reduced observation space

The observation distributions are all Gaussians centered at nominal observations, made assuming that the object is at one of the canonical poses \bar{w} for $w \in W$. We can take this discrete set of nominal observations (which in our implementation has about 24000 elements) as our set of possible observations. However, if we are to do lookahead search that branches on observations, then we must group and subsample the observation space so that it is much smaller.

In this section, we describe a very aggressive process for finding a small set of canonical observations to branch on during search. We always use the full observation, with the model described above, when doing belief-state estimation during execution.

The purpose of the lookahead search is to select actions that are most useful for gaining information. Reducing uncertainty in the orientation of the object, for example, is just as useful, no matter what the object's position is. Therefore, we can ignore the arm position at contact. For efficiency, but with a potential significant loss of effectiveness, we also ignore the information gained from ψ , the swept-volume aspect of the observation. So, we focus on reducing the space of possible contact observations c .

Clustering: We start by clustering directly contact observation vectors, that are close enough in terms of both Euclidean distance between the contact locations and the angle between the contact normals; these distances are assumed to be infinite if one of the contact measurements is *None*. Each of the resulting clusters of observations is represented by its most likely observation.

Sub-sampling: Next, we prune observation clusters that are unlikely to occur throughout the state estimation process. In the early parts of the estimation process, we expect our belief distributions to be quite diffuse. Later in the estimation process, we expect the belief distributions to be concentrated around particular values of w , but, of course, we

don't know which w . We do, however, know that later in the process, \hat{w} will be near the true w , so many observations will be unlikely because they result from actions that are ill-matched to the true hypothesis.

We consider three different levels of uncertainty: the initial diffuse Gaussian, an intermediate one in which the standard deviations are reduced by 1/5 in each dimension, and a focused one in which the deviations are quite small. We prune any observation clusters that have less than 0.01 probability of occurring under any of the three distributions. This reduces the observation cluster space in our examples to on the order of 50 elements.

Belief-dependent clustering: In the context of a particular search step, we have a current belief state b , and an action $\tau(\hat{w}(b))$. We can cluster observations based on their effect in this particular belief state. We do a further agglomerative clustering, grouping observations that lead to belief states that have similar variances in each dimension (on the canonical observations of the clusters from the previous two steps), where the distance metric is

$$d(c^1, c^2) = d_b(\text{SE}(b, \tau(\hat{w}(b))), \langle \phi, c^1 \rangle), \text{SE}(b, \tau(\hat{w}(b))), \langle \phi, c^2 \rangle),$$

where SE is the belief-state update process, and d_b is a distance metric on belief states, defined as

$$d_b(b_1, b_2) = |H(b_1) - H(b_2)| + \sum_{r \in \{x, y, \theta\}} |V_r(b_1) - V_r(b_2)|,$$

where H is entropy, and V_x , V_y , and V_θ are variances in the specified dimensions. In the current implementation, this process is only done for the initial belief state, where it gives the greatest leverage and cuts the number of clusters in half.

Belief-dependent sub-sampling: Finally, for any branch of the search tree, we sort all possible observation clusters by their probability in this belief state, and consider only the k most likely observations, whose summed probability is greater than 0.5. This is very aggressive, but it results in a manageable branching factor of between 1 and around 7; in our experiments, increasing the number of considered observations did not noticeably change the performance.

4.5 Transition model

The transition model specifies $\Pr(W_{t+1} = w_j | W_t = w_i, A_t = \tau_c)$. We treat two cases separately: information and goal WRTs, which are not intended to change the object's pose (but which might do so inadvertently), and re-orientation WRTs, which explicitly attempt to move the object.

Before we incorporate information from the observation, the transition distribution is fairly diffuse: there is a chance the robot will miss the object entirely (and therefore leave it in the same pose), that the robot will graze it with one finger

(and cause it to rotate), or that the robot will give it a solid shove. We compute a transition distribution that is already conditioned on some aspects of the observation, because by looking at all of the contact information jointly, we can estimate the force and torques that were applied to the object, and use that information to modulate the transition probabilities. This approach risks over-weighting the observation information, but seems to work well.

When no contacts are observed, we assume that the object was not contacted by the robot, and therefore was not moved. When contacts are observed, the transition distribution is a mixture of two Gaussians, one centered at the object's initial pose and one centered at a pose to which the object may have been "bumped" by the contacts:

$$\Pr(w'_j | w_i, \tau_c) \propto (1 - p_b) \mathcal{G}(d_p(\bar{w}_i, \bar{w}_j); 0, \sigma_s^2) + p_b \mathcal{G}(d_p(\text{bump}(\bar{w}_i, \phi, c), \bar{w}_j); 0, \sigma_b^2),$$

where p_b is the probability that the object will be bumped, d_p is a distance metric on poses that weights 1 cm of distance in position the same as 0.1 radians in rotation, and bump is a function that computes the most likely bump outcome from the old pose and the observation.

The bump pose is determined as follows. For each observed contact $c_i(\phi)$, we compute a unit force vector v_i applied at $l_i(\phi)$, in the direction $-n_i(\phi)$. We determine the center of mass of the object (assuming uniform density), compute the summed force and torque, and assume the object will translate a fixed distance per unit force in the direction of the net force, and rotate a fixed rotation per unit torque.

Because re-orientation has a moderately high probability of failure, we use a modified transition model, which is a mixture of three possible outcomes: the re-orientation fails entirely and the object stays in its initial pose, the re-orientation succeeds exactly, and the re-orientation leaves the object somewhere in between the start and goal poses. Each of the modes has a larger standard deviation than the corresponding standard deviation for other WRTs.

4.6 Single WRT example

Having outlined all the components, we will now show a simple example: Fig. 5 shows the operation of the system while grasping a rectangular box, using a single goal-achieving WRT. The robot attempts to execute a grasping trajectory, relative to the most likely element of that belief state. The first image in the top row shows where the robot thinks the most likely state of the box is relative to its hand. The second image shows the location of the robot at the first waypoint in that trajectory: we can see that the object is actually not in its most likely position (if it were, the robot's hand would be centered above it). The third image in the first row

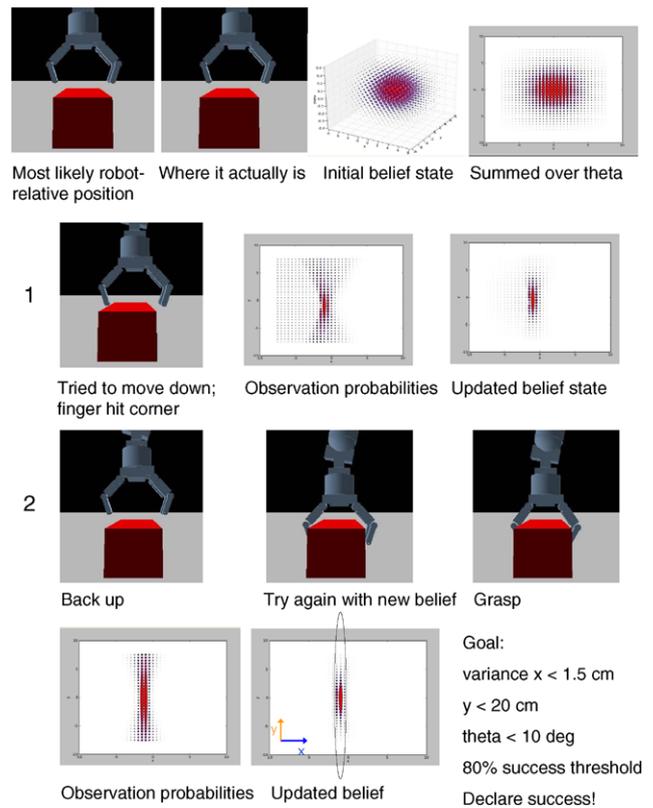


Fig. 5 Execution of WRT and (x, y, θ) belief state update

shows the initial belief state, with probabilities depicted via the radius of the balls shown at grid points on the (x, y, θ) state space of the box. Subsequent belief state images show the belief state summed over θ for easier visualization, as in the fourth image in the first row. In action 1 (row 2), the hand executes a guarded move toward the next waypoint in the trajectory, and is terminated by a fingertip contact on the corner of the box, as shown in the first figure in the second row. The middle figure in the second row shows the probability of observing that fingertip contact in each world configuration. Combining this information with the initial belief state, we obtain the updated belief state shown in the third figure in the second row. It is clear that the information obtained by the finger contact has considerably refined our estimate of the object's position.

The third and fourth rows of figures show a similar process. The same WRT is executed a second time, now with respect to the most likely state in the updated belief state. This time, the hand is able to move all the way down, and the fingers close on the box, with the resulting belief state shown in the final figure. Now, given a goal condition such as having the box centered between the fingers within 1.5 cm and oriented within 10 degrees of being straight, but not being concerned where along the box the fingers are grasping (shown by the oval in the updated belief state), we can evaluate the probability that it holds in the current belief state.

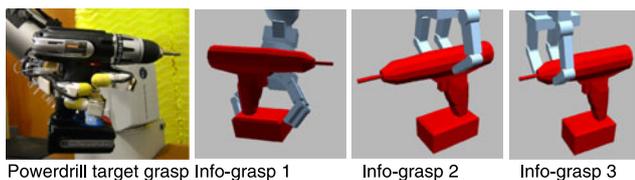


Fig. 6 Goal and information grasps for the power drill

In this case, it holds with probability ≥ 0.8 , so if δ were 0.2, we would terminate.

If, on the other hand, the goal condition also specified where the hand should be along the length of the box (for instance, if the box is heavy and might tip within the hand, or if the placement location requires the hand to grasp at a particular location), then this result would not have satisfied the goal condition, we would have no real way to improve the situation through further execution of our goal-seeking WRT, and the control loop would run forever. In this case, we would need to add an information-gathering WRT that can gather information about the location of the end of the box.

5 Experiments

Our experiments used 10 different objects, shown with their goal grasps in Figs. 1 and 6. The goal region for each object was hand-chosen to guarantee that being within the goal region ensures a stable grasp of the object. These regions are much larger for some objects than for others (for example, the goal region for the can is large, since the hand only has to envelop it), and the goal regions for the rotationally symmetric objects ignore the object orientation. In all experiments, the maximum number of actions allowed was 10; after the 9th action, if the goal criterion was not reached, the goal WRT was executed.

5.1 Simulation

In our simulation experiments, we compare three different strategies:

- *Open-loop*: execute the goal WRT once and terminate;
- *RRG-Goal*: execute the goal WRT repeatedly on the most likely state, terminating when the risk threshold is met; and
- *RRG-Info*: choose among goal, re-orientation, and information WRTs, with a depth k lookahead-search decision procedure, terminating when the risk threshold is met.

Each simulation experiments was carried out with at least 100 trials.

Figure 7 shows the results for experiments carried out in simulation with initial belief state a discretized Gaussian

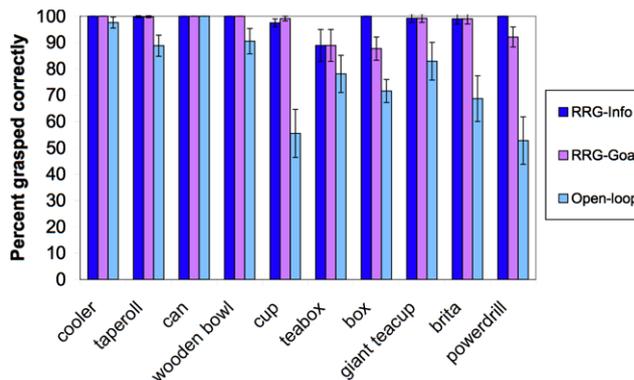


Fig. 7 (Color online) Simulation results for all objects at low uncertainty

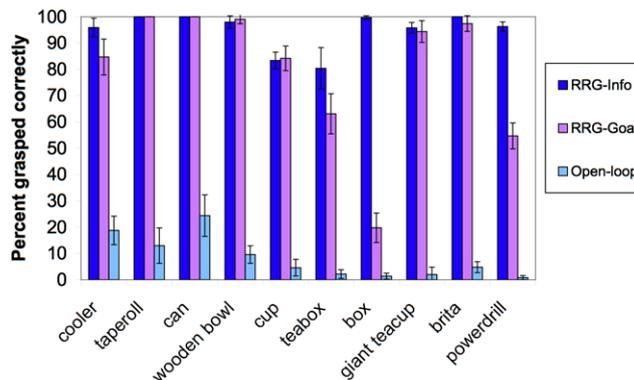


Fig. 8 (Color online) Simulation results for all objects at high uncertainty

with standard deviations of 1 cm in x , 1 cm in y , and 3 degrees in θ , and lookahead-search depth $k = 2$. The chart shows the percentage of grasps that were executed successfully (with 90% confidence bounds), for each object placed at random positions drawn from the initial belief distribution, for the three algorithms. The risk threshold was chosen to correspond to a target success rate of 90%. Even at this low level of uncertainty, executing the goal WRT open-loop fails frequently for many of the objects. Using RRG-Goal allows us to succeed nearly all of the time, and using RRG-Info brings the success rate above 97% for all objects except the tea box, for which the decision procedure only selects the goal WRT, because it recognizes that it will still reach the target success rate.

Figure 8 shows the simulation results for higher levels of initial uncertainty (standard deviations of 5 cm in x , 5 cm in y , and 30 degrees in θ), again with depth $k = 2$. At this level of uncertainty, $\sim 14\%$ of object positions are more than 10 cm away from the initial estimated position, and executing the goal WRT open-loop seldom succeeds. Using just RRG-Goal is sufficient for all of the objects except the box (which is the object used in the example discussed earlier, except with a stringent goal in both x and y , thus requir-

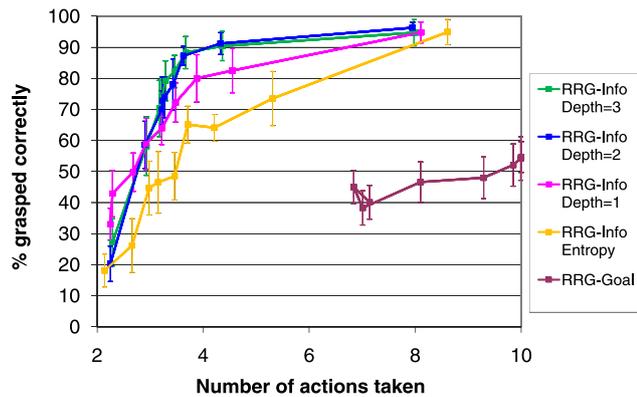


Fig. 9 (Color online) Varying termination criterion trades off between grasp correctness and number of actions required; shown for power drill

ing an information-gathering WRT), and the power drill, for which the goal WRT grasps a nearly-cylindrical handle that gives it little information about the orientation for pressing the trigger. Using RRG-Info brings our success rate above 95% for all objects except the cup and tea box. The lower performance with these two objects is due to the relatively coarse state-space grid in our implementation, which is on the same order of size (1 cm) as the goal set for these relatively small objects. With a grid of such coarseness, it is sometimes the case that the grid point with the highest observation probability is farther from the true object pose than a nearby, lower-probability grid point. This is not an issue if the goal region is relatively large, but does become an issue when the goal region is small. We expect that using an adaptive grid size would improve the performance.

Figure 9 shows the percentage of successful grasps (with 90% confidence bounds based on an assumption of binomial distributions) for the power drill in simulation at the high level of uncertainty (5 cm/30 degree), where the target estimated level of success before termination was varied (from 10% to 90%) to generate data that shows the trade-off between the number of actions executed and the actual success rate. Note that in the left part of the graph we are using a low target level of success, which accounts for the low percentage of actual success. The five strategies used here are RRG-Goal, RRG-Info with lookahead-search depths of 3, 2, and 1, and RRG-Info-Entropy, which is like RRG-Info with lookahead-search of 1, but using entropy of the belief state at the leaves of the search tree (as in Hsiao et al. 2008) instead of risk. Each point on the graph represents the average number of actions taken before termination and the percent success over more than 100 simulated runs. Just executing the goal WRT repeatedly does not work well for this object, whereas searching with a depth of 1 works reasonably well. Note that using risk values at the leaves leads to a substantial improvement over using belief entropy.

Recall that at lookahead-search depth 1, the decision procedure is considering plans with two actions, for example, an information WRT followed by the goal WRT. Increasing the depth to 2 causes the decision procedure to choose actions that may result in a lower probability of success after just 2 actions, but that pay off in terms of a higher probability of success later on. This is due to the fact that, although information WRT 1 (shown in Fig. 6) provides information about all three dimensions at once and information WRT 2 only provides information about two dimensions, information WRT 2 for the power drill acts as a “funnel” for information WRT 1, enabling it to be effective more often. Increasing the lookahead-search depth to 3 yields no additional benefit.

Although, for the drill, lookahead search with depth 1 has a success probability essentially identical to using a depth of 2 or 3 after 10 actions, searching deeper reduces the number of actions needed to succeed more than 90% of the time from an average of 7 actions to an average of 4 actions, which is a dramatic speedup. However, for most objects, using a depth of 1 works as well as using a depth of 2. This is significant since increasing the lookahead-search depth increases the action-selection time exponentially. In our (non-optimized Python) implementation, selecting the first action from among the 5 available power drill WRTs takes 3 seconds for a depth of 1; using a depth of 2 takes 10 times longer, and using a depth of 3 takes 60 times longer.

5.2 Real robot

On the real robot, we ran 10 experiments using RRG-Info with lookahead-search depth of 2, for the Brita pitcher and the power drill with high initial uncertainty. Both objects were grasped stably and lifted successfully 10 out of 10 times, with the trigger being pressed successfully on the power drill and the Brita pitcher being grasped properly by the handle. An example sequence of grasping the power drill is shown in Fig. 10.

For the other 8 objects, we ran four experiments each: one at low uncertainty levels (1 cm/3 deg) and three at high uncertainty levels (5 cm/30 deg). 27 out of the 32 experiments succeeded. Two of the 5 failures (for the cooler and the can) were due to the robot contacting the object in a part of the hand with no sensors. One failure each (for the cup and tea box) were due to the coarseness of the state grid (as in the simulation results in Sect. 5.1). One failure with the giant tea cup was due to due to inaccurate collision depth calculations for objects with thin features (like the tea cup handle) in our swept path computations, which are somewhat coarsely discretized for speed.

Videos of our real robot experiments can be seen here: <http://people.csail.mit.edu/kjhsiao/wrtpomdps>.

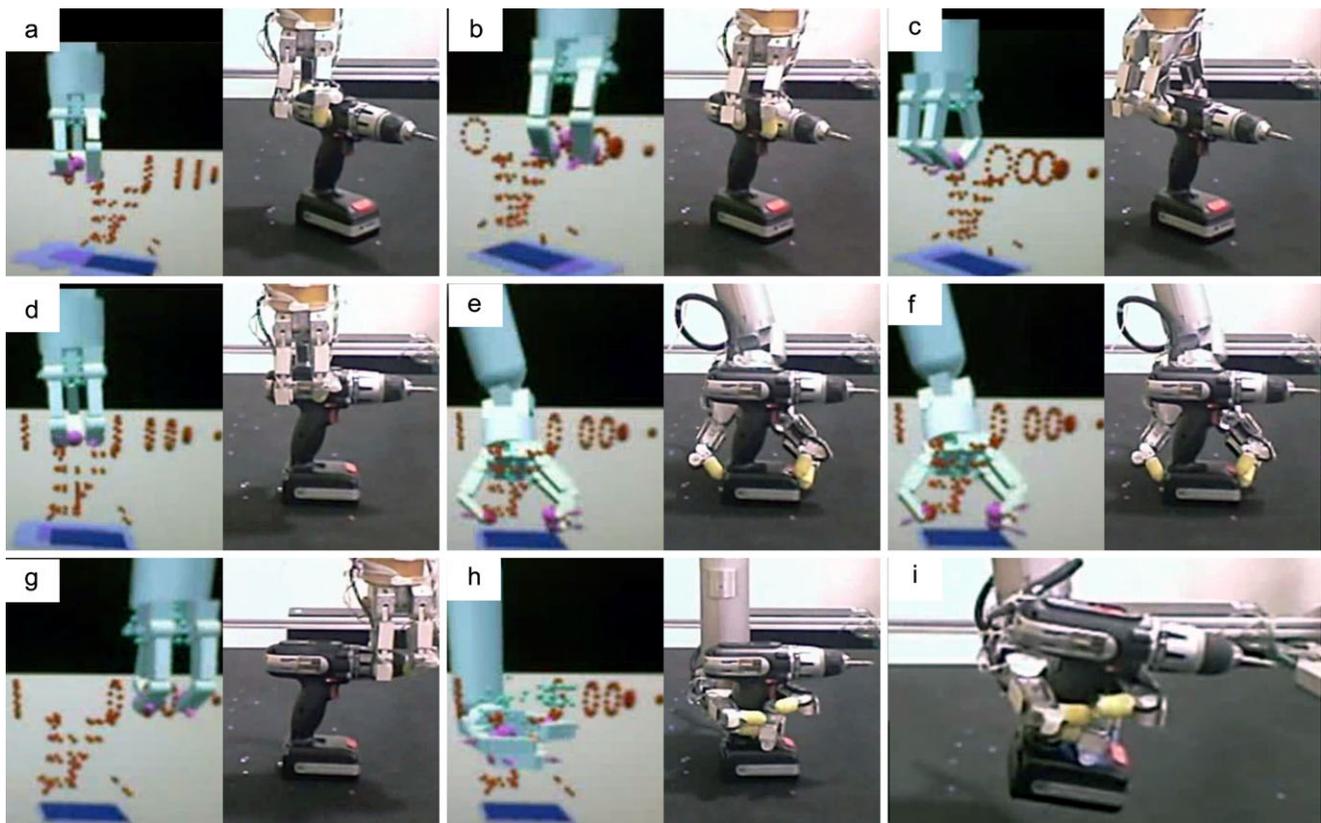


Fig. 10 (Color online) An example sequence of grasping the powerdrill: *the right panel in each pair shows the actual grasp that just occurred, while the left panel in each pair shows the robot's resulting belief state. In the left panel, the dark blue box on the table and the red points outlining the powerdrill show the current most likely state; the light blue boxes show states that are 1 standard deviation in each dimension away from the mean (shown in purple). The pink spheres and arrows show the observed hand contacts. In panels (a)–(c): info-grasp*

2 is used three times in a row, to narrow down the general position and orientation of the drill; panel (d) shows the result of reorienting the drill, which is necessary because the end grasp is infeasible before doing so; panels (e) and (f) show info-grasp 1 being used twice to narrow down the remaining major axis of uncertainty; panel (g) shows info-grasp 3 being used to precisely nail down the rotation of the drill; panel (h) shows the goal grasp being used on the drill; panel (i) shows the power drill having been lifted and the trigger pressed successfully

6 Conclusion

We can draw several conclusions from this work. First, just updating the belief state using observations and repeating the goal action until we are confident that success is likely is already sufficient to add a great deal of robustness to many situations. Second, even when information-gathering actions are necessary, a small search depth is effective in our framework; a depth of 1 is usually sufficient, and a depth greater than 2 is generally not useful. This means that action selection does not have to be very expensive. Third, we can choose effective actions despite our aggressive observation clustering, designed to bring the observation branching factor to a manageable level. Fourth, the quality of the observation and transition models limit the effectiveness of our system; this presents substantial opportunities for further research. In particular, a more predictive transition model that could more accurately estimate how objects move when we bump into them could further improve our results. It could

also enable us to add actions that purposely push objects in order to gain information, by, for instance, pushing them against walls. Another important area for further research is generalizing this work to more objects and more than three degrees of freedom per object. Promising approaches to explore include combinations of factoring, sampling, and adaptive-resolution grids.

We believe this work forms a step toward more general integration of tactile sensing and manipulation, ultimately supporting complex tasks such as multi-step assemblies.

Acknowledgement This research was supported by the National Science Foundation under Grant No. 0712012.

References

- Akella, S., & Mason, M. T. (1999). Using partial sensor information to orient parts. *The International Journal of Robotics Research*, 18(10), 963–997.

- Allen, P. K., & Bajcsy, R. (1985). Object recognition using vision and touch. In *Intl. joint conf. on artificial intelligence (IJCAI)*.
- Allen, P. K., & Michelman, P. (1990). Acquisition and interpretation of 3-D sensor data from touch. *IEEE Transactions on Robotics and Automation*, 6(4), 397–404.
- Alterovitz, R., Simeon, T., & Goldberg, K. (2007). The stochastic motion roadmap: a sampling framework for planning with markov motion uncertainty. In *Robotics science and systems (RSS)*.
- Baum, C. W., & Veeravalli, V. V. (1994). A sequential procedure for multihypothesis testing. *IEEE on Information Theory*, 40(6). doi:10.1109/18.340472.
- Burns, B., & Brock, O. (2007). Sampling-based motion planning with sensing uncertainty. In *IEEE intl. conf. on robotics and automation (ICRA)*.
- Cameron, A., & Durrant-Whyte, H. F. (1990). A Bayesian approach to optimal sensor placement. *The International Journal of Robotics Research*, 9(5), 70–88.
- Censi, A., Calisi, D., Luca, A. D., & Oriolo, G. (2008). A bayesian framework for optimal motion planning with uncertainty. In *IEEE intl. conf. on robotics and automation (ICRA)* (pp. 1798–1805).
- DeGroot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.
- Diankov, R., & Kuffner, J. (2008). Openrave: a planning architecture for autonomous robotics (Tech. Rep. CMU-RI-TR-08-34). Robotics Institute, CMU.
- Ellis, R. E. (1992). Planning tactile recognition paths in two and three dimensions. *The International Journal of Robotics Research*, 11(2), 87–111.
- Erdmann, M. (1998). Shape recovery from passive locally dense tactile data. In *Workshop on algorithmic foundations of robotics (WAFR)*.
- Erickson, L. H., Knuth, J., O’Kane, J. M., & Lavelle, S. M. (2008). Probabilistic localization with a blind robot. In *IEEE intl. conf. on robotics and automation (ICRA)*.
- Gadeyne, K., Lefebvre, T., & Bruyninckx, H. (2005). Bayesian hybrid model-state estimation applied to simultaneous contact formation recognition and geometrical parameter estimation. *The International Journal of Robotics Research*, 24, 615.
- Gaston, P. C., & Lozano-Perez, T. (1984). Tactile recognition and localization using object models: The case of polyhedra on a plane. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 257–265.
- Gonzalez, J. P., & Stentz, A. (2005). Planning with uncertainty in position an optimal and efficient planner. In *IEEE/RSJ intl. conf. on intelligent robots and systems (IROS)* (pp. 2435–2442).
- Hsiao, K. (2009). *Relatively robust grasping*. Ph.D. thesis, Massachusetts Institute of Technology.
- Hsiao, K., Kaelbling, L. P., & Lozano-Perez, T. (2007). Grasping POMDPs. In *IEEE intl. conf. on robotics and automation (ICRA)*.
- Hsiao, K., Kaelbling, L. P., & Lozano-Perez, T. (2008). Robust belief-based execution of manipulation programs. In *Workshop on algorithmic foundations of robotics (WAFR)*.
- Hsiao, K., Kaelbling, L. P., & Lozano-Perez, T. (2010). Task-driven tactile exploration. In *Robotics science and systems (RSS)*.
- Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101. doi:10.1016/S0004-3702(98)00023-X.
- Krause, A., & Guestrin, C. (2007). Near-optimal observation selection using submodular functions. In *Conf. on artificial intelligence (AAAI)*.
- Latombe, J. (1991). *Robot motion planning*. Dordrecht: Kluwer Academic.
- LaValle, S. M. (2006). *Planning algorithms*. Cambridge: Cambridge University Press.
- Lozano-Perez, T., Mason, M. T., & Taylor, R. H. (1984). Automatic synthesis of fine-motion strategies for robots. *International Journal of Robotics Research (IJRR)*, 3(1). doi:10.1177/027836498400300101.
- Melchior, N. A., & Simmons, R. (2007). Particle rrt for path planning with uncertainty. In *IEEE intl. conf. on robotics and automation (ICRA)*.
- Okada, T., & Tsuchiya, S. (1977). Object recognition by grasping. *Pattern Recognition*, 9(3), 111–119.
- Okamura, A. M., & Cutkosky, M. R. (2001). Feature detection for haptic exploration with robotic fingers. *The International Journal of Robotics Research*, 20(12), 925–938.
- Petrovskaya, A., & Ng, A. Y. (2007). Probabilistic mobile manipulation in dynamic environments, with application to opening doors. In *Intl. joint conf. on artificial intelligence (IJCAI)*.
- Prentice, S., & Roy, N. (2007). The belief roadmap: efficient planning in linear POMDPs by factoring the covariance. In *Intl. symposium on robotics research (ISRR)*.
- Skiena, S. (1989). Problems in geometric probing. *Algorithmica*, 4(4), 599–605.
- Taylor, R. H., Mason, M. T., & Goldberg, K. Y. (1988). Sensor-based manipulation planning as a game with nature. In *Intl. symposium on robotics research (ISRR)*.
- Wald, A. (1945). Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2). doi:10.1214/aoms/1177731118.



Kaijen Hsiao received her B.S.E. degree in Mechanical Engineering in 2002 from Princeton University and her Ph.D. degree in Computer Science in 2009 from MIT. She is now a research scientist at Willow Garage, where her research interests lie in grasping and manipulation.



Leslie Pack Kaelbling is Professor of Computer Science and Engineering and Research Director of the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology. She has previously held positions at Brown University, the Artificial Intelligence Center of SRI International, and at Teles Research. She received an A.B. in Philosophy in 1983 and a Ph.D. in Computer Science in 1990, both from Stanford University. Prof. Kaelbling has done substantial

research on designing situated agents, mobile robotics, reinforcement learning, and decision-theoretic planning. In 2000, she founded the Journal of Machine Learning Research, a high-quality journal that is both freely available electronically as well as published in archival form; she currently serves as editor-in-chief. Prof. Kaelbling is an NSF Presidential Faculty Fellow, a former member of the AAAI Executive Council, the 1997 recipient of the IJCAI Computers and Thought Award, a trustee of IJCAI and a fellow of the AAAI.



Tomás Lozano-Pérez is the School of Engineering Professor of Teaching Excellence at MIT, where he is a member of the Computer Science and Artificial Intelligence Laboratory. Professor Lozano-Pérez has all his degrees (S.B.'73, S.M.'76, Ph.D.'80) from MIT in Computer Science. He has been Associate Director of the Artificial Intelligence Laboratory and Associate Head for Computer Science of MIT's Department of Electrical Engineering and Computer Science. Professor Lozano-Pérez's research has been in

robotics (configuration-space approach to motion planning), computer vision (interpretation-tree approach to object recognition), machine learning (multiple-instance learning), medical imaging (computer-assisted surgery) and computational chemistry (drug activity prediction and protein structure determination from NMR & X-ray data).