Task-Driven Tactile Exploration

Task-Driven Tactile Exploration

Kaijen Hsiao
Willow Garage
Menlo Park, CA
Email: hsiao@willowgarage.com

Leslie Pack Kaelbling
Tomás Lozano-Pérez
MIT CSAIL
Cambridge, MA 02139
Email: lpk@mit.edu, tlp@mit.edu

Abstract— **This paper presents a decision-theoretic approach to problems that require accurate placement of a robot relative to an object in the world, including grasping and insertion. The decision process is applied to a robot hand with tactile sensors, to localize the object and ultimately achieve a target placement by selecting among grasping and information-gathering trajectories. The process is demonstrated in simulation and on a real robot.**

I. 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.

Vision and range sensors can estimate the pose of an object, but there is still residual uncertainty. Tactile sensing, combined with proprioception, can give highly reliable information about object position. It is expensive to map out an entire object with tactile sensing, so we 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 terminate and execute a final goal-achieving action, 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.

We consider several approximate decision procedures, based on *world relative trajectories* (WRTs) [11], which are robot arm trajectories parameterized with respect to a pose of an object in the world. 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. 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.

II. Related work

Using tactile sensing to recognize and/or locate objects has a long history [16, 3, 7, 1], 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 [2, 17]; the other uses sparse data directly via "contact probes" [9, 19]. Within the probe paradigm, there has been substantial work on "active" probing, choosing motions that best disambiguate among possible objects or poses [6, 19]. However, these probing motions have not typically been integrated into the goals of an overall manipulation task.

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, [15, 20, 13]) and in probabilistic formulations (for example, [14]). Several previous approaches have used probabilistic state estimation to represent uncertainty and integrate observational information in manipulation problems [18, 8]. Hsiao et al. [10] frame the decision-making problem as a partially observable Markov decision process and solve it near-optimally, but can only address small problems. Cameron et

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

al. [5] take a hypothesis-testing approach, applied to simple probes of a two-dimensional object.

This work builds on that of Hsiao et al. [11], who define WRTs and use them to limit the branching factor in lookahead search. That work showed, in simulation, that one-step lookahead, with respect to entropy reduction of the belief state, is an effective strategy for choosing grasping motions under uncertainty. However, reducing entropy in the belief state ignores the task requirements; we replace this criterion with estimated probability of task success. We also characterize the information conditions under which one can expect this type of strategy to produce correct results. We demonstrate the effectiveness of our methods both through simulation results and experiments with a real robot.

III. Action Selection

For concreteness, the rest of the paper discusses the problem of grasping an object using a pre-specified grasp (as shown in Figure 1), when there is uncertainty about the position of the object with respect to the robot, but it is important to keep in mind that the basic formulation is more broadly applicable.

A. Problem formulation

A grasp specification, G, consists of a set of relative poses for the hand and object, any of which constitutes 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. Thus, 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; the space of possible actions is enormous, if viewed as a space of trajectories or velocity commands. We use an action space that consists of a small set of WRTs, each of which is a trajectory for the robot arm described relative to the pose of the object, typically ending with closing the fingers. Together, an object pose $w \in W$ and a WRT τ determine a robot trajectory $\tau(w)$. An action (a) consists of moving the arm to a common starting configuration, and executing $\tau(w)$ until it completes without making a contact or until a contact is detected, at which point the arm stops moving and each finger closes until it makes contact or is fully closed. Note that every action starts from the same starting configuration. A single observation (o) is composed of: joint-angle sensing, a 6-axis force-torque sensor on each of three fingers, and binary contact sensors on the palm and hand.

B. State estimation

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

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

The state update depends on the observation model, $\Pr(o|s', a)$, and the state transition model, $\Pr(s'|s, a)$.

C. 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 could assign a cost to taking actions, and then seek a policy 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 [12], and can be computationally very difficult to solve in the general case.

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

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, the decision procedure is to: 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. 2) Obtain observation o_{t+1} and update the belief state. 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 section III-D, 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.

2) Multiple WRT: 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 will be 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 such cases, the robot must grasp the object using a different grasp and re-orient it in such a way that the that goal WRT is executable.

Given a set $\mathcal{T}$ of possible WRTs, the decision procedure is based on *finite-lookahead search*:

1) We construct a search tree, which is much like a game-search tree, except that instead of alternating between action choices of two players, the tree's layers alternate between action choices of the robot and stochastic outcome "choices" of the environment. Such a tree is shown in figure 2. The root node is the current belief state, b. It branches on the choice of actions a, and then, for each action, considers each possible observation o. The node reached from b via action a and observation o is a new belief state $b' = \mathrm{SE}(b, a, o)$. This process can be carried out to any depth; the figure shows part of a depth-2 tree.

For computational simplicity, the only actions that will be considered in belief state b are $\tau_j(\hat{w}(b))$, for each $\tau_j \in \mathcal{T}$. That is, we consider each possible WRT only with respect to the world state that is the most likely in b. Our observation space is actually continuous; section IV-A.2 discusses how we discretize it into a small number of canonical observations.

The tree has belief states at its leaves, which are assigned risk values $\rho_0(b) = \rho(\tau^*(\hat{w}(b)), b)$. The tree is then evaluated by backward induction: at observation nodes, the risk is computed based on the observation probabilities, and at action nodes, the minimum of the risks of the children is computed. Letting $\rho_k(b)$ be the value of being in belief state b with k action steps to go, and abbreviating $\tau_j(\hat{w}(b))$ as $\tau_j(b)$,

$$\rho_k(b) = \min_j \sum_o \Pr(o \mid b, \tau_j(b)) \rho_{k-1}(\mathrm{SE}(b, o, \tau_j(b))) \ .$$

2) Select the WRT τ_j that minimizes risk at the root node, and execute it with respect to $\hat{w}(b)$.

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

4) Terminate when a criterion on b is met, grasp the object using $\tau^*(\hat{w}(b))$, 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 $k = 1$, 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 section V, looking deeper can cause the termination criterion to be reached faster.

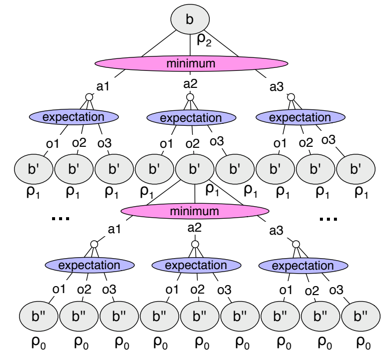

Fig. 2. Parts of a depth-2 search tree and associated risk computation.

D. 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.

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 [21] 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 [4]. These results hold whether the observation space is discrete or continuous. 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 $a = \tau^*(w_l)$ when the true world state is w_k to be

$$f_k^l(o) = \Pr(O = o \mid W = w_k, A = \tau^*(w_l))$$

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 not spatially overlapping with either w_i or 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 $j \neq k$:

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

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 (1) are less than 1, we can apply the MSPRT theorem as follows: In any sequence $a = a_0, \ldots, a_N$ where $n_j(a)$ of the trials in a are of action $\tau^*(w_j)$, then the probability that the decision procedure will not have terminated is a constant factor times $\max_j \left(I_{jk}^j\right)^{n_j(a)} \left(I_{jk}^k\right)^{n_k(a)}$. 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.

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 1 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 $j \in W \backslash W_G$ 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) 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.

4) Multiple WRTs: We must increase the set of WRTs when the requirements from (1) are not satisfied for τ^*. Given a set of WRTs $\mathcal{T}$, it must be that, for any pair of world states $j \in W \backslash W_G$ and $k \in W_G$ that 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.

IV. 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.

We used WRTs of three different types, for each target: 1) the goal WRT, τ^*, that grasps the object correctly if executed with respect to the correct world state, 2) 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 3) 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.

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 $\bar{w}$.

A. 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.

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

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

The size and complexity of the underlying state and observation spaces makes the modeling quite difficult. For tractability we make several assumptions: 1) There are no actual contacts that are not noticed or false triggering of the contact sensors. 2) The information gained from each aspect (arm, fingers) of the WRT is independent given the world state and action. 3) The swept path information is independent from the position and contact information. 4) 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 $\Pr(\phi, c, \psi \mid w, \tau_c)$ as

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

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:

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

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

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

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:

$$\begin{aligned} \Pr(\phi, c, \psi \mid w, \tau_c) &\approx \max_{\psi^*} \Pr(\psi^* \mid w, \tau_c) \Pr(\phi, c \mid \psi^*, w, \tau_c) \\ &\approx P_f(\lfloor \psi \rfloor, w) \max_{\phi^*} \Pr(\phi, c \mid \phi^*, w) \end{aligned}$$

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

The fingertip sensors allow us 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

$$\begin{aligned} \Pr(\phi, c, \psi \mid w, \tau_c) &\approx P_f(\lfloor \psi \rfloor, w) \mathcal{G}(d_l(l_i(\phi), f^*); 0, \sigma_l^2) \cdot \\ &\quad \mathcal{G}(d_n(n_i(\phi), f^*); 0, \sigma_n^2) \end{aligned}$$

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. 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.

2) Reduced observation space: The space of possible observations is continuous and high-dimensional. 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.

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 take this discrete set of nominal observations (which in our implementation has about 24000 elements) as our starting set of possible observations.

Clustering: We start by clustering directly on contact observation vectors, using an agglomerative clustering method, with a distance metric that is a sum over the contacts of a weighted combination of the 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. 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 7.

B. Transition Model

The transition model specifies $\Pr(W_{t+1} = w_j \mid W_t = w_i, A_t = \tau_c)$. 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:

$$\begin{aligned}\Pr(w'_j \mid w_i, \tau_c) \quad \propto \quad & (1-p_b)\mathcal{G}(d_p(\bar{w_i}, \bar{w_j}); 0, \sigma_s^2) \\ & + p_b\mathcal{G}(d_p(bump(\bar{w_i}, \phi, c), \bar{w_j}); 0, \sigma_b^2) \ ,\end{aligned}$$

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.

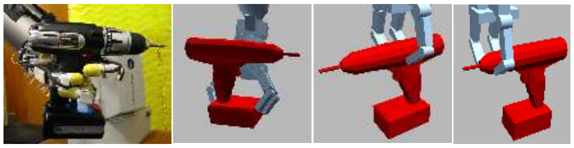

Fig. 3. Goal and information grasps for the power drill.

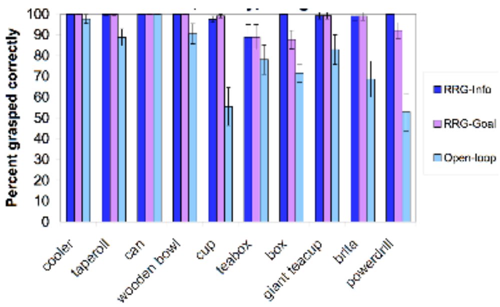

Fig. 4. Simulation results for all objects at low uncertainty.

V. Experiments

Our experiments used 10 different objects, shown with their goal grasps in Figures 1 and 3. 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 9^{th} action, if the goal criterion was not reached, the goal WRT was executed.

A. Simulation

In our simulation experiments, we compare three different strategies: 1) **Open-loop**: execute the goal WRT once and terminate; 2) **RRG-Goal**: execute the goal WRT repeatedly on the most likely state, terminating when the risk threshold is met; and 3) **RRG-Info**: choose among goal, re-orientation, and information WRTs, with a depth D lookahead-search decision procedure, terminating when the risk threshold is met. Each simulation experiments was carried out with at least 100 trials.

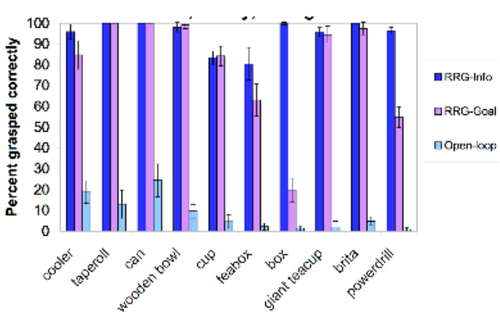

Fig. 5. Simulation results for all objects at high uncertainty.

Figure 4 shows the results for experiments carried out in simulation with initial belief state a discretized Gaussian with standard deviations of 1 cm in x, 1 cm in y, and 3 degrees in θ, and lookahead-search depth $D = 2$. 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 5 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 $D = 2$. At this level of uncertainty, ∼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, as discussed earlier, 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. We expect that using an adaptive grid size would improve the performance.

Figure 6 shows the percentage of successful grasps (with 90% confidence bounds) 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 [11]) 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

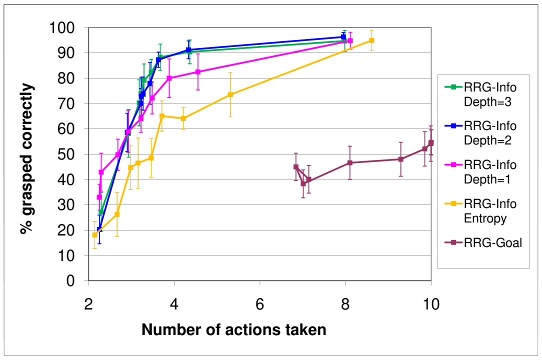

Fig. 6. Varying termination criterion trades off between grasp correctness and number of actions required; shown for power drill.

of success later on. This is due to the fact that, although information WRT 1 (shown in Figure 3) 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.

B. 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. 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 simulations). One failure with the giant tea cup was due to a flaw in our algorithm for computing collision depth for objects with thin features (like the tea cup handle).

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

VI. Conclusion

We can draw several conclusions from this work. First, 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. Second, we can choose effective actions despite our aggressive observation clustering, designed to bring the observation branching factor to a manageable level. Third, 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.

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

[1] S. Akella and M. T. Mason. Using partial sensor information to orient parts. *IJRR*, 18(10):963–997, 1999.
[2] P. K. Allen and R. Bajcsy. Object recognition using vision and touch. In *IJCAI*, 1985.
[3] P. K. Allen and P. Michelman. Acquisition and interpretation of 3-D sensor data from touch. *IEEE Trans. on RA*, 6(4):397–404, 1990.
[4] C. W. Baum and V. V. Veeravalli. A sequential procedure for multihypothesis testing. *IEEE Trans. on Information Theory*, 40(6), 1994.
[5] A. Cameron and H. F. Durrant-Whyte. A Bayesian approach to optimal sensor placement. *IJRR*, 9(5):70–88, 1990.
[6] R. E. Ellis. Planning tactile recognition paths in two and three dimensions. *IJRR*, 11(2):87–111, 1992.
[7] M. Erdmann. Shape recovery from passive locally dense tactile data. In *WAFR*, 1998.
[8] K. Gadeyne, T. Lefebvre, and H. Bruyninckx. Bayesian hybrid model-state estimation applied to simultaneous contact formation recognition and geometrical parameter estimation. *IJRR*, 24:615, 2005.
[9] P. C. Gaston and T. Lozano-Perez. Tactile recognition and localization using object models: The case of polyhedra on a plane. *IEEE Trans. on PAMI*, 6:257–265, 1984.
[10] K. Hsiao, L. P. Kaelbling, and T. Lozano-Perez. Grasping POMDPs. *ICRA*, 2007.
[11] K. Hsiao, L. P. Kaelbling, and T. Lozano-Perez. Robust belief-based execution of manipulation programs. In *WAFR*, 2008.
[12] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Art. Intell.*, 101, 1998.
[13] J.C. Latombe. *Robot Motion Planning*. Kluwer Academic, 1991.
[14] Steven M. LaValle. *Planning Algorithms*. Cambridge U. Press, 2006.
[15] T. Lozano-Perez, M. T. Mason, and R. H. Taylor. Automatic synthesis of fine-motion strategies for robots. *IJRR*, 3(1), 1984.
[16] T. Okada and S. Tsuchiya. Object recognition by grasping. *Pattern Recognition*, 9(3):111–119, 1977.
[17] A. M. Okamura and M. R. Cutkosky. Feature detection for haptic exploration with robotic fingers. *IJRR*, 20(12):925–938, 2001.
[18] A. Petrovskaya and A. Y. Ng. Probabilistic mobile manipulation in dynamic environments, with application to opening doors. *IJCAI*, 2007.
[19] Steven Skiena. Problems in geometric probing. *Algorithmica*, 4(4):599–605, 1989.
[20] R. H. Taylor, M. T. Mason, and K. Y. Goldberg. Sensor-based manipulation planning as a game with nature. In *ISRR*, 1988.
[21] A. Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2), 1945.