State Estimation in Contact-Rich Manipulation

Florian Wirnshofer¹, Philipp S. Schmitt¹, Philine Meister¹, Georg v. Wichert¹ and Wolfram Burgard²

Abstract—This paper introduces a Bayesian state estimator for contact-rich manipulation tasks with application in non-prehensile manipulation, industrial assembly or in-hand localization. The core idea of our approach is to explicitly model both the contact dynamics and a torque-based robot controller as part of the underlying system model. Our approach is capable of estimating the state of movable objects for various robot kinematics and geometries of robots and objects. This includes complex scenarios with multiple robots, multiple objects and articulated objects. We have validated our approach in simulation and on a physical robot. The experiments show that multi-modal distributions of six degrees of freedom object poses can be accurately tracked in real-time in a complex manipulation scenario.

I. INTRODUCTION

Various automation tasks require robots that interact with objects in their environment. Tasks range from material handling and logistics to potential household applications in the future. In unstructured environments, robots must perceive the poses of manipulated objects. Perceiving objects with cameras is challenging as occlusions occur between robot and objects and among different objects. These occlusions happen at times where an accurate perception is most important: during interactions with objects, such as grasping or assembly. For this reason, it is important to estimate the state of manipulated objects based on sensor-feedback available in contact.

A natural approach to estimation is to employ Bayesian reasoning. However, several aspects of contact-rich manipulation render this a challenging problem:

1) Variety of manipulation tasks: Various robot geometries and kinematics exist, such as multi-robot systems or dexterous multi-fingered grippers. The same applies to geometry and number of objects.

2) Complex, dynamic interactions in contact: Predicting the interactions between multiple robots and objects, such as shown in Fig. 1, requires reasoning about a dynamical evolution of object states. To do so, the velocity of objects and forces stemming from gravity, contact and friction must be considered.

3) High-dimensional distributions: Estimating the state of motion of manipulated objects leads to inherently high-dimensional distributions. Each object has at least 12 degrees-of-freedom (DoF), six for its pose and six for its velocity.

To the best of our knowledge, there currently exists no approach that addresses all three of these aspects. The contribution of this paper is to propose a new probabilistic model for state estimation in contact-rich manipulation. This model comprises the physical dynamics of contact as well as an explicit model of a compliant, trajectory-following controller for the robot. We propose a particle filter-based approach to sequentially estimate the probability distribution of object states. The combination of particle representation together with an integrated model of a compliant controller and contact dynamics makes it possible to estimate a multi-modal distribution over object poses in real-time. We experimentally validated our approach for a dual-robot scenario and an in-hand localization scenario. Note that we assume the object geometries given. We do not use any feedback other than joint positions.

II. RELATED WORK

Tracking object poses during manipulation constitutes a demanding yet relevant challenge that has attracted substantial research interest throughout the past decades. Jia et al. [1] achieve object tracking during planar grasping using state observers known from deterministic control theory. Our work covers non-prehensile manipulation tasks, where large and non-linear uncertainties favor probabilistic approaches.

Bayesian state estimation provides an alternative to classical observer-based methods. In Bayesian estimation, the knowledge about a state is represented as a probability density conditioned on the available data [2]. Since the exact computation of the belief is intractable, assumptions must be made with regard to the representation of the involved probability distributions. Lowrey et al. [3] and Pfanne et al. [4] make use of Gaussian belief approximations in order...
to achieve real-time capable state estimation for general, constrained physical systems and in-hand object localization. We consider contact dynamics and non-prehensile manipulation, which tend to produce multi-modal distributions when the variance of the initial belief is sufficiently large. In these cases, uni-modal Gaussian approximations may diverge from the actual object poses.

Non-parametric belief-approximations, such as particle filters, have been employed in the context of manipulation, as they are able to approximate multi-modal distributions. The approaches mainly differ with regard to two aspects: how probabilistic object motion is modeled and how measurements are used to improve the estimate.

The works of Thomas et al. [5] and Petrovskaya et al. [6] model object motion by means of a random diffusion process, where hypotheses over object poses are drawn randomly within the vicinity of a prior estimate. A random selection of object poses can be used for static objects that have no physically plausible motion model, but suffers from two distinct disadvantages. Firstly, random sampling yields a substantial amount of physically implausible candidates, e.g. situations in which the objects float without contact or penetrate each other or the environment geometry. To provide a remedy for generating a large amount of implausible hypotheses, Nottensteiner et al. [7] introduce heuristically guided sampling strategies. Applications of [7] have demonstrated reliable object localization in industrial assembly scenarios [8, 9]. Secondly, these methods do not account for motion caused by gravity, interactions of multiple objects or between robot and objects. In contrast, we make use of a physically consistent motion model, taking into account constraint forces from contact, friction, damping and actuator limits. Vondrak et al. [10] introduce a physically consistent, contact-aware motion model for vision-based tracking of a humanoid robot. The works of Duff et al. [11] and Grundmann et al. [12] utilize rigid-body simulators in order to improve estimates during vision-based object tracking. These works rely on the availability of visual feedback and do not account for robot-object interaction. Similarly, Zhang et al. [13] utilize a physically consistent motion model in combination with observations from tactile contact and visual sensors in a planar manipulation scenario. The authors propose a more general approach that further allows for the estimation of model parameters [14]. Unfortunately, discrete measurements with low noise, as obtained from contact sensors, have been found to quickly induce particle depletion [2, 15]. The manifold particle filter (Koval et al. [15]) circumvents the particle depletion issue, yet is based on a quasi-static, planar physics model. By contrast, we do not employ projections to lower-dimensional manifolds and allow for object motion, even when there is no direct interaction between object and manipulator.

III. PROBABILISTIC MODEL FOR COMPLIANT MOTION IN CONTACT

The core contribution of this paper is a probabilistic motion model for compliant interaction. We model robots and objects as a dynamical system of rigid bodies with contact forces and friction (Section III-A). Instead of considering robot torques as inputs to this system, we consider a compliant torque-based controller as part of our probabilistic model (Section III-B). Based on these two components we formulate the estimation problem in Section III-C.

A. Rigid Body Motion and Contact Dynamics

We model our system as a set of articulated, rigid bodies with contact dynamics. We denote the configuration of our system as \( q = \begin{bmatrix} q_r, q_o \end{bmatrix} \) that is composed of the robot configuration \( q_r \) and the configuration of the objects \( q_o \). In the case of the dual arm manipulator of Fig. 2, \( q_r \) comprises 14 axis-positions. The dimensionality of \( q_o \) depends on the number and structure of objects. For the case of the single, freely moving rigid object in Fig. 2, the object configuration denotes an element from the special Euclidean group \( q_o \in \text{SE}(3) \). Note that \( q_o \) may also represent multiple or articulated objects. A system configuration \( q \) evolves according to the system dynamics:

\[
M(q) \ddot{q} + c(q, \dot{q}) + g(q) = \tau_u + \tau_c,
\]

with

\[
M = \text{diag} \{ M_r, M_o \}, \quad g = \begin{bmatrix} g_r, g_o \end{bmatrix} \in \mathbb{R}^{n_r+n_o}, \quad c = \begin{bmatrix} c_r, c_o \end{bmatrix} \in \mathbb{R}^{n_r+n_o}, \quad \tau_u = \begin{bmatrix} \tau_r, \tau_o \end{bmatrix} \in \mathbb{R}^{n_r+n_o},
\]

We denote the dimensionality of robot- and object-configurations as \( n_r \) and \( n_o \). Vectors \( g \) and \( c \) denote the composite gravitational and Coriolis terms and \( M \) is the block-diagonal inertia matrix. The constraint torques \( \tau_c \) summarize effects from external perturbations, most importantly contact-forces. We denote the generalized input forces as \( \tau_u = \begin{bmatrix} \tau_r, \tau_o \end{bmatrix} \). In the dual arm example of Fig 2, \( \tau_r \) comprises the 14 axis torques generated by the respective motors. Solving (1) for \( \ddot{q} \) yields the forward dynamics \( f(\cdot) \):

\[
\ddot{q} = f(q, \dot{q}, \tau_u).
\]
The proximal control input to our model is therefore the desired configuration \( \mathbf{q}_d \). A compliant controller \( k(\cdot) \) computes the robot input forces \( \mathbf{\tau}_r \). The forward dynamics \( f(\cdot) \) compute the acceleration of the entire system.

**B. Compliant Interaction and Controller Model**

We aim to gather information through contact interaction with unstructured or partially known environments. Doing so, the path a manipulator follows may be constrained by the environment geometry. When stiff, position-controlled manipulators are employed, the resulting contact forces lie well beyond the physical capabilities of either the robot or its surroundings. This can be addressed by using a compliant, trajectory-following controller. Based on the deviation from the desired trajectory point \( \mathbf{q}_d \) and its current velocity \( \mathbf{q}_r \), the robot is controlled via

\[
\mathbf{\tau}_r = k \left( \mathbf{q}_r, \dot{\mathbf{q}}_r, \mathbf{q}_d, \dot{\mathbf{q}}_d \right).
\]

Here, a stable, compliant controller \( k \) computes the generalized input forces of the robot \( \mathbf{\tau}_r \). Examples for such a compliant controller include joint-level impedance control or Cartesian impedance control. Refer to Fig. 3 for a block diagram of the complete system. The proximal control input to our system is therefore the desired configuration \( \mathbf{q}_d \). This desired configuration may evolve in an arbitrary, twice differentiable trajectory which we assume as given. Motion generation is not within the scope of this paper.

Using a compliant controller as part of our system model has two advantages. The first advantage is that the controller stabilizes the robot configuration \( \mathbf{q}_o \), which prevents implausible drift due to the double integrator dynamics of (1). A second advantage is that (1) and (3) may be integrated efficiently in a physics simulator. This decouples the operating frequency of the actual torque-based robot controller from that of the state estimator. To estimate object poses in real-time, we may use the frequency at which \( \mathbf{q}_d \) is computed, which typically can be lower by orders of magnitude than the internal control frequency of the physical robot.

**C. Sequential Bayesian Estimation**

With models of system dynamics (1) and controller (3) in place, we can define the estimation problem. Note that for the remainder of this work we will annotate the discrete-time equivalent of all continuous variables with subscript \( (\cdot)_t \). Let therefore state \( \mathbf{x}_t \) be the time-discrete equivalent of all system configurations and velocities \( \mathbf{x}_t = (\mathbf{q}_t, \dot{\mathbf{q}}_t) \). During contact interaction, the belief over the system state \( \text{bel}(\mathbf{x}_t) \) evolves according to the motion model

\[
p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t),
\]

where \( \mathbf{x}_{t-1} \) is the state at time \( t - 1 \) and \( \mathbf{u}_t \) is the system input at time \( t \). The input \( \mathbf{u}_t \) encodes the evolution of \( \mathbf{q}_d \) between \( t - 1 \) and \( t \).

This motion model is computed by integrating (1) and (3) and, without modification, would yield a deterministic evolution of the state. However, we consider two sources of uncertainty in our dynamics model similar to prior work [14, 16]: external forces acting on the objects and parametric modeling errors of the objects. External forces on the objects are modeled as generalized object input force \( \mathbf{\tau}_o \). The application of a random force offers a significant advantage compared to a direct modification of \( \mathbf{q}_o \), namely that the physically consistent forward simulation of an applied wrench will in turn yield a physically plausible object configuration \( \mathbf{q}_o \). To incorporate modeling errors, we assume that the dynamic parameters of the objects such as mass, inertia and friction follow a suitable random process.

Using this motion model yields the predicted belief:

\[
\tilde{\text{bel}}(\mathbf{x}_t) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1}, \mathbf{u}_t) \text{bel}(\mathbf{x}_{t-1}) \, d\mathbf{x}_{t-1}.
\]

Once an observation \( z_t \) becomes available, we update our belief, following the Bayesian laws of probability as

\[
\text{bel}(\mathbf{x}_t) = \eta p(z_t | \mathbf{x}_t) \tilde{\text{bel}}(\mathbf{x}_t),
\]

with \( \eta \) being a normalization constant and \( p(z_t | \mathbf{x}_t) \) a suitable measurement model. In the example of the dual-arm manipulator of Fig. 2, \( z_t \) comprises 14 measured axis positions.

**IV. PARTICLE FILTER FOR CONTACT-RICH MANIPULATION**

In this section, we outline our concrete realization of the Bayes-filter of the previous section. As (4) and (5) are computationally intractable, we approximate the exact computation using a particle filter [17, 18], briefly explained in Section IV-A. We review the filter’s key components, namely the motion model and the measurement model, in Section IV-B and Section IV-C, respectively.
A. Particle Filter for Contact Manipulation

The general structure of a particle filter is outlined in algorithm Particle Filter. The non-parametric filter represents the belief over a system state \( \mathbf{bel}(\mathbf{x}_t) \) by means of a set \( X_t \) of \( N \) weighted particles \( < [i]\mathbf{x}_t, [i]w_t >, i = 1 \ldots N \). Procedure SAMPLE MOTIONMODEL (line 3) propagates each particle according to the system model (1) controlled by the compliant controller (3). Based on the measurement \( z_t \), procedure MEASUREMENTMODEL (line 4) assigns an importance weight to each of the particles. We assume that our observation \( z_t \) corresponds to some noisy measurement of the robot configuration \( q_{m,t}^{\text{nom}} \). The particles are re-sampled with regard to the importance weights (line 7), whenever the effective sample size \( N_{\text{eff}} = \sum_{i=1}^N [i]w_t^2 \) falls below a \( \gamma \)-fraction of the number of particles \( N \) [19].

B. Motion Model

Procedure SAMPLE MOTIONMODEL \([i]x_{t-1}, u_t \) computes a random sample \( [i]x_t \sim p(x_t \mid x_{t-1}, u_t) \). Via forward dynamics (2) the acceleration of a particle can be computed:

\[
[i] \ddot{q}_t = f \left( [i]q_t - 1, [i]\dot{q}_t - 1, [i]u_t \right),
\]

(6)

The generalized input forces \( [i]\tau_{u,t} = [i]\tau_{r,t} \) and objects \( [i]\tau_{o,t} \). According to (3), the robot input forces are computed by a compliant controller \( k(\cdot) \) based on a desired configuration \( q^{\text{nom}} \). We implement \( k(\cdot) \) as a joint-level impedance controller with gravity compensation

\[
[i]\tau_{r,t} = K \left( [i]q_t - [i]q^0 \right) + D \left( [i]\dot{q}_t - [i]\dot{q}^0 \right) + g_{r,t} \left( [i]q_t \right),
\]

where \( K \in \mathbb{R}^{n_r \times n_r} \) is a diagonal stiffness matrix and \( D \in \mathbb{R}^{n_r \times n_r} \) is a diagonal damping matrix. Our choice of the controller \( k(\cdot) \) compensates the robot dynamics and imposes a joint-wise spring damper system with known stiffness and damping. This decouples the physical realism of our simulation from the accuracy of the calibrated model-parameters (mass, inertia, friction and gear-ratios). In practice, obtaining precise model parameters requires significant effort. With our approach, the robot model may be a black box, as long as the imposed external behavior (spring-damper for impedance control) is known and provided by the manufacturer.

Note that each particle evolves according to an individual value of \( [i]q_t \) and \( [i]\dot{q}_t \) and individual \( [i]\tau_{r,t} \) and \( [i]\tau_{o,t} \). However, the motion model input \( u_t \) that encodes the evolution of \( \dot{q}_r,t, \dot{q}_{o,t} \) is shared across all particles.

Given the particle’s predicted acceleration \( [i]\ddot{q}_t \), we can evaluate the velocities and positions of \( [i]x_t \) by means of semi-implicit Euler integration over the step size \( \Delta T \).

For an undisturbed and perfectly modeled system, \( \tau_{o,t} \) is equal to zero and SAMPLE MOTIONMODEL is deterministic. As described in Section III-A, we consider two sources of motion noise: external forces on objects and noisy model parameters.

The generalized input forces on objects are computed via

\[
\tau_{o,t} = J_{CoM} h_t,
\]

where \( h_t \in \mathbb{R}^6 \) is a random wrench, applied at the center of mass (CoM) of each of the objects. The Jacobian \( J_{CoM} \in \mathbb{R}^{6 \times n_o} \) maps the input wrench \( h_t \) to the generalized force coordinates of \( \tau_{o,t} \). We generate the random wrench \( h_t \) by means of an Ornstein-Uhlenbeck stochastic process [20] with a stationary distribution \( h_t \sim \mathcal{N} \left( 0^6; \text{diag} \left\{ \sigma_F^2 1^3, \sigma_T^2 1^3 \right\} \right) \). Here, \( \sigma_F \) and \( \sigma_T \) are the force and torque noise deviations, respectively.

Additionally, we add noise to the parameter set of our dynamics model (1), which can also be regarded as a way of addressing plant-model mismatch through the motion model. In this context, let \( m_{\text{nom}} \) denote the nominal mass of one of the objects composing \( q_o \). We modify the object masses after each simulation period \( \Delta T \) by adding an offset \( \min \{ -0.9 m_{\text{nom}}, \hat{m} \} \), where \( \hat{m} \sim \mathcal{N} \left( 0, (0.1 m_{\text{nom}})^2 \right) \).
The truncation of \( \tilde{\eta} \) is required since a negative mass is physically implausible.

Fig. 5 shows the behavior of our motion model when no interaction between robot and objects is present. Starting from an arbitrary initial distribution, the contact dynamics lead to physically plausible hypotheses about object configurations, where the object is resting on one of its sides. It can be seen that the resulting distribution forms several manifolds in configuration space in which probability densities are peaked. These manifolds are three-dimensional and not 12-dimensional as the full configuration space of the object.

**C. Measurement Model**

For our measurement model, we use measurements \( q_i^{xyz} \) of the robot configurations as measurement input \( z_i \). For the dual-arm robot in Fig. 2, this measurement comprises 14 axis-positions. We assume that these measurements follow a multivariate normal distribution centered at the true robot configuration \( \tilde{q}_{r,t} \). This leads to the measurement model:

\[
[w_i]_t = p(\tilde{z}_t | [i]x_t)[w_i]_{t-1},
\]

with

\[
p(\tilde{z}_t | [i]x_t) = p(q_i^{xyz} | [i]q_{r,t}) = \mathcal{N}([i]q_{r,t}, \Sigma),
\]

where \( \Sigma = \text{diag} \{1^n/\sigma_q, 1^{18} \} \) denotes the variance of measurements of axis-positions. Using only position measurements and no torque measurements has a practical advantage for the implementation of our approach since obtaining plausible torque measurements from the simulated particles would again require precisely calibrated robot model-parameters.

Fig. 4 shows the operation of the particle filter. Step I shows the belief after all particles, sampled from the prior distribution, have come to rest within the simulation. The following steps show how simulation of contact dynamics and interaction with the impedance-controlled robots change the belief. Eventually the position of the object is accurately determined. Since the manipulated object has rotational symmetries, the distribution of rotations has eight peaks.

**V. Results**

This section introduces the set of benchmark problems to assess our approach (Section V-A) and gives implementation details (Section V-B). In Section V-C, we analyze results from simulated experiments. Finally, we discuss the results gathered from real-world experiments in Section V-D.

**A. Benchmark Scenarios**

We evaluated our approach on the six notably different benchmark scenarios depicted in Fig. 6. The benchmarks demonstrate our method’s applicability to scenarios with multiple robots and multiple or articulated objects. As torque-controlled robots we used either two 7-DoF arms or one 8-DoF robot hand with \( \sigma_\theta = 0.01 \text{ rad} \) (see (7)).

The CUBE** benchmarks show pose estimation during non-prehensile manipulation, where actions can drastically increase state uncertainty (e.g., when pushing an object off a box). The Labyrinth task consists of two spheres of equal size within a labyrinth. The position of each sphere has to be distinguished through their different inertial properties. Finally, we demonstrate our method’s applicability to in-hand localization of non-convex and articulated objects.

We constructed the initial belief \( x_0 \) as follows: the initial robot configuration \( q_{r,0} \) of each particle was set to match the initial configuration of the real manipulator. The initial object configuration \( q_{0,0} \) was sampled from a normal distribution, the only exception being object orientations, which were chosen uniformly at random on SO (3). Below the benchmark scenarios of Fig. 6, the dimensionality of the object configuration \( q_o \), the object mass \( m \) and the translational standard deviation \( \sigma_{xyz} \) are shown. For all benchmarks, we chose the standard deviation of the random forces \( \sigma_F \) that were applied in our motion model as one tenth of the gravitational forces \( \sigma_F = \frac{1}{10} m 9.81 m/s^2 \), with mass \( m \). We did not apply a random torque i.e. \( \sigma_T = 0 \text{ Nm} \).
TABLE I

<table>
<thead>
<tr>
<th></th>
<th>CUBE</th>
<th>CUBE 2</th>
<th>CUBE BOX</th>
<th>LARY- RINH</th>
<th>HAND STANFORD</th>
<th>HAND-HINGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>800</td>
<td>800</td>
<td>800</td>
<td>800</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>( t_{\text{motion}} )</td>
<td>3.21</td>
<td>3.38</td>
<td>4.70</td>
<td>4.81</td>
<td>4.14</td>
<td>3.13</td>
</tr>
<tr>
<td>( t_{\text{meas}} )</td>
<td>0.160</td>
<td>0.203</td>
<td>0.166</td>
<td>0.172</td>
<td>0.0964</td>
<td>0.151</td>
</tr>
<tr>
<td>( t_{\text{total}} )</td>
<td>3.42</td>
<td>3.63</td>
<td>4.92</td>
<td>5.04</td>
<td>4.29</td>
<td>3.32</td>
</tr>
</tbody>
</table>

B. Implementation

The simulation of the system dynamics with contact (1) was implemented via the Mujoco simulator [21] at an integration step-size of \( \Delta T = 5 \) ms. We computed the individual particle motions in parallel on two Intel Xeon E5-2640v4 (2.4 GHz) ten-core processors. The real-world experiments were carried out using two joint-impedance controlled KUKA iiwa 7 R800 7-axis manipulators. We chose a stiffness of \( K = \text{diag} \{114100\} \) Nm rad and computed the damping term \( D \) using double diagonalization damping design [22]. The command rate of the trajectory points \( q^*_C \) was 200 Hz. Table 1 lists the number of particles used, the average time to evaluate the motion and measurement model (\( t_{\text{motion}}, t_{\text{meas}} \)), as well as the average total cycle time \( t_{\text{total}} \). Except for the labyrinth scenario, we are able to update the filter in less than one actuation interval, making our filter real-time capable.

C. Results - Simulation

An important property of a particle filter is its ability to improve the estimate of the object configuration \( q_o \) relative to the prior distribution. We conducted 100 simulated experiments for each of the benchmarks, comparing the simulated ground truth to the filter’s estimate. As estimate of the filter we use the mean of the particles. The initial object configuration of the simulated ground truth was sampled from the filter’s prior belief. We measured the deviation from the simulated ground truth before the first and after the final re-sampling step.

To measure the deviation of the estimate from the ground truth, we use two distance metrics. Between two configuration components \( p_1, p_2 \in \mathbb{R}^n \) (e.g. 3D translation, 1-DoF joint position), distance is defined as

\[
d_{\text{lin}}(p_1, p_2) = \| p_2 - p_1 \|_2.
\]

Orientations are parametrized by unit quaternions \( q \). The distance between two quaternions \( q_1, q_2 \) is given by the SO(3) metric [23]

\[
d_{\text{rot}}(q_1, q_2) = \| \log(q_2^{-1} \ast q_1) \|_2,
\]

where * and \( \log \) denote the quaternion-product and quaternion-logarithm, respectively.

Fig. 7 depicts a boxplot over the obtained errors. The results demonstrate the filter’s ability to improve the estimate in most experiments across all benchmarks. We used a single, predefined manipulator trajectory \( q^*_C \) for each of the experiments. For some of the initial ground truth configurations, the trajectory resulted in ambiguous contact situations.

VI. CONCLUSION

This paper introduces a novel approach to state estimation during contact-rich manipulation tasks. Our method is capable of tracking non-convex and articulated objects. We consider the full, dynamic object state including velocities and do not rely on tailored reductions to lower-dimensional manifolds. The key idea of our approach is a motion model that incorporates both contact dynamics and a compliant robot controller. This model is then used within a particle filter. Our motion model focuses particles on physically plausible regions of the state-space. This enables robust tracking with low numbers of particles and thus on-line state estimation. Thorough simulative experiments based on several benchmark scenarios from different fields of application demonstrate the effectiveness of our approach. In a real-world, dual-robot experiment, our approach was shown to provide accurate estimates of object poses that enable peg-in-hole assembly.
REFERENCES


