

The Application of Particle Filtering to Grasping Acquisition with Visual Occlusion and Tactile Sensing

Li (Emma) Zhang and Jeffrey C. Trinkle

Abstract—Advanced grasp control algorithms could benefit greatly from accurate tracking of the object as well as an accurate all-around knowledge of the system when the robot attempts a grasp. This motivates our study of the $G-SL(AM)^2$ problem, in which two goals are simultaneously pursued: object tracking relative to the hand and estimation of parameters of the dynamic model. We view the $G-SL(AM)^2$ problem as a filtering problem. Because of stick-slip friction and collisions between the object and hand, suitable dynamic models exhibit strong nonlinearities and jump discontinuities. This fact makes Kalman filters (which assume linearity) and extended Kalman filters (which assume differentiability) inapplicable, and leads us to develop a particle filter. An important practical problem that arises during grasping is occlusion of the view of the object by the robot’s hand. To combat the resulting loss of visual tracking fidelity, we designed a particle filter that incorporates tactile sensor data. The filter is evaluated off-line with data gathered in advance from grasp acquisition experiments conducted with a planar test rig. The results show that our particle filter performs quite well, especially during periods of visual occlusion, in which it is much better than the same filter without tactile data.

I. INTRODUCTION

Why is autonomous grasping and manipulation in unstructured environments still so hard for robots after 30+ years of research? For a robot to perform skilled grasping and manipulation, it has to have information that is hard to obtain with sufficient accuracy: the geometry of the object, estimates of physical quantities (such as weight and friction), and the positions of the object and contacts.

The best scenario for a robot is that it can get an accurate physical model from a database, has a vision system that can track the object, and tactile sensors that can aid tracking when the fingers occlude the visual tracking features. Even in this case, the localization errors of the perception system and positioning errors allowed by the control system may not be small enough to ignore during the process of grasp acquisition or subsequent manipulation. These errors can cause the hand to bump the object accidentally when reaching or curling the fingers, possibly causing the object to slip and tumble out of the grasp. The problems are magnified when the physical quantities are known only roughly, or vary over space and time as friction parameters are known to do.

This paper is motivated by the idea that one can dramatically advance the state of the art in grasping and manipu-

lation by designing algorithms that can estimate a physical model¹ of a grasping system (composed of hand, object, and environment), while accurately tracking the object. We will refer to this problem as the $G-SL(AM)^2$ problem.

The $G-SL(AM)^2$ problem is to autonomous robotic grasping, what the SLAM problem is to autonomous robotic mobility. The G stands for Grasping. $SL(AM)^2$ stands for: Simultaneous Localization, and Modeling, and Manipulation. The word “Modeling” implies that the robot will use its sensor systems (tactile, visual, and kinesthetic) to build and improve a model of the object. “Manipulation” implies that the robot will physically manipulate the object to help accomplish the modeling task. “Localization” implies that the robot will track the pose of the object during grasp acquisition and manipulation. “Simultaneous” implies that localization, modeling, and manipulation will all occur together – in real time.

The $G-SL(AM)^2$ problem is filtering problem that presents special challenges peculiar to systems with intermittent contact: a nonsmooth dynamic model, a high-dimensional state-space, and unknown contact friction parameters that vary unpredictably. To complicate things further, the dimension of the state-space varies in time as contacts form and break, and each such event effectively changes the structure of the dynamic model.

Our approach to the $G-SL(AM)^2$ problem is based on Bayesian filtering, in particular, particle filtering [4]. A straight forward application would be to directly sample over the whole state space, which is doomed to *fail* in our problem, because in grasping, large portions of the state space are invalid. Specifically, particles should not be chosen that correspond to overlap between the geometric models of the bodies. Some may argue that an approach that checks validity and rejects invalid ones will help to recognize the correct samples, while we insist that such a method would lead to rarely any usable samples when contact exists, as the dimension of the valid state space in this case is lower. This would cause very poor estimation.

To shed light on these issues, we present a case study of a particular scaled-down version of the $G-SL(AM)^2$ problem, using data from our planar grasp acquisition testbed shown in Fig. 1. The main advantages of using this testbed for our initial study are the lower-dimensional state space and the smaller number of unknown model parameters. In this paper, it is assumed that the geometric models and most parameters

This work was partially supported by NSF CCF-0729161 and DARPA W15P7T-12-1-0002.

Zhang is a PhD candidate in the Department of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA. zhangl15@rpi.edu

Trinkle is Professor of Computer Science, Rensselaer Polytechnic Institute, Troy, NY 12180, USA. trinkle@cs.rpi.edu

¹The physical model includes the shapes of the bodies and other quantities such as friction coefficients and the mass of the object.

of the physical model are constant and known; only four friction parameters are assumed unknown.

A. Related Research

In [7], Jia et al. investigated the problem of blindly determining the pose and motion of a planar object with known geometry from pushing the object. Their method used tactile data and geometric models during the pushing process to infer the locations of the contact points on the object and the object pose. In [5], Haidacher *et al.* presented an approach to locally estimate the pose of an object during grasp acquisition in 3D when visual servoing is obstructed by the gripper. The approach first refined the object description offline by characteristic relations between planar facets of the geometric model and stored those values in a description database. After receiving tactile measurements from the robotic hand, this database was searched for possible matching facet combinations to determine the position of the object relative to the hand. Both of these research efforts relied on tactile sensors to infer object pose when object geometry was given and when there was no visual information.

Michael Krainin et al. developed an approach to build 3D surface model of unknown object using data from a depth camera [8]. Their approach does not rely on either highly accurate depth sensor or highly precise manipulator as they use a Kalman filter to help track the arm’s and object’s positions. All these research works focused on estimating object geometries, while we are interested in building a complete model of the system which includes geometry as well as physical parameters. It appears to be infeasible to incorporate such information into their work.

Yuval Tassa et al. formulated a Stochastic Linear Complementarity Problem (SLCP) to apply nonlinear control methods requiring local linearization to frictional contact problems [12]. However, as the authors pointed out, the actual distributions of the system state propagated by the true dynamics became multi-modal upon contact, requiring either nontrivial parametrization or a mixture of samples where the dynamic model is still non-differentiable.

B. Organization of This Paper

Section II introduces a planar grasp testbed and its discrete-time dynamic model in the form of an NCP [9]. The accuracy of this model and the robustness of dVC2d in simulating the grasping process has been shown in [14]. Section III develops a particle filter algorithm that effectively deals with the difficulties of intermittent contact and friction in a general way, allowing application to other contact problems such as assembly problems. In Section IV, we apply our particle filter to a series of planar grasping experiments to demonstrate its ability to track the object and friction parameters, with and without visual and tactile data. Section V summarizes the work and points out future directions.

II. EXPERIMENT AND DYNAMIC MODEL

A. Experimental Set-Up and Assumptions

In our experimental environment shown in Fig. 1, one linear pusher (or thumb) and three fixels (or fixed fingers) are

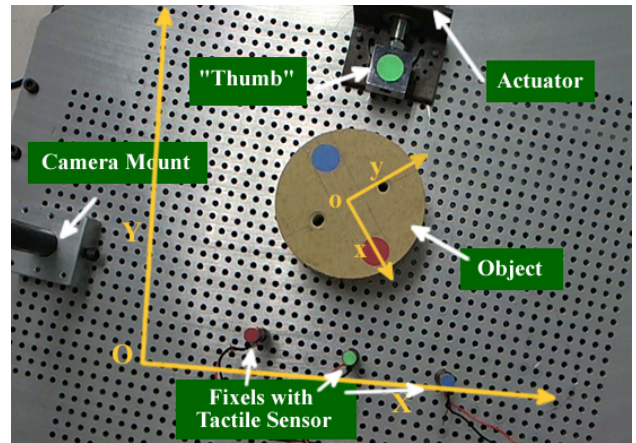


Fig. 1. Planar Grasping Testbed

mounted on an aluminum plate to act as a simple one-degree-of-freedom “hand.” The object is shown in configuration $q = (x, y, \theta)$ defined by the x - and y -coordinates of the origin of the object-fixed frame, and the angle between the x -axes of the base and object frames. During grasp acquisition, the object is pushed toward the fixels, until fixel contacts halt the motion or the pusher reaches its travel limit.

As the system moves, frames action are captured by an overhead camera at 30Hz and are post-processed to extract the configuration trajectory of the object and pusher. After calibration, the camera provides positions and orientations with resolutions of about 1mm in displacement and 1° in rotation. The thumb’s actuator is very stiff with a settable “constant” speed, so its motion is tracked using visual data and a simple Kalman filter.

Each fixel is equipped with a binary tactile sensor. Since we are especially interested in exploring the tactile data’s potential impact on tracking because, in 3D systems, the robot’s view of the object is occluded by the hand. In our testbed, the camera’s view is never occluded. This is a benefit for this study, since we can choose the periods of visual occlusion.

Our main modeling assumptions are as follows. Since gravity acts perpendicular to the support plane and the pusher moves parallel to the plane, we can reasonably assume that the object will maintain face-to-face contact with the support plane. In addition, all bodies are assumed rigid and all contacts are assumed to be discrete points where Coulomb’s Friction Law applies. The face-to-face contact between the object and the support plane is represented by three points of support rigidly fixed to the object. The points form an equilateral triangle whose center is directly below the object’s center of mass. This model greatly simplifies a highly complex friction process that depends on details of the geometries of the object and support surface, material properties, dirt, relative humidity, and more.

B. Dynamic Model

The dynamic model is composed of the Newton-Euler equation (restricted to the plane), non-penetration constraints,

and Coulomb's Friction Law. Frictional contact occurs between the object and fixels, the object and pusher, and the object and the support surface. For the former two, the contact normals lie in the plane of motion, so the friction cones reduce to 2D linear cones. For the support contacts, the contact normals are perpendicular to the plane, so the full 3D quadratic cone is retained. To reduce the problem size, we use a tripod to model the support contacts. Normal and friction forces are assumed to be applied at its three vertices. The resulting dynamic model can be formulated as a mixed nonlinear complementarity problem (mixed NCP) and solved by the `path` solver [3]. This mixed NCP could be linearized by approximating the quadratic cones with polyhedral cones, but previous results show that the NCP model is more accurate and can be solved just as quickly as the linearized model [2].

Below, the mixed NCP implementing the discrete-time, dynamic model is presented in equations (1) and (2) (see [1] for a complete derivation). Let Δt be the (constant) simulation time-step and let $\ell = \{0, 1, \dots, N\}$ denote the index of the current time step. In the NCP formulation, the unknowns are the contact impulses $p^{\ell+1}$, the object velocity $v^{\ell+1}$, and contact slip indicators $\sigma^{\ell+1}$, where the superscript is not an exponent, but rather indicates time $t = \Delta t(\ell + 1)$. All the other terms in the equations are constructed from information available at the current time: p^ℓ , v^ℓ , σ^ℓ , input forcing functions, collision detection algorithms, and physical properties of the system. To simulate the grasping process, this NCP is formulated and solved at each discrete time point, $\ell = 0, \dots, N - 1$. The new velocity $v^{\ell+1}$ is used to update the object's pose with an Euler time-stepping rule.

$$\begin{aligned}
0 &= -Mv^{\ell+1} + Mv^\ell + W_n^\ell p_n^{\ell+1} + W_f^\ell p_f^{\ell+1} \\
&\quad + W_{ts}^\ell p_{ts}^{\ell+1} + W_{os}^\ell p_{os}^{\ell+1} + p_{app}(t) \\
\rho_n^{\ell+1} &= (W_n^\ell)^T v^{\ell+1} + \frac{\psi_n^\ell}{h} + \frac{\partial \psi_n^\ell}{\partial t} \\
\rho_f^{\ell+1} &= (W_f^\ell)^T v^{\ell+1} + E \sigma^{\ell+1} \\
s^{\ell+1} &= U p_n^{\ell+1} - E^T p_f^{\ell+1} \\
s_s^{\ell+1} &= U_s^2 p_{ns}^{\ell+1} \circ p_{ns}^{\ell+1} - p_{ts}^{\ell+1} \circ p_{ts}^{\ell+1} - p_{os}^{\ell+1} \circ p_{os}^{\ell+1} \\
0 &= (U_s p_{ns}^{\ell+1}) \circ (W_{ts}^\ell v^{\ell+1}) + p_{ts}^{\ell+1} \circ \sigma_s^{\ell+1} \\
0 &= (U_s p_{ns}^{\ell+1}) \circ (W_{os}^\ell v^{\ell+1}) + p_{os}^{\ell+1} \circ \sigma_s^{\ell+1}
\end{aligned} \tag{1}$$

$$0 \leq \begin{bmatrix} \rho_n^{\ell+1} \\ \rho_f^{\ell+1} \\ s^{\ell+1} \\ s_s^{\ell+1} \end{bmatrix} \perp \begin{bmatrix} p_n^{\ell+1} \\ p_f^{\ell+1} \\ \sigma^{\ell+1} \\ \sigma_s^{\ell+1} \end{bmatrix} \geq 0. \tag{2}$$

In equation (1) above, the first equation is the Newton-Euler equation. The second imposes linearized non-penetration constraints, the third and fourth are the linear friction laws for the object contacts with the pusher and fixels, and the last three equations encode the friction laws for the object contacts with the support surface (the \circ operator is the

Hadamard product of two vectors). In addition, the subscripts $_n$ and $_f$ indicate that a quantity is related to the normal and frictional components of the contact impulses, respectively. The subscript $_s$ indicates relationship to the support surface, and the subscripts $_{ts}$ and $_{os}$ refer to the two perpendicular components of the support friction impulse in the plane. The other quantities in the formulation are the object's inertia matrix M , the contact constraint Jacobian matrix W , the contact pair distance ψ , the diagonal matrix U of friction coefficients, the selection matrix E , and the impulse p_{app} applied to the body by all non-contact sources .

A key point is that the underlying instantaneous model is *nonsmooth* - not just nonlinear with inequality constraints, but nonsmooth. Kalman filters (which assume a linear dynamic model) do not apply. Model nonsmoothness manifests as impulsive contact forces and discontinuous velocities. These are a result of transitions in and out of contact and between sticking and slipping at sustained contacts. At these times, derivatives with respect to the state variables do not exist. This fact makes the application of extended Kalman filters (which assume differentiability) to our model impossible. Strictly speaking, a Kalman filter or extended Kalman filter could be used, but only after removing the contact terms from the dynamic model and adding noise. This would mean eliminating complementarity conditions (2) entirely and all but the first equation in equation (1). This single equation left is Newton's second law with all contact forces treated as disturbances in p_{app} .

Notice also that the first four equations of system (1) are linear in the unknowns. The last three are bi-linear or quadratic. The derivatives of these expressions with respect to the state variables exist everywhere, so it might appear as though one could use an EKF for the $G-SL(AM)^2$ problem. However, the combination of system (1) with the complementarity conditions expressed in equation (2) causes nonexistence of the needed derivatives.

To form the NCP at the current time, we need estimates of the state of the object $X^\ell = [q^\ell, v^\ell]$ (where q is the object's configuration and v is its time derivative), the physical parameters β^ℓ , and the input u^ℓ . Solving the NCP yields the system state $X^{\ell+1}$ at the end of the current time step. For the planar grasp testbed, the mass and moment of inertia of the object and the shapes of the pusher, object, and fixels may be considered constant and known. However the friction model parameters vary in space and time. These parameters are the radius of the virtual support tripod d_{tri} and the coefficients of friction between the object and pusher μ_p , the object and fixels μ_f , and the object and support μ_s . These four parameters are the elements of β . For the remainder of this paper, we will denote the mixed NCP as Γ , and write the time-stepping subproblem in the following compact form:

$$X^{\ell+1} = \Gamma(X^\ell, u^\ell, \beta^\ell). \tag{3}$$

Note that this form is generally applicable to systems of bodies with intermittent frictional contact, not just planar grasping problems.

III. A PARTICLE FILTER FOR GRASPING

A recursive Bayesian filter that uses a parametric dynamic model of manipulation and data from all available sensors (*e.g.*, visual, tactile, and kinesthetic sensors) can help us to track the object while also estimating model parameter vector β . With this approach, if all the sensors fail, the model will be as accurate as possible, so error accrual during a sensor blackouts will be minimized.

A particle filter is a simulation-based Bayesian filter that aims to sequentially estimate the distribution of the state vector $X^{\ell+1}$ given an observed sequence of output vectors $Y^{1:\ell+1}$ on-line². The estimation process works by iteratively applying a model to predict the state one time step in the future, and then using observation data to improve the prediction. A general discrete-time Bayesian state-space model is given by two conditional probability density functions:

$$\begin{aligned} X^{\ell+1} &\sim P_{X^{\ell+1}|X^\ell}(X^\ell, u^\ell) \\ Y^\ell &\sim P_{Y^\ell|X^\ell}(X^\ell) \end{aligned} \quad (4)$$

where Y^ℓ denotes the observation at $t = \ell\Delta t$. $P_{X^{\ell+1}|X^\ell}(\cdot)$ is the state transition model and $P_{Y^\ell|X^\ell}(\cdot)$ the observation model, or sensor model. Particle filters assume that the system dynamics is a first-order Markov process, which means that past and future data are independent if one knows the current state X^ℓ . Also the observation data is conditionally independent if the state X^ℓ is given. Most multibody dynamic models satisfy these conditions.

The condensation algorithm is a popular particle filtering method. It approximates the state distribution by a weighted set of particles $\langle^{(i)}X^\ell, ^{(i)}g^\ell\rangle$. Each $^{(i)}X^\ell$ is a possible system state, or particle, and $^{(i)}g^\ell$ is the importance weight. The sum of the weights over all the particles must be one. The weighted particles can be used to approximate expectations and higher-order moments of various functions with respect to the approximated state distribution.

A. State Space Definition

To track the object's pose and estimate the unknown model parameters, one may define the filter's state vector as the object's state, the model parameters, and contact states with the fixels. In our problem, the filter state is defined as follows:

$$X = [x, y, \theta, \dot{x}, \dot{y}, \dot{\theta}, \mu_s, \mu_p, \mu_f, d_{\text{tri}}, c_1, c_2, c_3]^T,$$

where c_1 , c_2 , and c_3 are binary contact states between the object and each of the three fixels (Note that it is also possible to incorporate contact force data if the tactile sensors can provide it.). In the next section, if an experiment uses tactile data, the filter state is 13-dimensional otherwise, it is 10-dimensional.

It is worth noting here that the control of the pusher is very stiff. Therefore, in the dynamic model, the pusher is viewed as a position source and *not* a dynamic object. As such, the pusher's position and velocity are not part of the

state in equation (3). Rather they are included in the system input u . The only time when the pusher does not behave as a position source is when the object touches the fixels in a way that causes the pusher to jam (*i.e.*, when the grasp achieves frictional form closure [6], [10]). A final point here is that in our experimental study, we used the actual motion of the physical pusher to drive the simulation model.

B. Issues in Applying Particle Filters to Grasping

A commonly used model in tracking problems is:

$$\begin{aligned} X^{\ell+1} &= f(X^\ell, u^\ell) + w^\ell \quad \text{System} \\ Y^\ell &= h(X^\ell, u^\ell) + \varepsilon^\ell \quad \text{Observation} \end{aligned} \quad (5)$$

In this model, w denotes the process noise and ε the observation noise. Here both w and ε are mutually independent distributed sequences with known probability density functions. The functions, $f(\cdot)$ and $h(\cdot)$, are known deterministic state transition and observation functions.

If we use this form of the system model (by replacing f with Γ), the process noise w , which corresponds to random rigid body displacements and velocity disturbances in the ambient state space, will be added to the state predicted by the deterministic dynamic model of the grasping system at each particle. Physically, this means that in configurations involving contact, majority of the particles will correspond to configurations in which bodies are overlapping or are separated when the deterministically solved result corresponds to bodies in touching. And only the configuration involving touching is valid. Theoretically, all valid samples should come from a lower-dimensional manifold defined by the contact constraints. As a result, the particles and weights sampled from this model will be a very poor approximation of the state distribution. An approach that performs contact detection and rejects invalid particles will be able to eliminate particles in invalid configurations. However, we argue that such an approach cannot avoid the problem of having too few effective particles and this would cause very poor state representation.

As to parameter estimation, previous work in [13] introduced a way of estimating unknown parameters by composing a parameter dynamic model in addition to the system dynamic model, thereby forming an expanded state transition model. The main complications here are that there are no physically-motivated dynamics for the friction parameters, and the system model cannot generate useful estimates of parameters that are not currently impacting the system. As an example, before the object contacts a fixel, the object-fixel friction coefficient μ_f cannot be estimated, so computations associated with its estimation would be wasteful. Our approach to handling this problem as well as the previous issue will be described in the next two subsections.

C. Solution

We designed a specific particle filter based on the condensation algorithm. The process noise is broken into two

²A tutorial on particle filtering methods and a mathematical derivation can be found in [11].

components, external force noise w_{app} and parameter noise w_{β} . The force noise enters the dynamic model as follows:

$$M\mathbf{v}^{\ell+1} = M\mathbf{v}^{\ell} + W^{\ell}p^{\ell+1} + p_{\text{app}} + w_{\text{app}} \quad (6)$$

$W^{\ell}p^{\ell+1}$ represents the sum of all contact and friction impulse terms from the first equation of equation (1) and we denote the new probabilistic dynamic model as Γ' . Adding noise in this way guarantees that the nonpenetration and friction constraints will be satisfied at the end of the time-step at each particle, because the probabilistically predicted state is obtained by solving the NCP problem and thereby satisfies the constraints at every particle. The parameter noise is added to the parameter vector at each particle to allow its estimate to evolve as the state estimate evolves. In particular, we generate an intermediate value $\hat{\beta}^{\ell+1}$ as: $\hat{\beta}^{\ell+1} = \beta^{\ell} + w_{\beta}$. Now, the expanded state transition model can be written as:

$$\begin{aligned} \mathbf{v}^{\ell+1} &= \Gamma'([q^{\ell}, \mathbf{v}^{\ell}], u^{\ell}, \hat{\beta}^{\ell+1}, w_{\text{app}}) \\ \beta^{\ell+1} &= \Phi(\beta^{\ell}, \hat{\beta}^{\ell+1}) \\ q^{\ell+1} &= q^{\ell} + \Delta t \cdot \mathbf{v}^{\ell+1} \end{aligned} \quad (7)$$

where $\Phi(\cdot, \cdot)$ encodes rules that update only those parameters which are actively involved in the evolution of the system state according to the rules in Table I. When the state includes c_1 , c_2 , and c_3 , their values will be obtained from collision detection at every time step.

As we pointed out earlier, directed sampling over the ten (thirteen including contact states) dimensional space is doomed to fail, because the state space corresponding to physically valid configurations is constrained to a lower-dimensional manifold. By propagating the system state with random general applied force disturbances through the non-linear complementarity formulation, which explicitly models the non-penetration constraint and stick-slip behaviour, we guarantee that the samples are constrained in this valid subspace throughout the process. As implicitly-involved system state components during the MCP formulation, the contact variables c_1 , c_2 , c_3 are also effectively constrained because they are determined by performing collision detection and thereby agree with the samples of the position. By exploiting these constraints, we greatly reduced the number of samples required. Note that $2^{10} = 1024$ samples are only enough to cover two values of each variable if we adopt grid sampling over the whole space, yet we found 500 particles good enough for our system.

The observation model is hybrid because it involves both continuous visual sensor and discrete tactile sensor. The continuous part of Y is defined as $[x, y, \theta]^T + \varepsilon$, where ε is a Gaussian noise. The discrete part is defined as $[\hat{c}_1, \hat{c}_2, \hat{c}_3]$, where \hat{c} represents real tactile sensor reading. The discrete probability model of each tactile sensor $p(\hat{c}|c)$ was approximated through experiments.

D. Algorithm

We assume the initial distribution of q is normal with mean equal to the first observation data. The variances were carefully designed to balance granularity and diversity of

TABLE I
SELECTIVE STRATEGY OF ϕ

contact between the pusher and object exists	choose $\hat{\mu}_p^{\ell+1}$
contact between the pusher and a fixel exists	choose $\hat{\mu}_f^{\ell+1}$
$\ \mathbf{v}^{\ell}\ > 0$ or $\ \mathbf{v}^{\ell+1}\ > 0$	choose $\hat{\mu}_s^{\ell+1}$ and $d_{\text{tri}}^{\ell+1}$

the particles. The system is initially at rest and the object does not contact the fixels. All initial particles generated share common initial parameter values, namely a nominal parameter set $\beta_0 = [\mu_{s0}, \mu_{p0}, \mu_{f0}, d_{\text{tri}0}]$, and are assigned equal weights. The parameter values could be obtained via offline calibration or from relevant databases.

The general algorithm framework is defined as:

Algorithm: Grasp Acquisition Particle Filter

For each time step $\ell = 1, \dots, N$:

 If resample condition satisfied, then

 Resample all particles

 For each particle $i = 1, \dots, N_p$:

 Run the system transition model defined in eq (7)
 if sensory data is available, then

 Update particle weight ${}^{(i)}g^{\ell}$ based on $P_{Y^{\ell}|X^{\ell}}(\cdot)$

 else

 Particle weight is unchanged

The resample condition adopted here is that the effective number of particles $N_{\text{eff}} = \frac{1}{\sum_{L=1}^{N_p} ({}^{(L)}g^{\ell})^2}$ is less than the threshold $N_{\text{thr}} = 0.3N_p$.

IV. APPLICATION TO REAL EXPERIMENTS

To demonstrate the effectiveness of the proposed approach, we now present the results from using our particle filter to process the visual data gathered in a series of experiments in which object geometries, initial object poses and fixel configurations were varied.

A. Overall Comparison Results

Tracking results of the particle filter (PF) were compared to the results from deterministic simulations using fixed nominal values of the unknown parameters, which we call nominal simulation (NS), as well as results from Kalman filter (KF) using the linearization as discussed in II-B. As presented above, our main interests lie in incorporating different sensor data to produce robust filtering estimates, hence we tested the particle filter with different sensor availabilities. We intentionally removed visual data during the last phase of the experiments to create visual blackouts. The starting time of visual occlusion was selected before the object made contact with any fixel, taking into account that real fingers can block the view when forming contact with the object. During the visual blackout, we tested the particle filter with and without binary tactile data. For brevity, we will refer to them as PF_(t only) and PF_(none). Please note that the description in the parenthesis only refers to the sensor availability during the blackout. Similarly, we name the original PF as PF_(v&t).

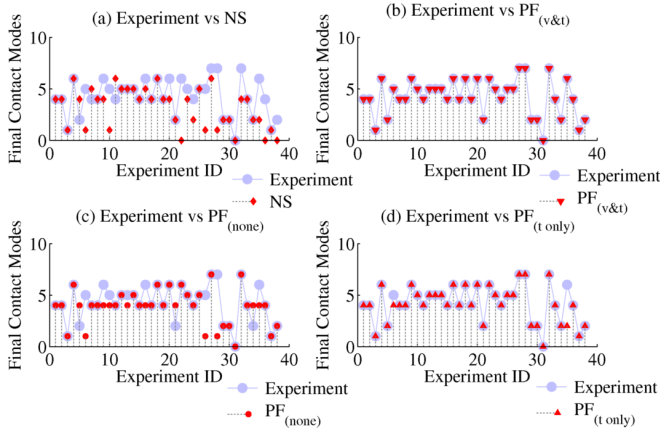


Fig. 2. Final Contact Modes Comparison from 38 experiments

During blackouts, $PF_{(t \text{ only})}$ updates all particles only based on binary tactile data, while $PF_{(none)}$ updates all particles using the system transition model without adjusting the particle weights or introducing parameter noise. As a result, the friction parameter estimates do not change since blackout started for $PF_{(none)}$.

A contact mode defines how the object formed contact with the three fixels and final contact mode is constant after the pusher jammed object on any fixel and formed frictional form closure or pusher slid past the object. We used a single bit to represent the binary touch-or-not against each fixel and thereby formed an integer between 0 (000) - 7 (111) to represent all possible contact modes. It is desired that the tracking algorithm capture this value correctly not only out of pursuit of accuracy, but also because it will strongly affect the manipulation modeling after acquisition succeeded.

Fig. 2 shows the final contact modes from differently processed results compared against the experiment ground truth.

Since KF cannot easily incorporate any non-penetration constraints, during blackouts, KF will simply assume that the object keeps moving and never stops. Final configurations of the object always ended up overlapping with one or more fixels, which correspond to invalid contact modes. That is why the result was not listed here.

One can easily notice that $PF_{(v\&t)}$ was able to capture contact modes correctly for all experiments, while NS gave the worst performance. $PF_{(t \text{ only})}$ failed in two cases. After investigation, we found that both cases had very long visual occlusion range due to very early contact with one fixel and the missed fixel contact happened very late during the experiment. Therefore our inference is that the filter did not get good enough parameter estimates before the blackout and the missed tactile readings were too late to prevent the filter's divergence. Table II gives the percentage of wrong estimates which were calculated by dividing the amount of incorrectly estimated contact modes by the total number of experiments. One can notice that $PF_{(t \text{ only})}$ performs nearly as well as $PF_{(v\&t)}$ and much better than $PF_{(none)}$ and NS. Considering that we only used binary tactile sensors, it is

TABLE II
INCORRECT FINAL CONTACT MODE ESTIMATES PERCENTAGE

NS	$PF_{(v\&t)}$	$PF_{(t \text{ only})}$	$PF_{(none)}$
21.05 %	0 %	1.75 %	10.53 %

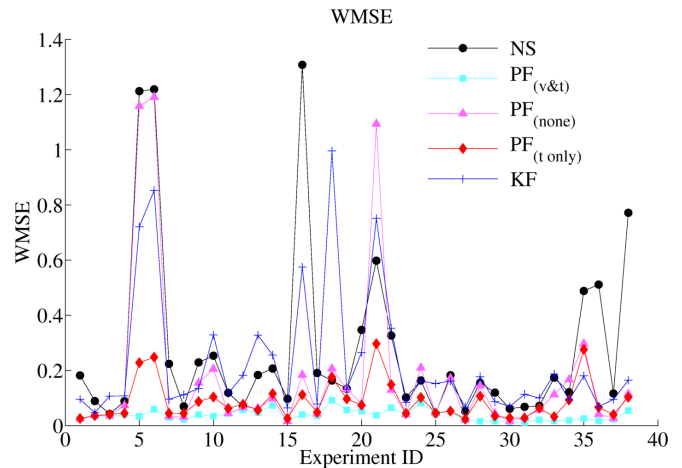


Fig. 3. WMSE Comparison from 38 experiments

already quite an improvement.

To compare the trajectory estimates as a whole and quantify the filters' accuracy with various data sources, we introduce a weighted statistical value WMSE (weighted mean square error):

$$WMSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (\hat{x}_i - \bar{x}_i)^2 + (\hat{y}_i - \bar{y}_i)^2 + [\eta(\hat{\theta}_i - \bar{\theta}_i)]^2}$$

where $\hat{\cdot}$ denotes the estimated or simulated trajectory component, $\bar{\cdot}$ denotes the observed trajectory component, and η is the so-called "virtual radius," which acts as a weighting coefficient to balance the orientation component and the two translational components.

Of course the observed trajectory doesn't equal to the ground truth because of the camera system's noise. But given the fairly high resolution of the overhead camera detection system, the smaller the WMSE value is, the more accurate tracking we will get.

The WMSE values from NS, PF with different sensor data availabilities and KF were given in Fig. 3. It is quite obvious that the two lowest curves are from the $PF_{(v\&t)}$ and the $PF_{(t \text{ only})}$. NS and KF share the highest position. It is pretty clear from the plot that, by incorporating tactile sensors, the particle filter works much better.

All processed results of the experiments are available at <http://www.cs.rpi.edu/~zhangl15/all/>.

B. A Typical Experiment

This experiment was a typical one from a set of experiments characterized by extended sliding against the fixels, ultimately ending with contact with all three fixels. Fig. 4 gives a series of frame captures from the experiment video.

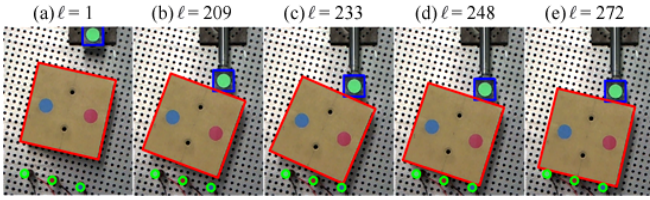


Fig. 4. Experiment Frameshots: (a) object at rest (b) object beginning to move (c) object touched one fixel (d) object sliding against the fixel (e) object stopped after touching all fixels

Colored plots on top of the image show the detected component edges. Visual information was taken away just before the first fixel contact. Besides comparing state trajectories numerically (see Fig. 5), the state estimates were mapped back onto the camera images to allow visual verification.

Six trajectory plots are shown in Fig. 5: raw data (green), NS (black), $PF_{(v\&t)}$ (cyan), $PF_{(t\ only)}$ (pink), $PF_{(none)}$ (red), and KF (blue). The pusher contacts the object at time-step $\ell = 150$. The first fixel contact occurs at $\ell = 233$ and the grasp is completed at $\ell = 272$. Visual occlusion was applied at $\ell = 210$. Fig. 6 shows predicted object poses on camera images at time $\ell = 272$. The colors of the projected object correspond to the filter trajectory colors.

Consider the pink plots in Fig. 5 and the corresponding images in Fig. 6. One can see from the former that at the time of the first fixel contact, $PF_{(none)}$ estimates the object’s pose accurately. However, by the end of the grasping process, less than two seconds later, the error has become significant. The filter indicates contact with only one fixel.

Comparing images in Fig. 6, $PF_{(t\ only)}$ clearly outperforms $PF_{(none)}$ by the end of the grasping process. The trajectory plots in Fig. 5 also support this conclusion, which might not be so obvious with the x -error plot. Notice that this plot ends with nearly $2mm$ of error. This is because the binary tactile sensor data contains no information about the lateral displacement when sliding along all three contacts.

The performance of the KF in this experiment is not as good right after the object contacts the pusher ($\ell \approx 150$) and after visual occlusion begins. At the former time, a transient was caused by the contact impulse that the KF treats as disturbances. The reason that caused divergence during the occlusion has been explained in the previous subsection.

Fig. 7 shows the friction parameter trajectories estimated by $PF_{(v\&t)}$ and $PF_{(t\ only)}$. Parameter estimates from $PF_{(none)}$ are exactly the same as that of $PF_{(v\&t)}$ except that during occlusion, the estimated parameter values do not change, hence it is not shown here. Notice that the pusher friction coefficient μ_p and the support friction parameters, μ_s and d_{tri} , start to vary as soon as the pusher touches the object. However, the filter cannot gain any knowledge of μ_f until object touches a fixel. Except for μ_f , all three parameters vary within a relatively smaller range after some initial transients.

One should also observe the spikes in the plots, especially for d_{tri} . This is not an indication that the real parameters underwent drastic sudden changes. Actually, the spikes are caused by the fixel contact. Specifically, since the grasp sticks

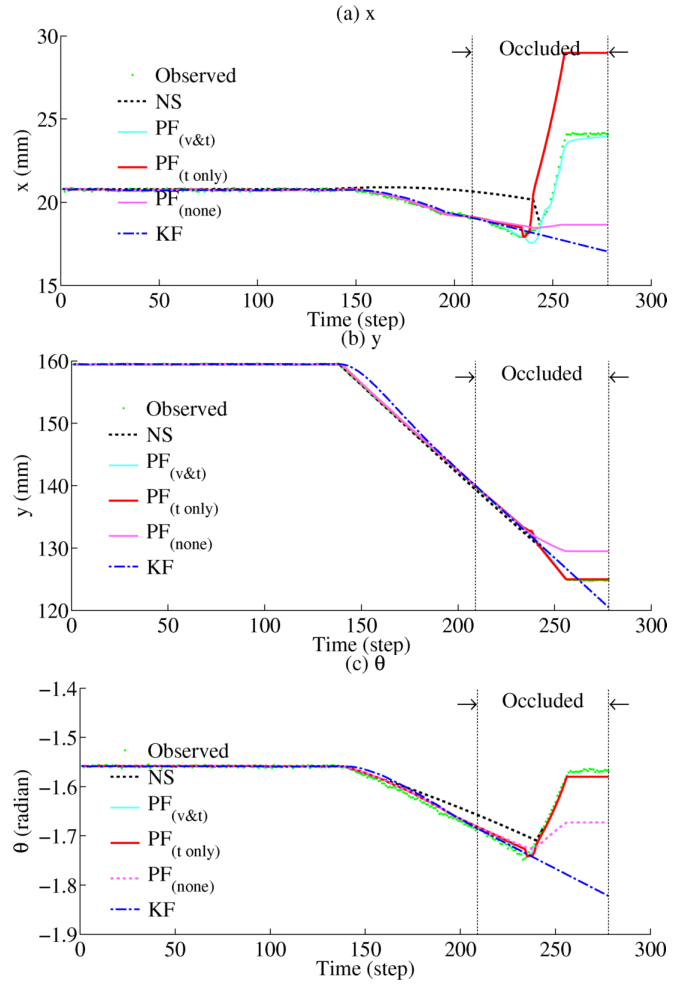


Fig. 5. Estimated Object Trajectory: (a) x (b) y (c) θ

to the two contacts, the dimension of the effective state space of valid particles is reduced. As a result our algorithm produced fewer effective particles than necessary because many of the particles generated at this time corresponded to frictional form closure and cannot proceed, i.e., the filter experienced “particle impoverishment” during this tightly constrained phase. The parameter estimate thereby undergoes a big jump. This scenario reveals that sometimes the effective state space is so tightly constrained that it is very hard to “squeeze” into this low-dimensional subspace and generate enough sample from it.

C. Computational Issues

In the current work, we found 500 particles to be sufficient for good filtering results. Each particle runs collision detection, forms NCP problems, and calls `path` [3] to solve the dynamic model independently of the other particles. Due to this parallel nature, we adopted parallelism when computing state transitions for the particles. However, processing a pre-collected 20-second experiment currently requires 10 minutes of cpu time on a i7-core desktop PC. This is far slower than real time, and if we expand the filter state to include geometric model parameters of the object, then we will

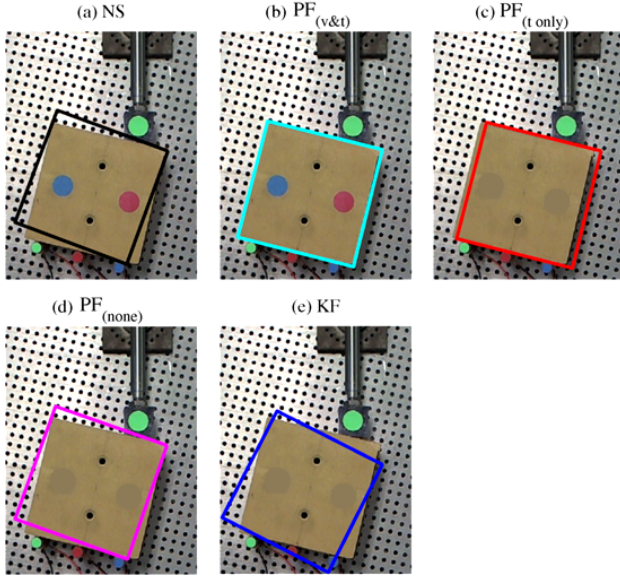


Fig. 6. Overlaid Estimated Object Pose at $\ell = 272$: (a) NS: object was stuck with one fixel (b) $PF_{(v\&t)}$: object correctly touched all three fixels (c) $PF_{(t \text{ only})}$: object correctly touched all three fixels (d) $PF_{(none)}$: object was stuck with one fixel (e) KF: object penetrated two fixels

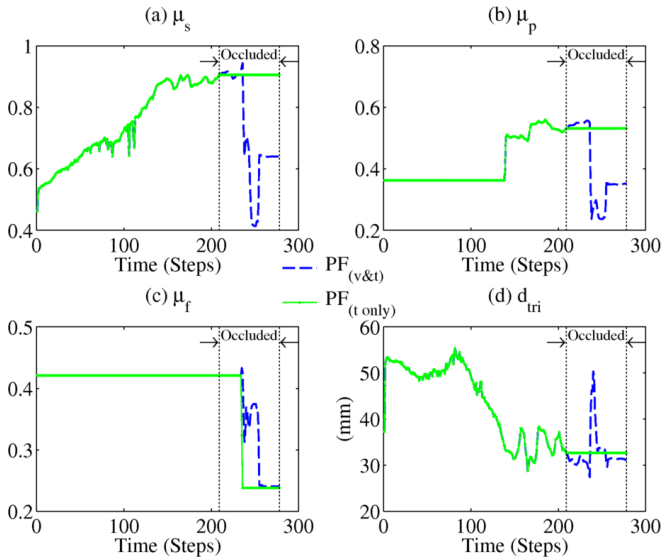


Fig. 7. Estimated Friction Parameters: (a) object-surface friction coefficient (b) object-pusher friction coefficient (c) object-fixels friction coefficient (d) “support tripod” diameter.

require even more particles and cpu time. Nonetheless, we have reason to be optimistic. Over 90% of the computational effort was spent on solving the dynamic model. New solvers based on new formulations of the dynamic model have the potential to be hundreds times of faster when deployed on multi-threaded CPUs and GPUs.

V. CONCLUSIONS AND FUTURE WORKS

In this paper we developed an approach to apply particle filtering to the $G\text{-SL(AM)}^2$ problem. The approach ensures that non-penetration and stick-slip friction constraints of the dynamic model are always satisfied with all particles

and also provides a reasonable model to refine the physical parameters. From the particular grasping experiments studied, one can conclude that the proposed approach can be effective in improving our knowledge of the system’s physical parameters while simultaneously tracking the object during visual occlusion. While this paper focused on the $G\text{-SL(AM)}^2$ problem, the approach taken is applicable to a wide variety of multibody systems with intermittent contact.

For future work, it will be important to solve the “sample impoverishment” problem that arises when the valid subset of state space is tightly constrained by complex contacts. We plan to investigate hybrid sampling from the sensory model. We also plan to incorporate geometric model parameters and extend the filtering scheme to three-dimensional problems. Applying the framework presented here directly to more complex problems could increase the computational burden unacceptably, as both the number of particles required and the sizes of the MCPs to be solved will grow. Therefore, we are actively looking into the possibilities of applying Rao-Blackwellization within our filtering framework and parallel computation for both MCP solutions and particle processing.

REFERENCES

- [1] S. Berard. *Using Simulation for Planning and Design of Robotic Systems with Intermittent Contact*. PhD thesis, Rensselaer Polytechnic Institute Department of Computer Science, 2009.
- [2] S. Berard, B. Nguyen, K. Anderson, and J.C. Trinkle. Sources of error in a simulation of rigid parts on a vibrating rigid plate. *ASME Journal of Computational and Nonlinear Dynamics*, 5(4):1–14, 2010.
- [3] M.C. Ferris and T.S. Munson. Interfaces to path 3.0: Design, implementation, and usage. *Journal of Computational Optimization and Applications*, 12(1-3):207–227, January 1999.
- [4] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-gaussian bayesian state estimation. *Radar and Signal Processing, IEE Proceedings F*, 140(2):107–113, April 1993.
- [5] S. Haidacher and G. Hirzinger. Estimating finger contact location and object pose from contact measurements in 3d grasping. In *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on*, volume 2, pages 1805 – 1810 vol.2, September 2003.
- [6] L. Han, J.C. Trinkle, and Z. Li. Grasp analysis as linear matrix inequality problems. *IEEE Transactions on Robotics and Automation*, 16(6):663–674, 2000.
- [7] Y.B. Jia and M. Erdmann. Pose and motion from contact. *Journal of Robotics Research*, 18(1):466–490, May 1999.
- [8] M. Krainin, P. Henry, X. Ren, and D. Fox. Manipulator and object tracking for in-hand model acquisition. In *Proceedings, IEEE International Conference on Robots and Automation*, 2010. Workshop on Mobile Manipulation and Best Practices in Robotics.
- [9] K.G. Murty. *Linear Complementarity, Linear and Nonlinear Programming*. Helderman-Verlag, 1988.
- [10] D. Prattichizzo and J.C. Trinkle. Grasping. In B. Siciliano and O. Khatib, editors, *Handbook of Robotics*. Springer-Verlag, 2008.
- [11] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House, 2004.
- [12] Y. Tassa and E. Todorov. Stochastic complementarity for local control of discontinuous dynamics. In *RSS*, 2010.
- [13] E. A. Wan and R. Van Der Merwe. The unscented kalman filter for nonlinear estimation. In *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. The IEEE 2000*, pages 153–158, August 2002.
- [14] L. Zhang, J. Betz, and J.C. Trinkle. Comparison of simulated and experimental grasping actions in the plane. In *First International Multibody Dynamics Symposium*, June 2010.