Inferring hand-object configuration directly from tactile data

Robert Platt Jr, Frank Permenter, Joel Pfeiffer

Abstract— **The potential utility of tactile sensors in manipulation has long been recognized. However, there are very few examples in the literature of systems that use tactile information any more complex than binary contact/no-contact sensors. This paper approximates a direct mapping between hand-object state and high-dimensional tactile measurements based on training data. Although it can be precise, tactile data provides only partial state information at any given time. Therefore, we integrate tactile information over time using a particle filter. Several variations on a basic measurement model are considered and the paper evaluates the relative utility of each one. Since we create our model from a training data set rather than building a geometric model of the manipulation interaction, the approach works well localizing difficult-to-model objects such as flexible materials or fabrics. We demonstrate that the approach can localize a haptic feature in a piece of flexible plastic to within one tenth of an inch using sensors mounted on the phalanges of human-sized fingers. All demonstrations are performed using Robonaut 2.**

Fig. 1. Robonaut 2 hand localizing a bump in a piece of flexible plastic.

I. Introduction

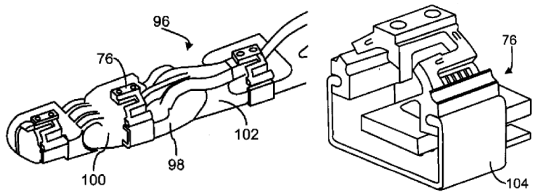
The problem of localizing hand-object configuration from tactile data is important to robot manipulation. While extensive data is available from head-mounted cameras or laser range sensors, this information suffers from occlusions and typically depends on precise calibration between the head and end-effector. Although tactile data does not provide information about the entire manipulation interaction in a single time step, it measures partial information regarding the state of the manipulation interaction very accurately. However, although tactile sensing has great potential, only a few methods have been proposed for using it to locate or characterize an object that a robot touches [1], [2], [3], [4]. These methods generally track the pose and shape of the object based on the position of the parts of the manipulator that touch the object. In this paradigm, tactile sensing is used only indirectly to measure the presence of contact. One possible exception is [3], where the authors augment contact position information by estimating the contact surface normal as well. Nevertheless, the three-dimensional surface normal encapsulates only a small portion of the total information that is potentially available from tactile measurements. It is also based on a geometric model of the interaction and as a consequence requires carefully calibrated load cells to estimate the surface normal. Moreover, since these methods rely on matching hand geometry to object geometry, they are completely unsuitable for state estimation with flexible or deformable materials.

This paper considers the problem of estimating hand-object configuration directly from tactile information. In particular, we focus on the problem of locating a haptic feature (a grommet, button, snap, *etc.*) embedded in a flexible material (Figure 1). The flexible material is compliantly squeezed between the fingers of a robot hand and pulled so that the material slides through the fingers. Tactile sensors embedded in the finger tips provide relevant measurements as they move over the grommet. We experimentally show that a characteristic signature is obtained when the fingers move over the haptic feature as described above. This signature traces out part of a "tactile map." Our approach is to construct the tactile map that is subsequently used during a training phase as a basis for localization. On the assumption of non-zero process noise (uncertain relative hand-object motion), we use a particle filter to track hand-object state. Essentially, the particle filter matches the haptic signature obtained during testing with an appropriate corresponding trajectory on the map. We find that the analytical form of the measurement model has important performance implications. We compare the performance of a single Gaussian model with a multi-Gaussian model. We also propose a method for averaging "featureless" measurements in the model in order to minimize overfitting. One advantage of the proposed approach is that it does not rely on precise robot calibration because the haptic map is constructed during a training phase. In fact, it is unnecessary to calibrate the tactile load cells at all; the training phase is capable of modeling raw sensor measurements as well as measured loads. The paper first describes our experimental scenario and the tactile

Robert Platt Jr. is with the Computer Science and Artificial Intelligence Laboratory at MIT. This work was performed while he was at NASA Johnson Space Center. rplatt@csail.mit.edu

Frank Permenter is with Oceaneering Space Systems. frank.n.permenter@nasa.gov

Joel Pfeiffer is with the Computer Science Department at Purdue University. jpfeiffer@purdue.edu



(a) (b)

Fig. 2. Each phalange of the R2 finger is equipped with a phalange load cell. (a) illustrates the finger with attached load cells. (b) shows the design of a single load cell.[2]

sensors involved. Then, we propose various forms of the measurement model and compare their relative performance.

II. Tactile interaction scenario

A significant amount of relevant tactile information can be measured by touching material with high quality tactile sensors. We summarize the design of the tactile sensors and introduce the tactile map.

A. Robonaut 2 experimental setup

.

Our approach to hand-object localization is predicated on manipulator compliance and high-quality tactile sensors. In general, these capabilities can be realized in a number of ways. The Robonaut 2 implementation involves a humanoid robot hand with active stiffness control in the fingers [5] and miniature load cells mounted on the finger phalanges [6]. Each load cell is comprised of a cap that captures and transmits applied loads to a spring element (Figure 2) instrumented with flea-sized strain gauges that actually measure the loads. The cylindrical load cell form factor allows it to be mounted on the phalanges of the Robonaut 2 fingers. The shape of the spring element is such that any wrench applied to the cap results in a bending moment in one or more of bars instrumented by strain gauges. The mechanics of the spring element are such that an affine relationship exists between the measured strains and the applied loads. The sensor cap also provides overload protection by making secondary contact with the finger itself when excessive loads are applied.

The experimental setup involved a flexible material (for example, a flexible sheet of plastic) loosely hanging from a test rig such that it is free to swing or otherwise move with respect to the robot. The material was held between the thumb and forefingers (index and middle fingers) of the Robonaut 2 hand as illustrated in Figure 1. The thumb was put in position control mode such that the thumb tip position remained roughly constant. The index and middle fingers were put in torque mode so that the material was squeezed with a constant and controlled force. During the experiments, the robot hand was pulled away from the test rig in controlled "sweeps." As the hand moved, the material was pulled through the fingers such that the fingers experienced aspects of the haptic feature.

[2]Images taken from US patent application number 20100077867 [6] and can be downloaded from the USPTO website, http://patft.uspto.gov.

B. Interacting with the flexible material

The basic mechanism for interacting with the flexible material is to pull it through the fingers by squeezing it between the thumb and forefingers as illustrated in Figure 3(a). At each time step during the interaction, each sensor generates a measurement vector (8-dimensional per phalange sensor) where each element corresponds to a voltage generated by a pair of strain gauges. When the fingers interact with a discontinuity (*i.e.* a haptic feature) in the surface of the material, a characteristic sensor signature is generated. Since each strain gauge pair responds differently to applied loads, this signature is different for each element of the measurement vector.

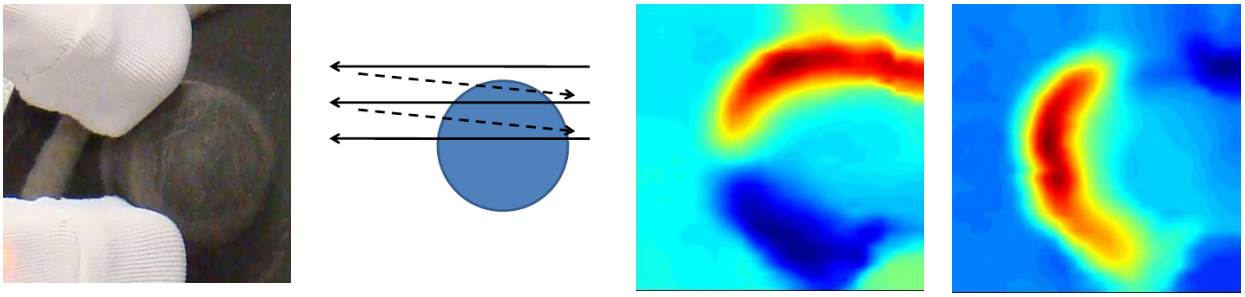
Figures 3(c) and 3(d) illustrate elements of the haptic map for the bump in flexible plastic (Figure 3(a)) acquired by comprehensively "scanning" the hand over the material surface. This process is illustrated in Figure 3(b). The hand makes a series of sweeps laterally across the material and gradually descending. Each sweep is approximately one tenth of an inch below the previous one. After the scanning process is complete, the maps illustrated in Figure 3 are constructed by locally interpolating the measurements at regular grid points. Each of the two maps shows the interpolation for one element of the measurement vector (for Robonaut 2, there are a total of eight elements in each load cell.) Notice that the two elements of the sensor signal respond to the interaction in different ways. The signal in Figure 3(c) responds depending upon where the load cell is with respect to the bump vertical dimension. The signal in Figure 3(d) responds depending upon where the load cell is vertically.

III. Hand-object state estimation

While manipulating the object, the unpredictable nature of the hand-object interaction results in non-zero process noise. As a result, it is not strictly correct to align measurements from different time steps into a single frame and estimate the maximum likelihood state. Instead, we use a particle filter to solve a Bayesian filtering problem. In order to apply the Bayes filter, it is necessary to estimate the likelihood of measurements given hand-object state. This paper considers several possible models starting with a single Gaussian model in this section.

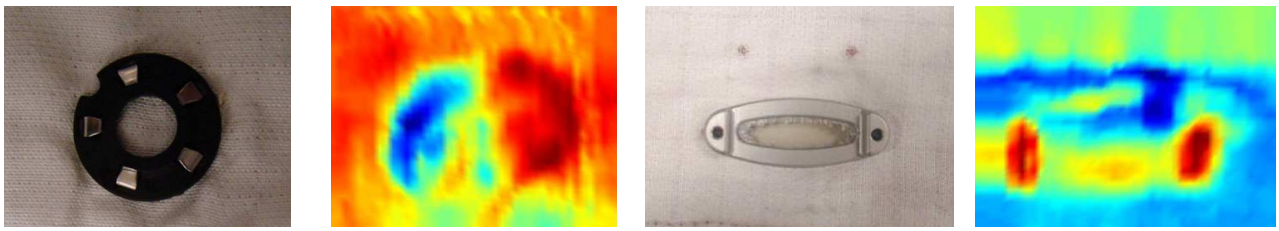
A. Bayesian filtering

The goal of Bayesian filtering is to track the state of a stochastic system as it changes. It is assumed that state, x, is a stochastic Markov function of time. At every time step, the measurements, z, depend only on the current state. Starting with a prior distribution over state, $P(x_0)$, Bayesian filtering recursively updates a posterior distribution, $P(x_t|z_{1:t})$, where x_t is the state at time t and $z_{1:t} = \{z_1, \ldots, z_t\}$ is the set of measurements between time 1 and time t. The update to the posterior (also called the "belief state") is



(a) (b) (c) (d)

Fig. 3. Creation of the "tactile map." The fingers make a series of lateral sweeps from right to left. (a) shows the fingers making such a sweep over a bump in flexible plastic. (b) shows the path of these sweeps. (c) and (d) show two elements (out of a total of eight per sensor) of the haptic map for the bump. Notice that the two elements of the sensor signal respond differently to the interaction.



(a) (b) (c) (d)

Fig. 4. Tactile maps for a snap (a) and (b) and a grommet (c) and (d).

accomplished in two steps. First, the prediction step updates the distribution by applying a system model:

$$P(x_t|z_{1:t-1}) = \int P(x_t|x_{t-1})P(x_{t-1}|z_{1:t-1})dx_{t-1} \quad (1)$$

The above uses the Markov assumption that $P(x_t|x_{t-1}, z_{1:t-1}) = P(x_t|x_{t-1})$. In the second step, the posterior distribution is updated in proportion to the likelihood of having generated the observed measurements, z_t:

$$P(x_t|z_{1:t}) = \eta P(z_t|x_t)P(x_t|z_{1:t-1}), \quad (2)$$

where

$$\eta = \frac{1}{P(z_t|z_{1:t-1})}$$

is a normalizing constant.

Equations 1 and 2 constitute an optimal solution to the problem of tracking state in a Markov system. However, they ignore the question of how the posterior distribution is represented. Two popular solutions to this problem are the Kalman filter and the particle filter. The Kalman filter is optimal, but makes strict (linear system, Gaussian noise) assumptions regarding the system and measurement models. The particle filter does not make these assumptions, but relies on Monte Carlo methods that depend on an adequate sampling the posterior distribution. This paper uses the sample importance resampling (SIR) version of the particle filter [7] to track hand-object state.

B. Single Gaussian measurement model

The single Gaussian model estimates the likelihood of the measurements, $P(z|x)$, as a Gaussian distribution. At a given point in time, the state, x, of the system is the position of the robot palm with respect to the fabric, and the measurement, z, is the vector of sensor measurements. We require the robot to interact with the flexible material in a training phase where the robot sweeps its fingers over the material in a structured way as illustrated in Figure 3(b). The goal of training is to construct a dataset, $\mathcal{D}$, consisting of state-measurement pairs, (x, z). Since state is defined relative to the material itself, it is important during training for the material to be fixtured relative to the robot base frame so that relative palm-material position can be inferred from the palm position relative to the base.

In order to evaluate the likelihood of $P(z_i|x_i)$, we take a locally-weighted approach where the moments of the Gaussian are estimated using the sample moments of the k nearest neighbors. Let $N_k(x_j) \subseteq \mathcal{D}$ be the k pairs, $(x, z) \in \mathcal{D}$ nearest x_i under a Euclidian distance metric. Then the likelihood is:

$$P(z_i|x_i) = \mathcal{N}(z|\hat{z}_i, \Sigma_i), \quad (3)$$

where

$$\hat{z}_i = \frac{1}{k} \sum_{(x,z)\in N_k(x_i)} z,$$

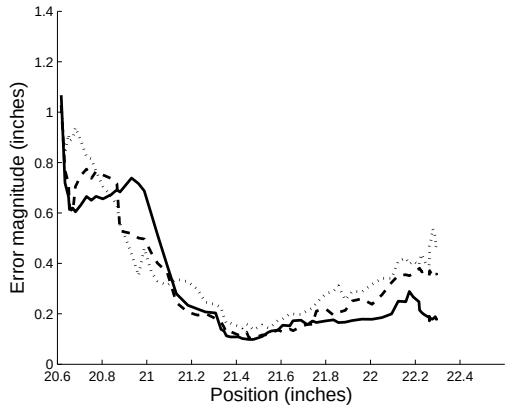


Fig. 5. Localization accuracy as a function of horizontal (*i.e.* direction of the sweeps) tactile sensor location averaged over 18 trials at different vertical locations in the state space. The dotted line shows localization performance for the fully coupled single Gaussian measurement model. The dashed line shows performance for the Gaussian model with inter-finger covariance set to zero. The solid line shows the performance of the sum-of-Gaussians model.

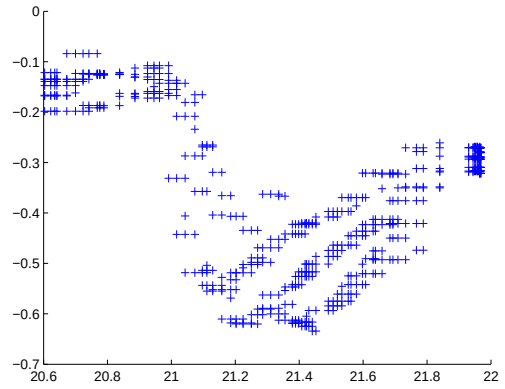


Fig. 6. Measurements as a function of finger position (one element of the measurement vector) for a single trajectory through state space. The figure shows measurements corresponding to the 8 nearest neighbors in state space for each point along the trajectory.

and

$$\Sigma_i = \frac{1}{k} \sum_{(x,z)\in N_k(x_j)} (z - \hat{z}_i)(z - \hat{z}_i)^T$$

are the sample mean and covariance over the k neighbors.

The localization error of particle filtering using the Gaussian model is illustrated by the dotted and dashed lines in Figure 5. The state space is a two-dimensional two inch square, $x = (x_1, x_2)$ with $x_1 \in [20.5, 22.5]$ and $x_2 \in [-13.5, -11.5]$. The particle filter uses 75 particles including 10 particles that are placed uniformly randomly on each update. The dataset, $\mathcal{D}$ is comprised of 16321 samples constructed by scanning the material as illustrated in Figure 3. Figure 5 gives results averaged over 18 test scans. After tracking grommet position using the particle filter in 18 different scan configurations, the error as a function of position was averaged for the 18 runs and reported. The dotted line shows average particle filter performance when Equation 3 uses a covariance matrix calculated across tactile measurements from two fingers (the index and middle fingers). The dashed line shows performance when it is assumed that there is zero covariance between the two fingers (off-diagonal covariance terms between measurements from different fingers are set to zero). Both measurement models perform approximately equally as well for positions less that 21.2 (before the fingers touch the bump.) After touching the bump, there appears to be a small advantage to requiring zero covariance between the fingers. This suggests that the inter-finger covariance found in the data is actually overfit and would disappear with a larger dataset. Since only a limited amount of data is used in the experiment, it helps to enforce this constraint. Notwithstanding this difference, it seems that localization accuracy generally degrades (between positions 21.6 and 22.2) after initially contacting the bump at 21.4. One can guess that the reason for this is illustrated in Figure 3(c) and (d). Notice that in these two elements of the measurement vector, the prominent features (red and dark blue) are clustered at the edges of the bump rather than in the middle. This suggests that localization occurs primarily during periods with prominent measurements. Then, when the fingers are on the bump and the measurements are less descriptive, localization error gradually integrates.

C. Sum-of-Gaussians model

Modeling the likelihood distribution as a single Gaussian ignores the possibility that the distribution is multi-modal. It is a bit unintuitive that the distribution would be multi-modal: one might expect a single relative hand-object configuration to cause the same mechanical reaction forces and result in the same vector of sensor measurements in all cases. However, the nature of deformable materials makes it difficult to obtain exactly the same results each time. Furthermore, one might imagine that the noise is not necessarily Gaussian because a stochastic deformation event could result in two different interaction modes. Moreover, the difficulty involved in controlling the data collection scenario precisely suggests that non-Gaussian variation could result simply from material slippage during training. A look at Figure 6 bears out this intuition. The plot shows measurements associated with the k nearest neighbors in state space ($k = 8$ in the figure) along a sample trajectory. Notice that between 21 and 21.8, there are three possible sensor measurement trajectories supported by the data. Clearly, the Gaussian model is not a good choice for this type of data. As a result, we consider a sum-of-Gaussians model.

The sum-of-Gaussians model calculates the likelihood distribution by marginalizing over the space of true state-observation pairs:

$$P(z_i|x_i) = \eta \sum_{(x,z)\in X\times Z} P(z_i, x_i|z, x),$$

where $\eta = \frac{1}{P(x_i)}$ is assumed to be constant at each time step, reflecting a uniform prior belief. X and Z denote the state space and measurement spaces, respectively. Approximating $X \times Z$ by $\mathcal{D}$, we have:

$$P(z_i|x_i) = \eta \sum_{(x,z)\in\mathcal{D}} P(z_i, x_i|z, x).$$

Assuming that the joint likelihood of state and measurements, $P(z_i, x_i|z, x)$, is corrupted by Gaussian noise, we have:

$$P(z_i|x_i) = \eta \sum_{(x,z)\in\mathcal{D}} \mathcal{N}\left[\begin{array}{c|c} x_i & x \\ z_i & z \end{array}, \begin{pmatrix} \Sigma_x & \mathbf{0} \\ \mathbf{0} & \Sigma_z \end{pmatrix}\right], \quad (4)$$

where Σ_x and Σ_z are assumed to be diagonal covariance matrices that are specified by hand.

The results of tracking grommet location using the sum-of-Gaussians model in conjunction with the particle filter is illustrated by the solid line in Figure 5. These experiments used position and measurement covariances of $\Sigma_x = diag(0.0075)$ and $\Sigma_x = diag(0.001)$ inches. This data is averaged over the same 18 test sweeps that were used to produce the Gaussian measurement model results. Notice that in position 21.2, when the fingers are just beginning to touch the bump and the sensor measurements are most prominent, there is little advantage to using the sum-of-Gaussians model over a Gaussian model. However, after the initial period of prominent measurements, localization error accumulates less quickly in the sum-of-Gaussians model. This suggests that there is subtle information in the measurements after the fingers touch the bump initially that can improve localization. However, it seems that the single Gaussian model does not capture this information as well as the sum-of-Gaussians model. This finding supports the intuition suggested in Figure 6. After touching the bump for the first time, the sensor is likely to obtain measurements drawn from a multi-modal distribution that are best modeled using the sum-of-Gaussians.

D. Modeling "featureless" regions separately

In order to use the localization results described thus far, it is important to be able to determine when localization has occurred. In the controlled experiments that have been presented, localization occurs when the finger just begins touching the bump. Of course, in an application scenario, this information is unknown and must be inferred. Ways of predicting localization in a particle filter scenario include calculating the marginal posterior likelihood of the latest measurements (sum of the particle weights) or calculating the sample covariance of the particle distribution. When the marginal measurement likelihood is high or the particle covariance is low, it is reasonable to assume localization convergence. Unfortunately, we have found that on some occasions, both of the above measures falsely signal localization convergence before the hand actually touches the bump. This is a result of the observed measurements having a low likelihood everywhere in the state space. In this situation, even unlikely states can appear to be likely enough to cause the particle filter to converge.

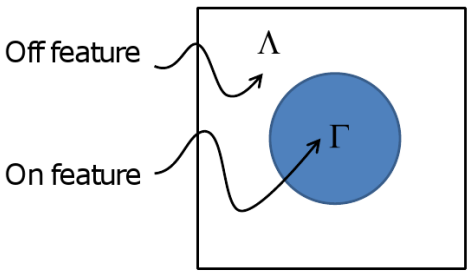


Fig. 7. Illustration of the on-feature and off-feature regions of state space. Measurements corresponding to points in Λ are combined into a single estimate of off-feature measurements.

Fundamentally, this problem is caused by overfitting the model to the training data. In principle, all data collected when the fingers are not interacting with a haptic feature should be identical. However, in fact there are always small variations in the training data. Given our prior knowledge that the tactile data collected in regions not near the relevant feature should be identical, we impose this constraint on the data by adjusting the model. In particular, we define two regions of state space (shown in Figure 7): the region $\Lambda \subseteq X$ known *not* to be touching the feature and the region $\Gamma \subseteq X$ known to contain the feature. Measurements corresponding to the off-feature states in Λ are modeled together while measurement corresponding to on-feature states in Γ are modeled as before.

For example, consider the case of two tactile sensors with positions a and b and corresponding measurement vectors z^a and z^b such that $z = \{z^a, z^b\}$. Whereas in earlier sections, the measurement likelihood was conditioned on the palm position, now marginalize over the two sensor positions:

$$P(z_i^a, z_i^b|x) = \sum_{a,b} P(z_i^a|a)P(z_i^b|b)P(a, b|x). \quad (5)$$

Define functions, $A(x)$ and $B(x)$, that evaluate to the position of sensors a and b, respectively, when the palm is at x. Approximate $P(a, b|x)$ to be 1 when $a \in A(x)$ and $b \in B(x)$ and zero otherwise. Then, Equation 5 becomes:

$$P(z_i^a, z_i^b|x) = \sum_{(a,b)\in A(x)\times B(x)} P(z_i^a|a)P(z_i^b|b). \quad (6)$$

If a is in Γ, then we estimate $P(z_i^a|a)$ the same as before:

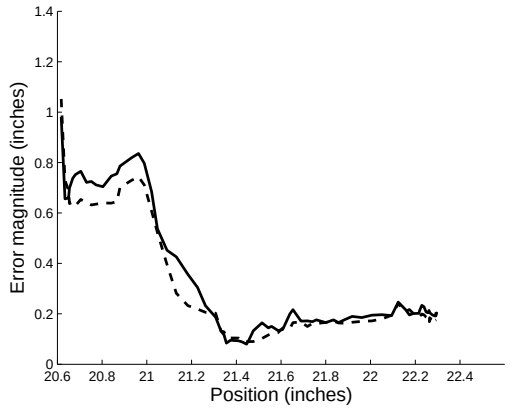
$$P(z_i^a|a) = \eta \sum_{(x,z)\in\mathcal{D}} \mathcal{N}\left[\begin{array}{c|c} a & A(x) \\ z_i & z \end{array}, \begin{pmatrix} \Sigma_a & \mathbf{0} \\ \mathbf{0} & \Sigma_z \end{pmatrix}\right].$$

If a is in Λ, then we estimate:

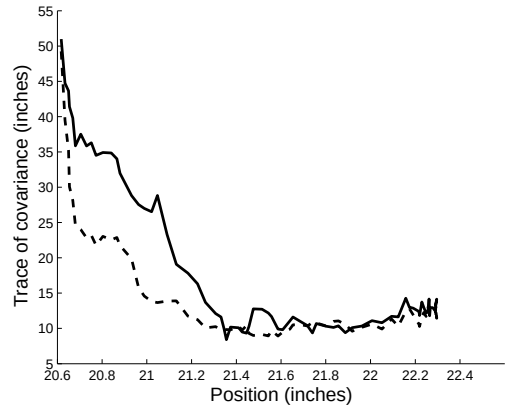
$$P(z_i^a|a) = \mathcal{N}(z_i^a|\hat{z}_{off}, \Sigma_{off}), \quad (7)$$

where $\hat{z}_{off}$ and Σ_{off} are the sample mean and covariance taken over all points in Λ.

Figure 8 illustrates the results of modeling featureless regions separately. In this experiment, training measurements



(a)



(b)

Fig. 8. Performance of modeling off-feature regions separately compared to the sum-of-Gaussians approach of Section III-C. The solid lines show the performance of the featureless averaging method. The dashed lines show the sum-of-Gaussians performance from Section III-C for comparison. (a) shows that localization error is comparable for both approaches. (b) shows that the particle distribution has higher entropy (more spread out) before contacting the bump when modeling the off-feature region separately.

taken when the fingers were in front of the bump (in this experiment, roughly when the palm is at a position less than 21.2 inches) were combined and used to calculate the sample mean and covariance in Equation 7. The solid line in Figure 8(a) shows the error for this approach averaged over the 18 trials (same as in earlier experiments). The dashed line repeats the sum-of-Gaussians error from Figure 5 for comparison. Notice that localization performance is about the same even though we have smoothed out all off-feature variation. Figure 8(b) shows the trace of the particle covariance (sum of the eigenvalues) averaged over the 18 trials. This essentially measures how peaked the particle distribution is. Notice that smoothing over the featureless region of state space has increased the eigenvalues of the covariance matrix for those points during localization prior to touching the bump. Essentially, this improves our ability to measure the convergence event without harming localization accuracy.

IV. Conclusion

This paper considers the problem of localizing objects that a robot touches based on tactile data. In particular, we focus on the problem of locating haptic features in flexible materials such as grommets in fabric or bumps in flexible plastic. Rather than estimating the measurement likelihoods based on a geometric model of the interaction, this paper introduces the idea of modeling the state-measurement relationship directly from training data. This modeled state-measurement relationship constitutes a "tactile map" that defines what the robot "feels" as its sensors move over the haptic feature. Several variations on a basic measurement model are proposed are their relative merits evaluated. Our evaluation of this approach suggests that it effectively tracks manipulation state during manipulation interactions. The experiments show repeatable average localization errors of approximately one tenth of an inch using tactile sensors much larger in size (the sensors are approximately the same size as human fingertips.) Our success with localization in a low-dimensional space based on relatively simple state-measurement models suggests that the same type of approach could be applied to higher dimensional state spaces given more structured measurement models.

References

[1] C. Corcoran and R. Platt, "A measurement model for tracking hand-object state during dexterous manipulation," in *IEEE Int'l Conf. on Robotics and Automation*, 2010.

[2] K. Gadeyne and H. Bruyninckx, "Markov techniques for object localization with force-controlled robots," in *10th Int'l Conf. on Advanced Robotics*, 2001.

[3] A. Petrovskaya, O. Khatib, S. Thrun, and A. Ng, "Bayesian estimation for autonomous object manipulation based on tactile sensors," in *IEEE Int'l Conf. on Robotics and Automation*, 2006, pp. 707–714.

[4] S. Chhatpar and M. Branicky, "Particle filtering for localization in robotic assemblies with position uncertainty," in *IEEE Int'l Conf. on Intelligent Robots and Systems*, 2005, pp. 3610– 3617.

[5] R. Platt, M. Abdallah, and C. Wampler, "Joint-space torque and stiffness control of tendon-driven manipulators," in *IEEE Int'l Conf. on Intelligent Robots and Systems (submitted)*, 2010.

[6] C. Ihrke, M. Diftler, M. Linn, R. Platt, and B. Griffith, "Phalange tactile load cell," US Patent Application 20 100 077 867, 2008.

[7] S. Arulampalam, S. Maskell, G. N., and T. Clapp, "A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking," *IEEE Transactions on Signal Processing*, vol. 50, pp. 174–188, 2001.