Reactive Grasping Using Optical Proximity Sensors

Kaijen Hsiao, Paul Nangeroni, Manfred Huber, Ashutosh Saxena, Andrew Y Ng

Abstract— **We propose a system for improving grasping using fingertip optical proximity sensors that allows us to perform online grasp adjustments to an initial grasp point without requiring premature object contact or regrasping strategies. We present novel optical proximity sensors that fit inside the fingertips of a Barrett Hand, and demonstrate their use alongside a probabilistic model for robustly combining sensor readings and a hierarchical reactive controller for improving grasps online. This system can be used to complement existing grasp planning algorithms, or be used in more interactive settings where a human indicates the location of objects. Finally, we perform a series of experiments using a Barrett hand equipped with our sensors to grasp a variety of common objects with mixed geometries and surface textures.**

I. Introduction

Grasping is a basic and important problem in robotic manipulation. For robots to reliably grasp novel objects, they must be able to sense the object geometry sufficiently accurately to choose a good grasp. In this paper, we develop an optical proximity sensor, embedded in the fingers of the robot (see Fig. 1), and show how it can be used to estimate local object geometries and perform better reactive grasps.

Conventional grasp planning strategies rely heavily on long range vision sensors (such as cameras, LIDAR, and IR range finders) to detect and model objects, and to determine grasp configurations (e.g., [11], [15], [17]). While this has led to a number of successful robot systems, errors and uncertainty ranging from small deviations in the object's location to occluded surfaces have significantly limited the reliability of these open-loop grasping strategies. Indeed, in [15], we found that approximately 65% of the grasp failures were because we used only long range sensors and lacked a reactive controller with sufficient local surface pose information.

Tactile sensing has been employed as a means to augment the initial grasp and manipulation strategies by addressing inconsistencies in the contact forces during object contact and manipulation [20]. However, tactile sensors have to actually touch the object in order to provide useful information. Because current sensor technologies are not sensitive enough to detect finger contacts before causing significant object motion, their use is limited to either minor adjustments of contact forces at pre-computed grasp configurations, or to planning algorithms that require iterative re-grasping of the object in order to grasp successfully [8], [19], [3]. While the latter approach has shown substantial improvements in grasp reliability, it requires a significant amount of time and frequently causes lighter objects to be knocked over during the repeated grasp attempts.

Fig. 1: Three-fingered Barrett Hand with our optical proximity sensors mounted on the finger tips.

The limitations of tactile-sensing-augmented grasp planning can be overcome by 'pre-touch' sensing. This modality has recently become a popular means of bridging the gap in performance between long range vision and tactile sensors. In pre-touch sensing, gripper-mounted, short-range (0-4cm) proximity sensors are used to estimate the absolute distance and orientation (collectively called surface pose) of a desired contact location without requiring the robot to touch the object [18]. The vast majority of these pre-touch proximity sensors use optical methods because of their high precision [1], [7], [21], [9], [10]. Optical sensors, however, are highly sensitive to surface reflection properties. Alternatively, capacitive-based proximity sensors have also been used [22]. While invariant to surface properties, these capacitive-based sensors have difficulty detecting materials with low dielectric contrast, such as fabrics and thin plastics. Unfortunately, in both cases, present sensor calibration and modeling techniques have yet to produce pose estimates that are robust enough to be useful across the range of surface textures, materials, and geometries encountered in unstructured environments. Furthermore, the finger tips of typical robotic grippers are too small to accommodate the

K. Hsiao is with the Computer Science and Artificial Intelligence Lab at the Massachusetts Institute of Technology, Cambridge, MA 02139, USA kjhsiao@mit.edu. This work was done while Kaijen was visiting Stanford University.

P. Nangeroni is with the Department of Mechanical Engineering at Stanford University, Stanford, CA 94305, USA pauln1@stanford.edu

M. Huber is with the Department of Computer Science and Engineering at the University of Texas at Arlington, Arlington, TX 76019, USA huber@cse.uta.edu

A. Saxena and A.Y. Ng are with the Department of Computer Science at Stanford University, Stanford, CA 94305, USA {asaxena,ang}@cs.stanford.edu

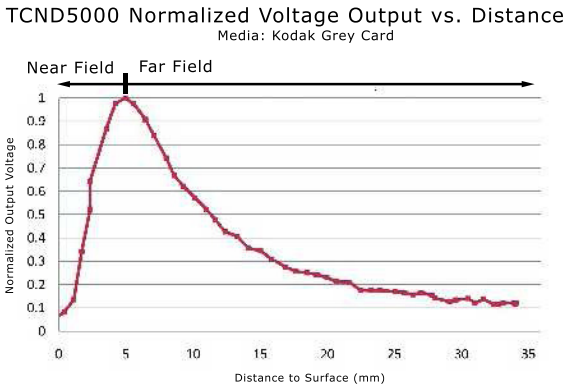

Fig. 2: Normalized Voltage Output vs. Distance for a TCND5000 emitter-detector pair.

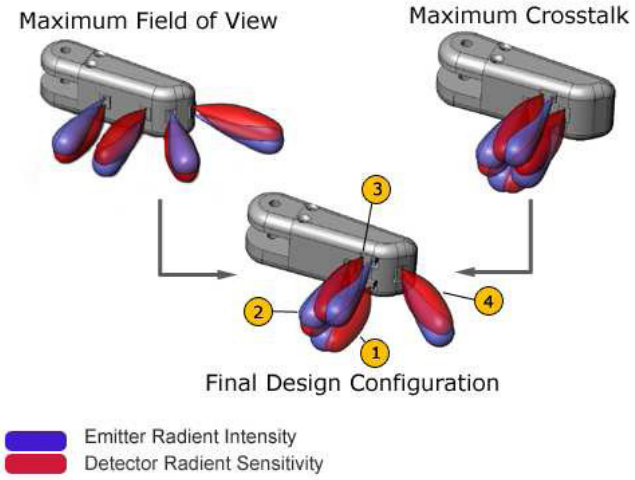

Fig. 3: Possible fingertip sensor configurations.

sensors used in previous work.

This paper presents an integrated approach that combines sensor design, probabilistic data modeling, and a reactive controller into a system that allows for on-line grasp adjustments to an initial grasp configuration without the need for premature object contact or re-grasping strategies. Specifically, a design for a low cost, pose-estimating proximity sensor is presented that meets the small form factor constraints of a typical robotic gripper.

The data from these sensors is interpreted using empirically derived models and a robust, belief-state-based surface pose estimation algorithm. The resulting pose estimates are provided to a reactive grasp closure controller that regulates contact distances even in the absence of reliable surface estimates. This allows the robot to move the finger tip sensors safely into configurations where sensor data for the belief-state update can be gathered. Simultaneously, the pose estimates of the sensor interpretation approach can provide the necessary information to adjust the grasp configuration to match the orientation of the object surface, increasing the likelihood of a stable grasp.

We perform a series of grasping experiments to validate the system using a dexterous robot consisting of a 7-DOF Barrett arm and multi-fingered hand. In these tests, we assume that an approximate location of the object to be grasped has already been determined either from long range vision sensors [14], or through human interaction [5]. From the initial grasp positions, the system exhibits improved grasping on a variety of household objects with varying materials, surface properties, and geometries.

II. Optical Sensor Hardware

The purpose of the optical sensor hardware is to provide data that can be used by the modeling algorithm to construct pose estimates of nearby surfaces. The design is driven by a series of constraints, including size, sensor response, and field of view, which are detailed in the following section.

A basic optical sensor consists of an emitter, photo-receiver, and signal processing circuitry. The light from the emitter is reflected by nearby surfaces and received by the photo-receiver. The amplitude and phase of the light vary as a function of the distance to the surface, its orientation, and other properties of the surface material (reflectance, texture, etc.) [1]. In amplitude-modulated proximity sensor design, the most commonly preferred method, these variations in amplitude can be converted into pose estimates by measuring the response from constellations of at least three receivers focused at the same point on the target surface [1], [12], [21].

Although conceptually simple, modeling the pose of unknown surfaces is difficult because of the non-monotonic behavior of the proximity sensor receivers. The response of a single sensor is a function of the distance to the target surface and the baseline between the emitter and receiver, as shown in Fig. 2. While the response in the far-field varies as the inverse square of the distance [6], the response of the sensor in the near field is far more complex. The decrease in received light energy in the near field is governed not only by the reflective properties of the surface, but also by the geometric baseline between the emitter and receiver. As distance approaches zero in the near field, the amount of energy that can be reflected between the emitter and receiver decreases because the overlap between the emitter and receiver cones decreases. This results in a sharp drop-off of received light intensity. To avoid complications in modeling the sensor data, most approaches, including ours, offset the sensors from the surface of the gripper to maximize the far field sensor response. Although this sacrifices the higher sensitivity of the near field, the longer range of the far field is better suited to grasping applications [1].

The selection of specific sensor hardware and signal processing electronics is severely constrained by the limited space in typical robotic fingers (for instance, our Barrett fingers have a 1cm x 2cm x 1cm cavity volume). Bonen and Walker used optical fibers to achieve the desired geometric arrangement of emitters and receivers in their respective sensors. However, the bend radius and large terminations of

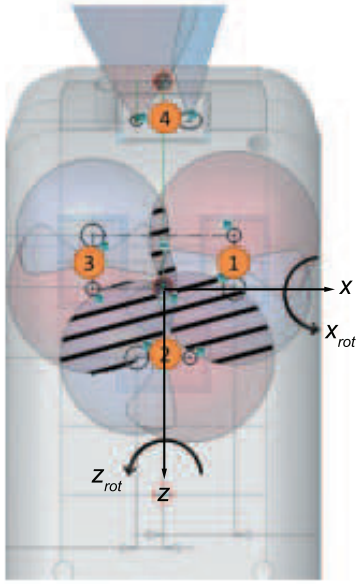

Fig. 4: Front view of sensor constellation. Hatches show the crosstalk between adjacent sensor pairs.

the fibers violate the space constraints in this application. Instead, our design uses four low-cost, off-the-shelf infrared emitter/receiver pairs (Vishay TCND50000) on each finger. The small size (6 x 3.7 x 3.7 mm) meets the volume requirements of the design and the range (2-40mm) is ideal for pre-touch sensing.

The arrangement of these sensors represents yet another design tradeoff between field of view and pose estimation accuracy. Large fields of view, both out from and in front of the face of the fingertip, as shown in Fig. 3a, are advantageous for detecting oncoming objects. Unfortunately, the sensor spacing needed to achieve this reduces the overlap between adjacent sensor pairs and lowers the signal to noise ratio. Conversely, arranging the sensors to focus the emitters and maximize crosstalk, as shown in Fig. 3b, improves local pose estimation accuracy at the expense of broader range data to nearby objects. The final configuration, shown in Fig. 3c, consists of three sensors arranged in a triangle on the face of the finger to estimate pose with a fourth sensor at a 45° angle to the finger tip to increase the field of view. Although surface area exists for placing additional sensors, the quantity is limited to four by the available space in the Barrett finger cavity for pre-amplifier circuitry. The sensors are inset into the finger to minimize near-field effects, and the aluminum housing is matte anodized to decrease internal specular reflection. The crosstalk between adjacent sensors is illustrated by the hatched area in Fig. 4.[1]

[1]While the crosstalk between sensors 1-3 can be useful when the primary sensor values saturate, which occurs on many light-colored surfaces, experimental testing showed the crosstalk to be below the noise floor on many of the object surfaces we encountered. Since this work assumes that the nature of the surface is unknown a priori, our current work ignores the crosstalk and only uses the primary sensor values.

The complete proximity sensing suite consists of twelve total emitter-detector pairs, four on each finger. The emitters are pulsed in sequence by a PIC 18F4550 micro-controller located on the wrist of the robot (as shown in Fig. 1) so that 16 readings are generated for each finger on each cycle. The collected sensor data is pre-amplified by circuitry in the finger tip and then sampled by a 10-bit A/D converter before streaming the data back to the robot over a serial link. In spite of the 950nm operating frequency, raw sensor readings remain sensitive to ambient light effects. Background subtraction was used to remove this ambient component and increase the signal to noise ratio.

III. Sensor Model

Given a series of observations from the sensors on each finger, our goal is to estimate the surface pose (distance and orientation) of the surface in front of the finger.

More formally, let $o = (o_1, o_2, o_3) \in \mathbb{R}^3$ be the readings from the three sensors grouped in a triangle,[2] and $s = (d, x_{rot}, z_{rot})$ be the surface pose of the local surface (approximated as a plane). Here, d is the straight-line distance from the center of the three sensors to the surface of the object (along the vector pointing outwards from the finger surface) sensors. x_{rot} and z_{rot} are the relative orientations of the surface around the finger's x-axis (pitch) and z-axis (roll) respectively, as shown in Fig. 4. (They are equal to 90° when the object surface is parallel to the finger surface.)

One of the major challenges in the use of optical sensors is that intrinsic surface properties (such as reflectivity, diffusivity, etc.) cause the relationship between the raw sensor signal and the surface pose to vary significantly across different surface types. For that reason, prior work using short-range proximity sensors to find surface pose has focused on using multiple direct models obtained by performing regression on empirical data. This data is gathered by recording sensor array readings for each relevant surface in turn, placed at a comprehensive set of known surface poses [1]. In particular, [1], [7], [21] both use a least-squares polynomial fit of data taken for a known surface or group of similar surfaces to directly estimate surface pose given sensor readings. However, acquiring enough data to successfully model a new surface is extremely time-consuming, and having to do so for every potential surface that might be encountered is prohibitive. In practical grasping scenarios, we need to be able to deal with unknown and never-before-encountered surfaces.

A. Calibration Data and Reference Forward Model

As opposed to fully characterizing every surface with a separate model, we use a single reference model that is scaled with an estimate of the object's IR reflectivity parameter in

[2]Sensor 4, which is offset by 45° to increase the field of view, is not included in this sensor model because it rarely focuses on the same flat surface as the other three. As a stand-alone sensor, it is nevertheless independently useful, particularly when the peak expected value (object surface reflectivity) is known. For instance, it can be used to prevent unexpected collisions with objects while moving, or even to move to a fixed distance from a table or other surface with known orientation.

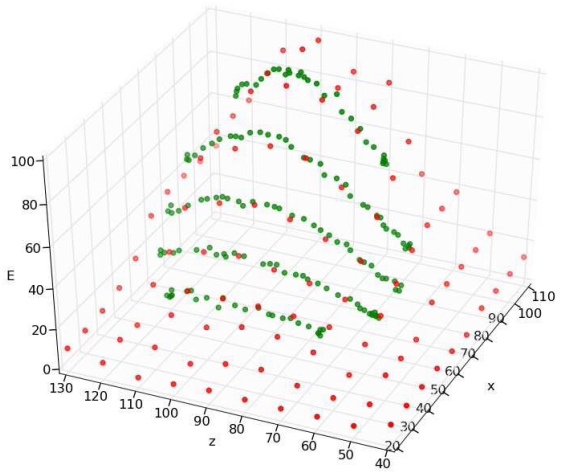

Fig. 6: Locally weighted linear regression on calibration data. The plot shows energy values for one sensor at a fixed distance and varying x and z orientation. Green points are recorded data; red points show the interpolated and extrapolated estimates of the model.

order to obtain an approximate forward model for the specific object of interest. [3] By scaling with the estimated peak value, the model becomes roughly invariant to surface brightness.

For this work, the calibration data was taken using a Kodak grey card, which is a surface often used to adjust white balance in photography. The data consists of 1904 samples of o taken at distance values ranging from 0.5 cm to 3.4 cm, x_{rot} values ranging from 30° to 110°, and z_{rot} values ranging from 40° to 140°. Fig. 5 shows the calibration data.

B. Direct Model Using Polynomial Fit

We first tried using a least-squares polynomial fit on the calibration data to estimate s given o, as in [1]. While the resulting polynomial function was a reasonable fit to the calibration data, its performance on estimating s on different surfaces was extremely poor. This is because even small changes in the true underlying function (caused by different brightness or reflection properties) results in large shifts in estimates for s for the same o, making the direct model very brittle in the context of different, unknown objects, and limiting its applicability to contexts where the *precise* object characteristics are known a priori.

C. Belief State Model

In order to address the sensitivity of the direct model to the object surface characteristics, we use an empirical forward sensor model and a probabilistic estimation process to derive the current best estimate for s. In our setting, as our robot hand is executing a grasp trajectory, we take many sensor readings with fingers at different (known) locations, and use the sensor readings, the finger positions, and our empirical sensor model to update a belief state (a probability distribution over possible values of s) at each time step.

Here, we assume that the object is static during a grasp. This is a reasonable assumption due to the fact that because we are using proximity or 'pre-touch' sensors, the actual grasp and sensing procedure does not make contact with the object prior to complete closure of the grasp and thus does not actively cause any object displacements. In addition, we assume that the surface seen under each finger is locally planar throughout the entire grasp process.[4]

For each finger, let S be the set of all possible states s, and let s_t, o_t and a_t be the state, sensor readings (observation), and hand pose (actions) at time t, respectively. At all times, we will track the belief state $b_t := P(s_0|o_1...o_t, a_1...a_t)$, the probability distribution over all possible surface poses in S seen at the first time step, given all observations and hand poses since. We assume that our sensor observations at different times are conditionally independent given the surface pose:

$$P(s_0|o_1...o_T, a_1...a_T) = \prod_{t=1}^{T} P(s_0|o_t, a_t) \qquad (1)$$

D. Observation Model

An essential part of finding the most likely surface pose is having a model for how likely it is that we could have seen the current sensor readings if that pose were true. Specifically, the quantity we are looking for is $P(o_t|a_t, s_t)$, the probability of seeing the sensor readings we obtained at time t (o_t) given the current hand configuration (a_t) and a particular surface pose (s_t). For this, we need a function mapping states to observations, $C(s) = o$. We use locally weighted linear regression on our scaled calibration data set to estimate o given s. An example of the values obtained using locally weighted linear regression is shown in Fig. 6, where the green points are actual data and the red points are estimated values. Each estimated point uses a plane derived from only the 8 closest actual data points. Also, because extrapolation using locally weighted linear regression is extremely poor, the estimated point is clipped to be no greater than the highest of those 8 values, and no less than the lowest.

We then assume that the estimated model values are correct (for a given s_t, we expect to see the model-estimated o_t), and that any deviations we see in the actual sensor readings, $\epsilon = o_t - C(s_t)$, are due to Gaussian noise, $\epsilon \sim N(0, \sigma^2)$. This assumption is wildly inaccurate, as the errors are in fact systematic, but this assumption nonetheless allows one to find the closest alignment of the observed o points to the scaled calibration model without any assumptions about surface characteristics. For our experiments, sensor readings

[3] Estimation of the surface reflectivity value is not particularly onerous to collect, as a single grasp of the object, particularly with the use of raw values to attempt to align at least one finger with the surface of the object, is sufficient to collect the required value. Alternatively, an estimate of the peak value could be obtained by observing the object with an IR camera or with a laser rangefinder that provides IR intensity values.

[4] Even when the surface under the finger changes and the old data becomes erroneous, the estimates would get progressively better as more data from the new surface is observed. It is also simple to place a higher weight on new data, to make the adjustment faster.

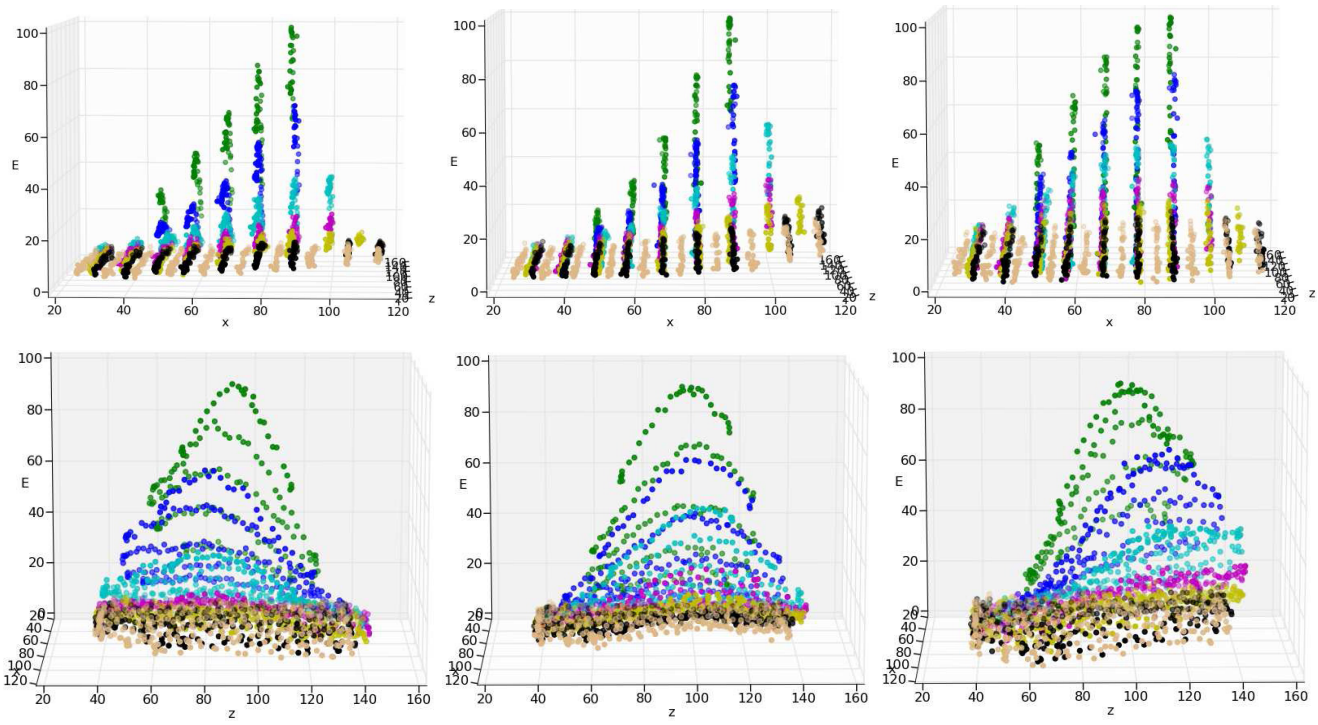

Fig. 5: Calibration Data. (Top row) plots xrot vs. energy for all three sensors, with binned distances represented by different colors (distance bin centers in mm: green=5, blue=9, cyan=13, magenta=19, yellow=24, black=29, burlywood=34). (Bottom row) shows zrot vs. energy for all three sensors.

were scaled to vary from 1 to 100, and σ was set to be 25, since we expect significant deviations from our model values.

E. Inference

At each time step, we wish to compute an estimate $\hat{s}_t$ of the current state. The states, S, are discretized to a uniform grid. We can represent b_t as a grid of probabilities that sum to 1, and update the belief state probabilities at each time step using our actions and observations as we would any Bayesian filter.

More specifically, the belief state update is performed as follows:

$$\begin{aligned} b_t =& P(s_0|o_1...o_t, a_1...a_t) \\ =& \frac{P(o_t|a_0, a_t, s_0)P(s_0|o_1...o_{t-1}, a_1...a_{t-1})}{P(o_t)} \\ =& \frac{P(o_t|a_t, s_t)b_{t-1}}{P(o_t)} \end{aligned} \tag{2}$$

We assume a uniform prior on the sensor readings, and thus the denominator can be normalized out. We then find the expected value of the state s_0 as follows:

$$\hat{s}_0 = E(s_0) = \sum_{s_0} P(s_0|o_1...o_t, a_1...a_t)s_0 \tag{3}$$

We can compute $\hat{s}_t$ from $\hat{s}_0$ using the hand kinematics and the known hand positions a_t and a_0.

The advantage of combining observations in this manner is that, while the actual sensor readings we expect to see vary greatly from surface to surface, readings across different sensors in the array vary in a similar way with respect to orientation changes for all surfaces. Thus, a number of observations over a wide range of different values of s can be used to align a new set of observations with a model that is not quite right. The s with expected (model) observations that align best with the actual observations is generally the one with the highest likelihood, even with large model error.

IV. Reactive Grasp Closure Controller

To verify the usefulness of our 'pre-grasp' proximity sensors, a reactive grasp closure controller was designed and implemented that uses the proximity data to move the fingers such that: 1) they do not collide with the object, and 2) they close simultaneously, forming a symmetric grasp configuration around a pre-specified grasp point. This ability is aimed at achieving successful grasps even in the presence of significant uncertainty about the object geometry, and at allowing the sensor interpretation approach to safely collect data to improve the surface estimates.

The hierarchical control architecture used here composes two reactive control elements that run asynchronously and that control separate aspects of the grasp closure process. At the bottom level, a finger distance controller controls the fingers to maintain approximately equal distances to the object surface. On top of this control element, a kinematic conditioning controller controls the arm to center the object within the hand and to cause the fingers to be parallel to the surface.

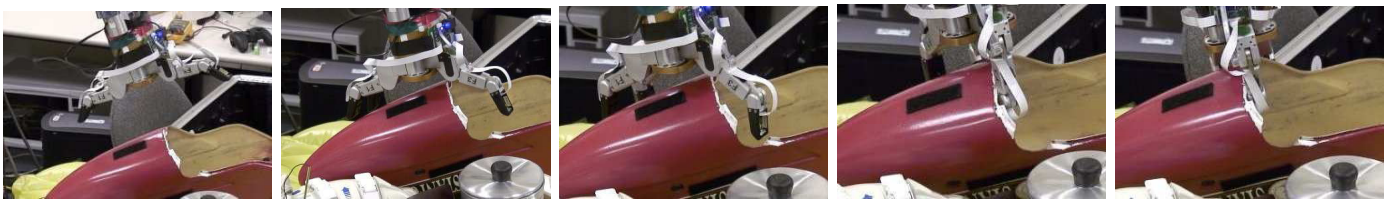

Fig. 7: A sequence showing the grasp trajectory chosen by our algorithm. Initially, the fingers are completely open; as more data comes, the estimates get better, and the hand turns and closes the finger in such a way that all the fingers touch the object at the same time.

A. Finger Distance Controller

At the bottom level, the finger distance controller moves just the fingers in an attempt to keep all fingers at a given distance from the object surface while closing them around the initial grasp pose. To avoid collisions even in the initial stages of grasping, the finger distance controller starts out using just raw proximity sensor values normalized for the estimated object reflectivity (expected peak value). When available, it switches to using distance information from the belief estimation process. Either raw values or distance estimates can be used to achieve and maintain a desired proximity from the surface for all fingers, at the rate required by the basic hand controller. This allows for the efficient gathering of proximity sensor readings during the grasp, which can be used to improve the surface pose belief state, which in turn can improve the finger distance controller's ability to keep the fingers at a desired distance. Steadily decreasing the desired surface distance allows one to close the fingers such that they simultaneously touch the object surface.

B. Kinematic Configuration Controller

Given the current grasp pose and finger configuration, the kinematic configuration controller moves the arm and hand to optimize the available kinematic workspace and contact force capabilities of the grasping system. To achieve this, it selects actions that locally minimize a kinematic configuration error function, $\epsilon_K = \sqrt{\sum_i \epsilon_{K_i}^2}$, by descending the gradient, $\frac{\partial \epsilon_K}{\partial x_j}$ of this metric with respect to the hand frame pose parameters x_j that determine the position and orientation of the hand. ϵ_{K_i} is the kinematic error for finger i, which is a measure of the distance from the preferred kinematic configuration (which in our case is the middle of the finger joint range).

For the Barrett hand, the result of descending this gradient is a hand/arm configuration in which all fingers are bent to the same degree and the hand is centered over the current contact configuration as determined by the finger distance controller. When used in combination with the finger distance controller, the result is that the two fingers on the same side of the hand end up parallel to the surface under them (assuming they see the same or a similar surface), and the hand ends up centered around the object. For a completely novel object, these two controllers together can be used to perform an initial grasp to collect data identifying the object's reflectivity parameters, since the peak sensor value is only seen when the finger is parallel to the object surface.

TABLE I: Model Test Experiment Error

	DIST(CM)	$x_{rot}(^\circ)$	$z_{rot}(^\circ)$
BROWN WOOD BOWL	.35	4.4	13.5
BEIGE PLASTIC BOWL	.34	4.4	23.0
BLACK PLASTIC BOX	.49	10.3	16.5
BLUE PLASTIC BOX	.59	5.3	21.7
CARDBOARD BOX	.20	3.8	14.4
YELLOW BOX	.43	3.6	17.3
AVERAGE	.40	5.3	17.7

Given a correctly chosen initial grasp point, this two-element controller hierarchy provides robustness to the grasping process and can significantly increase the success rate for grasping tasks. However, it still relies heavily on a good choice of initial grasp pose. To alleviate this and to fully utilize the capabilities of the sensors, the presented reactive grasp closure hierarchy could be augmented with a reactive grasp configuration controller that uses the surface normal estimates provided by the sensors' belief state estimation system. Having normal estimates would allow such a controller to adjust the grasp configuration to optimize the local grasp geometry. In particular, one could add a reactive local control component based on wrench residuals [2] to the grasp closure controller using a hierarchical control composition approach [4]. Such a component would use the local surface normal information to optimize the contact configuration, allowing one to achieve reliable force and moment closure grasps of objects even when no two fingers see the same planar face [13], [2].

V. EXPERIMENTS

We will first evaluate the ability of our algorithm to model sensor data in Section V-B. Then we will use the model to perform grasping experiments in Section V-C.

A. Robot Hand

The sensors described in Section II were mounted onto the finger tips of a Barrett hand. This manipulator has a total of 4 degrees of freedom (three fingers each with one degree of freedom, and a finger "spread"). The hand was attached to a 7-dof arm (WAM, by Barrett Technologies) mounted on the STAIR (STanford AI Robot) platform.

B. Prediction Model

As a test of our estimation method, we placed the Barrett hand at known positions relative to a number of commonly

Fig. 8: Different objects on which we present our analysis of the model predictions. See Section V-B for details.

found objects of various materials. (See Fig. 8 for their pictures.) Our test data consisted of three different orientations for each object: parallel to the fingers, $+10°$ rotation, and $-20°$ rotation. The fingers were then made to close slowly around the object until each finger nearly touched the surface. Sets of 100 raw sensor readings were taken every 1.2 degrees of finger base joint bend, and for every 1.2 degrees of finger spread (up to at most 17 degrees) and averaged. Results showing errors in the final estimates of (d, x_{rot}, z_{rot}) are in Table I.

Table I shows that our model is able to predict the distances with an average error of 0.4cm. We are also able to estimate x_{rot} reasonably accurately, with an average error of $5.3°$. The high error in x_{rot} in the case of the black plastic box can be attributed to the fact that the first finger in one run did not see the same surface the entire time, which is a requirement for reasonable predictions with our model. Note that these objects have significantly different surface properties, and other than the peak sensor value, no other surface characteristics were assumed. High-reflectance surfaces tended to do worse than matte surfaces due to significant deviation from the calibration model. Nonetheless, our experiments show that we are able to make reasonable predictions by the time the fingers touch the surface.

The higher errors in z_{rot} are a consequence of the movements used during the data collection and estimation process. Since the sensor interpretation approach uses belief updates to determine the maximum likelihood orientation, the quality of the resulting estimates depends on the proximity data encountered along the fingers' trajectories, with larger variations in the local geometry resulting in more accurate estimates. In the procedure used here to gather the test data, a wide range of x_{rot} angles was encountered within the useful range of the sensors due to the strong curling of the Barrett fingers. On the other hand, only a very limited range of z_{rot} angles were correctly observed due to the significantly smaller variation available using the hand's spread angle. Furthermore, most spread actions moved the fingers to positions significantly further from the object, often resulting in sensor readings that were no longer observable. While this is a problem in the test data and illustrates the advantage of active sensing strategies, its cause should be largely alleviated when using the grasp closure controller to establish the initial grasp, due to the ability of the finger distance controller to maintain the fingers at a distance that provides usable results throughout the finger spread operation. Additionally, the inclusion of arm motions through the use of the kinematic conditioning controller should further enhance the range of z_{rot} angles encountered during a grasp, and thus allow for somewhat better z_{rot} estimates.

C. Grasping Experiments

Our goal was to focus on the final approach of grasping using proximity sensors. Our system could be used in a variety of settings, including the point-and-click approach of [5], where a laser pointer is used to highlight an object for a robot to pick it up, or combined with long range vision sensors that select optimal grasp points [14].

In this experiment, we placed a variety of objects (weighing less than 3 pounds) in known locations on a table (see Fig. 10), with some objects flush against each other. These objects are of a variety of shapes, ranging from simple boxes or bowls, to more complicated shapes such as a ski boot.

The robot moves the hand to the approximate center of the object and executes a grasp strategy,[5] using our controller to move the hand and the fingers in response to estimated distances from the sensors and the pose estimation algorithm. The robot then picks up the object and moves it to verify that the grasp is stable. (See Fig. 7 and Fig. 10 for some images of the robot picking up the objects.)

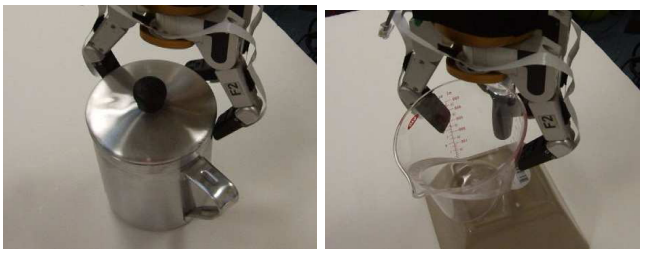

Fig. 9: Failure cases: (a) Shiny can, (b) Transparent cup

Note that our objects are of a variety of shapes and made of a variety of materials. Out of the 26 grasping trials performed on 21 unique objects,[6] our grasping system failed three times. Two of these failures were due to extreme surface types: a

[5]One could envision using better strategies, such as those based on vision-based learning [15], [16], for moving the hand to a grasping point on the object.

[6]*Procedure used to choose the objects*: We asked a person not associated with the project to bring objects larger than about 8 inches in length and less than 3 pounds from our lab and different offices. No other selection criterion was specified, and therefore we believe that our objects represented a reasonably unbiased sample of commonly found objects.

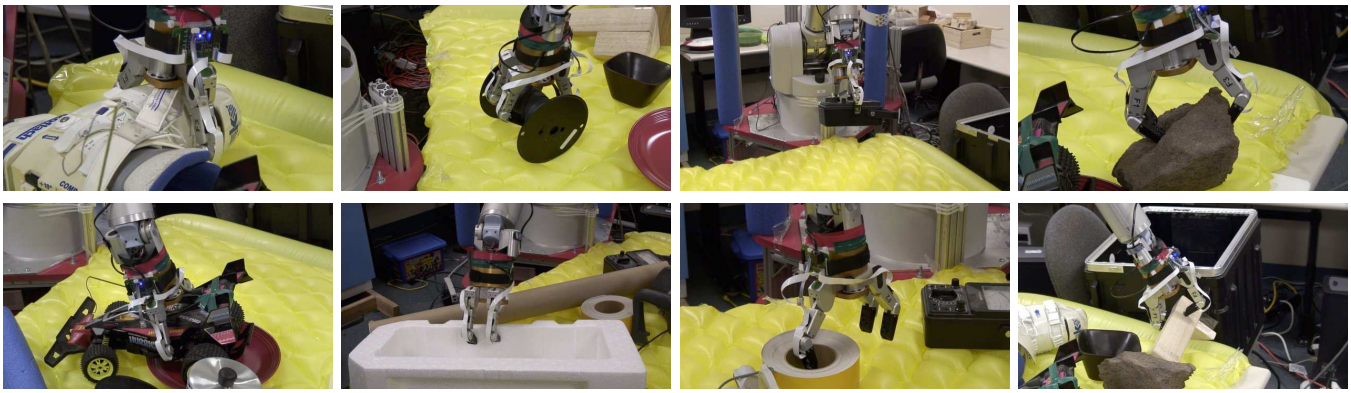

Fig. 10: Snapshots of the robot picking up different objects.

transparent cup and a highly reflective aluminum can (Fig. 9). To address these cases, our optical-based proximity sensors could be combined with a capacitive-based system that is good at detecting metallic or glass surfaces. The third object was a ski-boot, for which our approach worked perfectly; however, due to the object's weight, our robot was unable to lift it up.

A video of the grasp experiments is available at the following url:

http://stair.stanford.edu/proximitygrasping/

In the video, the hand rotates the fingers in many cases to be approximately parallel to the object surface and causes the fingers to contact at nearly the same time, thereby improving the grasp.

VI. Conclusion

In this paper, we presented a system that enables stable grasping of common objects using pre-touch pose estimation. We designed a novel low-cost optical proximity sensor to meet the space constraints of typical robot grippers using off-the-shelf components. We converted the data provided by these sensors into pose estimates of nearby surfaces by a probabilistic model that combines observations over a wide range of finger/hand configurations. We also created a hierarchical reactive controller to perform the grasp while optimizing finger locations to maintain distance and improve the estimation model. Finally, we validated the system by experiments with a Barrett hand, which showed improvements in reactive grasping.

Acknowledgements

We thank Kenneth Salisbury for helpful discussions. We also thank Aaron Rosekind and Hee-Tae Jung for help in the experiments. This work was supported in part by the DARPA transfer learning program under contract number FA8750-05-2-0249. Support from Willow Garage is also gratefully acknowledged.

References

[1] A. Bonen. Development of a robust electro-optical proximity-sensing system. *Ph.D. Dissertation, Univ. Toronto*, 1995.
[2] J. A. Coelho Jr. and R. A. Grupen. A control basis for learning multifingered grasps. *JRS*, 14(7):545–557, October 1997.
[3] K. Hsiao, T. Lozano-Perez, and L. P. Kaelbling. Robust belief-based execution of manipulation programs. In *WAFR*, 2008.
[4] M. Huber and R. A. Grupen. Robust finger gaits from closed-loop controllers. In *IROS*, 2002.
[5] C. C. Kemp, C. Anderson, H. Nguyen, A. Trevor, and Z. Xu. A point-and-click interface for the real world: Laser designation of objects for mobile manipulation. In *HRI*, 2008.
[6] O. Partaatmadja, B. Benhabib, E. Kaizerman, and M. Q. Dai. A two-dimensional orientation sensor. *Journal of Robotic Systems*, 1992.
[7] O. Partaatmadja, B. Benhabib, A. Sun, and A. A. Goldenberg. An electrooptical orientation sensor for robotics. *Robotics and Automation, IEEE Transactions on*, 8:111–119, 1992.
[8] A. Petrovskaya, O. Khatib, S. Thrun, and A. Y. Ng. Bayesian estimation for autonomous object manipulation based on tactile sensors. In *ICRA*, 2006.
[9] G. Petryk and M. Buehler. Dynamic object localization via a proximity sensor network. In *Int'l Conf Multisensor Fusion and Integration for Intelligent Systems*, 1996.
[10] G. Petryk and M. Buehler. Robust estimation of pre-contact object trajectories. In *Fifth Symposium on Robot Control*, 1997.
[11] J. H. Piater. Learning visual features to predict hand orientations. In *ICML Workshop on Machine Learning of Spatial Knowledge*, 2002.
[12] P. P. L. Regtien. Accurate optical proximity detector. *Instrumentation and Measurement Technology Conference, 1990. IMTC-90. Conference Record., 7th IEEE*, pages 141–143, 1990.
[13] J. Salisbury. Interpretation of contact geometries from force measurements. In *Proceedings 1984 Conference on Robotics*, pages 240–247, Atlanta, Georgia, March 1984. IEEE.
[14] A. Saxena, J. Driemeyer, J. Kearns, and A. Y. Ng. Robotic grasping of novel objects. In *NIPS*, 2006.
[15] A. Saxena, J. Driemeyer, and A. Y. Ng. Robotic grasping of novel objects using vision. *IJRR*, 27:157–173, 2008.
[16] A. Saxena, J. Driemeyer, and A. Y. Ng. Learning 3-d object orientation from images. In *International Conference on Robotics and Automation (ICRA)*, 2009.
[17] A. Saxena, L. Wong, and A. Y. Ng. Learning grasp strategies with partial shape information. In *AAAI*, 2008.
[18] J. R. Smith, E. Garcia, R. Wistort, and G. Krishnamoorthy. Electric field imaging pretouch for robotic graspers. In *International Conference on Intelligent Robots and Systems (IROS)*, 2007.
[19] T. Takahashi, T. Tsuboi, T. Kishida, Y. Kawanami, S. Shimizu, M. Iribe, T. Fukushima, and M. Fujita. Adaptive grasping by multi fingered hand with tactile sensor based on robust force and position control. In *ICRA*, 2008.
[20] J. Tegin and J. Wikander. Tactile sensing in intelligent robotic manipulation - a review. *Industrial Robot: An International Journal*, pages 264–271, 2005.
[21] S. Walker, K. Loewke, M. Fischer, C. Liu, and J. K. Salisbury. An optical fiber proximity sensor for haptic exploration. *ICRA*, 2007.
[22] R. Wistort and J. R. Smith. Electric field servoing for robotic manipulation. In *IROS*, 2008.