The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Learning Grasp Affordances with Variable Centroid Offsets

Thomas J. Palmer and Andrew H. Fagg

Abstract— **When grasping an object, a robot must identify the available forms of interaction with that object. Each of these forms of interaction, a grasp *affordance*, describes one canonical option for placing the hand and fingers with respect to the object as an agent prepares to grasp it. The affordance does not represent a single hand posture, but an entire manifold within a space that describes hand position/orientation and finger configuration. Our challenges are 1) how to represent this manifold in as compact a manner as possible, and 2) how to extract these affordance representations given a set of example grasps as demonstrated by a human teacher.**

In this paper, we approach the problem of representation by capturing all instances of a canonical grasp using a joint probability density function (PDF) in a hand posture space. The PDF captures in an object-centered coordinate frame a combination of hand orientation, grasp centroid position and offset from hand to centroid. The set of canonical grasps is then represented using a mixture distribution model. We address the problem of learning the model parameters from a set of example grasps using a clustering approach based on expectation maximization. Our experiments show that the learned canonical grasps correspond to the functionally different ways that the object may be grasped. In addition, by including the grasp centroid/hand relationship within the learned model, we eliminate this as a hard-coded parameter and the resulting approach is capable of separating different grasp types, even when the different types involve similar hand postures.

I. INTRODUCTION

Manipulating one's world in very flexible ways is a skill that is shared only by a small number of species. Humans are particularly skilled at applying their manipulation abilities in novel situations using a range of effectors, from hands and other parts of the body, to tools. How can robots come to organize and learn knowledge representations for solving grasping and manipulation problems in unstructured environments? J. J. Gibson [9], [10] suggests that these representations should be partitioned into *what* can be done with particular objects and *why* an object should be manipulated in a certain way. The first of these, which Gibson terms *object affordances*, captures the details of what can be done with the object by the agent. The latter captures information about how individual manipulation skills are to be put together in order to solve a specific task. The task-neutral affordance representation is important in that it can provide an agent with a menu of actions or activities that are possible with a given object – whether the current task is well known or not.

In this paper, we examine the grasp affordance question. For a given object, we would like to compactly represent the feasible set of grasps that can be used with that object. These representations should be sufficient to enable a robot to execute the grasp, recognize the use of the grasp as made by other agents and even form a plan for how the grasp could subsequently be used in a task. For example, a cup might be grasped somewhere around its circumference using a ball type grasp, or a cereal box might be grasped along its thin side, to enable pouring, using the finger tips in opposition to the thumb.

Shape primitive approaches address this problem of associating objects with possible grasps by decomposing an object into a collection of volumetric primitives such as cylinders, rectangular prisms, spheres and cones (e.g., [1], [13]). Each primitive is associated *a priori* with a set of possible hand postures that can be used to grasp the component. Candidate grasps are then pruned based on a variety geometric and grasp quality constraints.

Visual feature approaches directly map identifiable visual features to particular hand postures (e.g., [11], [14]). Coelho et al. [3] and Piater et al. [15] explicitly learn the relationship between specific visual features and successful hand postures. In their work, the hand postures are discovered through a haptic exploration process. Hence, the resulting representations are rooted in the agent's own experiences with the objects.

Manifold approaches describe the feasible set of grasps in terms of a set of points within a space that captures hand position/orientation relative to the object and finger configuration (e.g., [16], [6]). De Granville et al. describe these manifolds using a mixture probability density function approach, in which each PDF is a joint PDF over hand position, orientation, and (in some cases) finger configuration [5], [4]. Because a nontrivial degree of hand position variation can be seen with small changes in finger configuration, but with little to no change in the location of contacts, the joint PDF captures the position of a *grasp centroid* rather than the hand explicitly. The grasp centroid is assumed to be at a constant offset from the hand location. De Granville et al. have shown that the parameters for an appropriate mixture distribution can be learned from a large set of grasps demonstrated by a human teacher, and that individual PDFs correspond roughly to the functional ways that the object may be grasped.

While the approach of de Granville can work well in some contexts, the relative position of the contacts and the hand can vary dramatically depending on the choice of grasp. For example, a cup might be grasped with fingertips in a

T. J. Palmer is a Ph.D. student and University of Oklahoma Foundation Fellow, University of Oklahoma, Norman, OK 73019, USA tjpalmer@tjpalmer.com

A. H. Fagg is an Associate Professor of Computer Science and Bioengineering, University of Oklahoma, Norman, OK 73019, USA fagg@cs.ou.edu

978-1-4244-3804-4/09/$25.00 ©2009 IEEE

precision grasp or enclosed in a power grasp. In this paper, we address this problem by allowing a an offset from the hand to the grasp centroid to be selected by the algorithm on a grasp-by-grasp basis. We show that the estimation of this grasp centroid offset can be performed as part of the expectation maximization (EM) algorithm [7] that is used to estimate the mixture PDF parameters. This eliminates a hard-coded parameter that can result in some poor-performing models. Furthermore, we show experimentally that by adding these extra degrees of freedom to the models, the learning algorithm is capable of separating different grasp types, even when the different types involve similar hand postures.

II. METHODS

The set of hand postures that correspond to feasible grasps of a particular object can be described as a manifold in hand posture and finger configuration space. Our challenges are 1) to generally represent these manifolds in as compact a manner as possible such that the representation makes explicit the functionally different ways that an object can be grasped and 2) to construct such a representation for a specific object given a set of examples of grasping it.

Consider grasping a cylindrical object from the side using a precision type grasp (e.g., as if to drink from a cup). Fig. 1a shows such a grasp, where the location of the hand is described in the object coordinate frame as ${}^{Ob}x$, and the orientation of the hand is described as ${}^{Ob}_{H}R$. Given all possible approach directions, the set of hand positions forms a ring around the object. The question is: how do we model this set of solutions in as simple a manner as possible? One possibility is to model the location of a *grasp reference point* instead of the hand location directly. In Fig. 1a, this grasp reference point, ${}^{Ob}y$, is modeled as a fixed translation from the hand, ${}^{H}S_1$. If this translation (or offset) is selected appropriately, the set of grasp reference points that results from all possible approach directions forms a compact set in Cartesian space at the center of the object. In contrast, when the object is grasped using a power (or palmar) grasp, the set of hand positions also forms a ring around the object, but at a smaller radius (Fig. 1b). By selecting an appropriately scaled translation, ${}^{H}S_2$, the set of grasp reference points also forms a compact set, which we refer to as the *grasp centroid*.

Because the grasp reference points form a compact set, it is convenient to describe this set using a Gaussian distribution. We approach the general problem of representing the set of hand configurations by using a joint probability density function (PDF) over the grasp reference points and hand orientations (we do not explicitly treat finger configuration in this paper). We capture multiple canonical grasps (e.g., the precision and power grasps of Fig. 1) using a mixture model of the joint PDFs. Given a set of example grasps, we can treat the learning problem as one of clustering in which the parameters of the PDFs are learned at the same time as individual samples are clustered into the component PDFs. We employ expectation maximization (EM) to perform this clustering process [7].

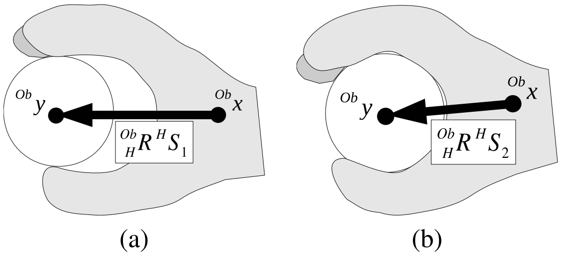


Fig. 1. Precision (a) and power (b) grasps, for example, have different offset vectors (${}^{H}S_j$) from the the hand (${}^{Ob}x$) to the grasp reference point (${}^{Ob}y$). The offset is expressed in the hand coordinate frame.

In practice, the grasp reference point is frequently contained within the contact points of the hand with the object, though in some cases it could be at some other displacement. For example, when grasping around the outside of a large disc, the grasp affordance centroid might be at the center of the disc, even if the fingers do not reach the center.

A. A PDF Representation of Grasp Affordances

Each demonstrated grasp posture i consists of the hand's position ${}^{Ob}x_i \in \mathbb{R}^3$ and rotation ${}^{Ob}_{H}R_i \in SO(3)$, both in the coordinate frame of the object. Given a set of sample postures representing valid grasps of the object, we desire to cluster these samples using a weighted mixture model of PDFs. The mixture PDF, h, representing the likelihood of a hand posture given that the agent is grasping the object, is given by:

$$h({}^{Ob}x_i, {}^{Ob}_{H}R_i|\Phi) = \sum_{j=1}^{M} w_j g_j({}^{Ob}x_i, {}^{Ob}_{H}R_i|\theta_j), \qquad (1)$$

where Φ is the full set of parameters, M is the total number of clusters, w_j is the weight of cluster j, g_j is the likelihood of the posture given cluster j, and θ_j is the parameter set of cluster j. Also, $\sum_{j=1}^{M} w_j = 1$, where each w_j can be interpreted as being the probability of a sample falling within cluster j.

Following de Granville et al., the PDF of each cluster is described as a joint PDF in both position and orientation. We assume that these two components are independent given the cluster:

$$g_j({}^{Ob}x_i, {}^{Ob}_{H}R_i|\theta_j) = p({}^{Ob}x_i, {}^{Ob}_{H}R_i|\theta_{pj}) f_j({}^{Ob}_{H}R_i|\theta_{fj}), \quad (2)$$

where $p(.)$ describes the position likelihood and $f_j(.)$ describes the likelihood of the orientation. The distribution parameters are split into position and orientation components θ_{pj} and θ_{fj}. Each f_j is one of two possible distributions in orientation space (and hence each is indexed by j).

The two types of distributions capture orientations in a unit quaternion space [12], [5], [4]. *Dimroth-Watson* distributions are Gaussian-like in their shape and are described by a "mean" rotation and a degree of allowable variation around this mean. *Girdle distributions* assign maximum likelihood to all rotations about some fixed, but arbitrary, axis. This likelihood drops as rotation deviates from this set. We refer the reader to de Granville et al. for more details [5], [4].

We model the distribution of grasp reference points by a multivariate Gaussian. However, we need to allow each cluster to have a unique offset from the hand to the grasp centroid. Therefore,

$$p({}^{Ob}x_i, {}^{Ob}_{H}R_i | {}^{Ob}\mu_j, V_j, {}^{H}S_j) = \frac{1}{(2\pi)^{3/2}|V_j|^{1/2}} \exp\left(-\frac{1}{2}\delta_{ij}^T V_j^{-1} \delta_{ij}\right), \quad (3)$$

where ${}^{Ob}\mu_j \in \mathbb{R}^3$ is the grasp centroid mean in the object's coordinate frame and $V_j \in \mathbb{R}^{3\times 3}$ a covariance matrix, ${}^{H}S_j \in \mathbb{R}^3$ is the offset from hand to grasp centroid in the *hand's* coordinate frame, and δ_{ij} is the vector from the cluster mean to the grasp reference point sample:

$$\delta_{ij} = {}^{Ob}_{H}R_i \; {}^{H}S_j + {}^{Ob}x_i - {}^{Ob}\mu_j \; . \quad (4)$$

B. Parameter Estimation

Given N sample hand postures, we use expectation maximization (EM) to find the parameters for a set of clusters defined by Eq. (1) and the probability α_{ij} that sample i belongs to cluster j. Sample membership estimation is the *expectation step,* and cluster parameter estimation is the *maximization step.* Here, we derive the effect of the offset parameter ${}^{H}S_j$ on the parameter estimation process.

The EM approach selects parameters to maximize the expected log-likelihood (ELL) of the joint event for all samples i and hidden variables. ELL is given by:

$$ELL = \sum_{i=1}^{N}\sum_{j=1}^{M} \alpha_{ij} \log\left(w_j g_j({}^{Ob}x_i, {}^{Ob}_{H}R|\theta_j)\right), \quad (5)$$

where $\sum_{j=1}^{M} \alpha_{ij} = 1$ for each sample i. The total number of clusters M is fixed for each use of EM. More specifically, a certain number of Dimroth-Watson and girdle clusters are specified in advance.

Focusing on the offset parameter ${}^{H}S_j$ for a particular cluster, substituting Eq. (2) into the above and simplifying yields:

$$\max_{{}^{H}S_j} ELL = \max_{{}^{H}S_j} \sum_{i=1}^{N} \alpha_{ij} \delta_{ij}^T V^{-1} \delta_{ij}. \quad (6)$$

We find the maximum likelihood estimate for ${}^{H}S_j$ by taking the derivative of ELL with respect to ${}^{H}S_j$ and setting to 0:

$$0 = \sum_{i=1}^{N} \alpha_{ij} {}^{Ob}_{H}R_i^T V_j^{-1} \left({}^{Ob}_{H}R_i \; {}^{H}S_j + {}^{Ob}x_i - {}^{Ob}\mu_j\right),$$

which yields the solution:

$${}^{H}\hat{S}_j = \left(\sum_{i=1}^{N} \alpha_{ij} \left({}^{Ob}_{H}R_i^T \; V_j^{-1} \; {}^{Ob}_{H}R_i\right)\right)^{-1} \sum_{i=1}^{N} \alpha_{ij} {}^{Ob}_{H}R_i^T \; V_j^{-1} \left({}^{Ob}\mu_j - {}^{Ob}x_i\right). \quad (7)$$

Similar derivations exist for the other parameters. These are roughly equivalent to standard maximum likelihood parameter estimates for the Gaussian distribution except that the position of the hand is replaced with the position of the grasp centroid:

$${}^{Ob}\hat{\mu}_j = \frac{\sum_{i=1}^{N} \alpha_{ij} \left({}^{Ob}_{H}R_i \; {}^{H}S_j + {}^{Ob}x_i\right)}{\sum_{i=1}^{N} \alpha_{ij}} \text{, and} \quad (8)$$

$$\hat{V}_j = \frac{\sum_{i=1}^{N} \alpha_{ij} \delta_{ij} \delta_{ij}^T}{\sum_{i=1}^{N} \alpha_{ij}}. \quad (9)$$

Note that the update rules for some parameters (including the offset) depend on the values of other parameters. We update all distribution parameters in parallel.

The update rule for cluster weight is the same as for other mixture-of-PDF approaches, and the rules for the orientation component of each cluster are unchanged from prior work by de Granville et al. [5].

C. Model Selection

EM is a gradient ascent method used here for maximizing ELL in order to discover estimates for distribution parameters and the probability of samples belonging to particular clusters. For our domain, many local optima exist, with results varying greatly depending on initial conditions. To address this issue, we perform many attempts of EM from different randomly-selected initial conditions. Rather than select the best global result by highest ELL, we instead employ metrics designed also to limit model complexity.

A metric that is very similar to ELL, but which purposely avoids rewarding overlapping clusters, is the *completed log likelihood* (CLL):

$$CLL = \sum_{i=1}^{N}\sum_{j=1}^{M} \hat{\alpha}_{ij} \log\left(w_j g_j({}^{Ob}x_i, {}^{Ob}_{H}R_i|\theta_j)\right), \quad (10)$$

where $\hat{\alpha}_{ij}$ is 1 if cluster j is the highest likelihood cluster for sample i and 0 otherwise. That is, due to $\hat{\alpha}_{ij}$, each sample's likelihood counts only for its best-fitting cluster.

Furthermore, we explicitly want to punish mixture models with excessive numbers of clusters. Fewer clusters means a smaller number of identified grasps on which to apply other algorithms. Therefore, to punish more complex mixture models, we employ the *Integrated Completed Likelihood* (ICL) metric [2]:

$$ICL = -2 \, CLL + \zeta\nu \log(N), \quad (11)$$

where ζ determines the magnitude of the complexity punishment and ν is the number of degrees of freedom (parameters) in the PDF model. By this measure, more complicated distributions are punished more than simpler ones. In a sense, each distribution has to pay for its complexity by providing a sufficient fit. Unlike CLL, *lower* ICL is better.

Of all EM attempts performed from different initial conditions, the retained model is that with the best ICL as calculated on a separate set of validation samples. Also, we do not know *a priori* how many clusters are appropriate for a given object. Following de Granville et al., we try mixture models with different numbers of clusters and different combinations of Dimroth-Watson and girdle distributions.

The model among the combinations with the best ICL on a second validation data set is selected as the final solution for the data set.

III. EXPERIMENTS

We evaluate the capabilities and performance of our algorithm using a few objects (a hammer handle, a spray bottle, a plate, and a spoon), each of which can be grasped in several ways. In particular, we compare our algorithm with that of de Granville et al., which assumes a fixed offset from the hand coordinate frame to the grasp centroid.

A. Data Collection

When demonstrating grasp postures, we need to measure the hand position and orientation relative to the object. To do this, we attached Polhemus FASTRAK sensors to each, giving the position and orientation of each in the global coordinate frame. From these, the relative measures could be calculated. Typically, the teacher used the non-instrumented hand to hold the object to enable quick demonstration of many grasp poses around the object. Also, for the fixed offset experiments, we estimated the offset by taking a mean of samples while directly handling the sensor using a variety of grasp types.

For our experiments, we used ICL punishment factor $\zeta = 4$. In practice, we had seen this choice to reduce model complexity while not leading to excessively simplified and degenerate solutions.

For each object, a human teacher demonstrated a certain number of grasp postures. We performed 30 independent experiments with each data set. For each experiment, we randomly subsampled from this total. Specifically, we chose 1000 training samples for EM, 250 different validation samples for evaluating multiple EM attempts by ICL, 250 additional validation samples for comparing the results of different numbers and types of clusters (again, by ICL), and 250 independent test samples for the evaluation of the resulting models by CLL.

In all experiments, every possible combination of Dimroth-Watson and girdle clusters was attempted up to a limit of B clusters, chosen in each case to be somewhat greater than the number of clusters expected (with the intent of avoiding ceiling effects). For each combination of clusters, 60 attempts from different starting conditions were performed, each with 20 EM steps. These numbers were chosen based on exploratory experiments.

B. Performance Measures

Because our data set is an unlabeled set of example grasps, there is no innate correct answer. Therefore, when assessing experimental results, we are concerned with whether the clusters match our expectations. That is, for each cluster found, is it expected or extraneous? Further, are any expected clusters missing? We are also interested in the overall quality of fit of clusters to the test data. Therefore, when comparing results here, we emphasize the following measures:

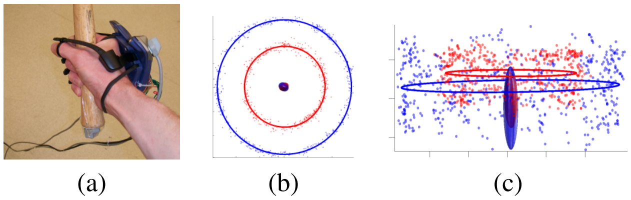


Fig. 2. Hammer handle (a) used for precision and power grasps. An example of approximately expected results is shown from top (b) and side (c). The point clouds show the measured hand positions. Offset from the hand points are ellipsoids representing the 3D Gaussians that capture the grasp centroids for each cluster. The orientation component of each cluster is a girdle distribution, as indicated by the visible rings. The inner ring is for power grasps, and the outer ring is for precision grasps.

- True positive rate (TPR) describes how many expected clusters are found. $TPR = TP/(TP+FN)$, where TP is the number of true positive identifications (expected clusters found in the results) and FN is the number of false negatives (expected but not found).
- Precision (PRC) describes how many resulting clusters are expected. $PRC = TP/(TP + FP)$, where FP is the number of false positive identifications (found clusters that are not expected, often due to unwanted splits of expected clusters).
- CLL measures the quality of fit for samples against the learned model.

TPR and PRC are subjective metrics. However, they are evaluated with respect to a set of expectations that are determined before the grasps are demonstrated. On the other hand, CLL is an objective metric of model quality.

C. Hammer Handle

As a simple example for discovering grasp offset, we demonstrated precision and power grasps around a hammer handle. Similar handles or other rotationally symmetric grasp options exist for various objects. Thus, this experiment represented a fundamental case to test the basic applicability of our method. Because of the clearly distinct offsets and many different approach directions, we expected the use of variable offsets to outperform the use of a fixed offset.

This data set included 2000 samples, 1000 for each grasp type. We limited the maximum number of clusters to $B = 5$ for this experiment. An example of the expected results is shown in Fig. 2. Specifically, true positives, false positives, and false negatives (as defined above) are judged in relation to these expected results.

Our proposed algorithm, with learned offsets, consistently found at least one ring for each grasp type. One example solution is shown in Fig. 4a, in which there was a clear separation between the clusters corresponding to the precision (red) and power grasps (blue). In contrast, when the fixed offset was used, the fit to the data was poor, as shown in Fig. 4b. In particular, in order to represent the interior points, one cluster (green) expands dramatically in the lateral directions. This case was classified as a false positive because much of the space supposedly available for a grasp reference point would result in a failed grasp.

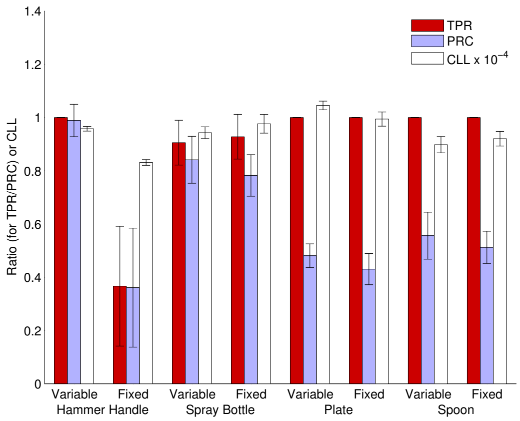


Fig. 3. TPR, PRC and CLL for each object and algorithm. Error bars show standard deviation of the respective measure.

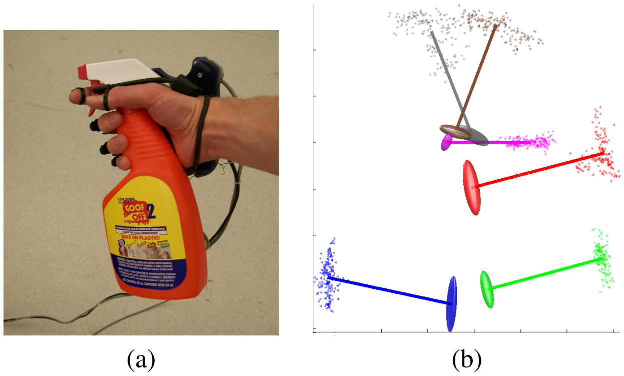


Fig. 5. Spray bottle (a) and example of approximately expected results (b). As for the expected hammer handle results, this result shows the basis for judgment of true positives, false positives, and false negatives. In this case, all expected clusters use a Dimroth-Watson distribution for their orientation component. The lines from hand point clouds to grasp centroid Gaussian means show the offset for each cluster.

Our new algorithm, with learned offsets, consistently found at least one ring for each grasp type. In 29 of 30 experiments, the algorithm discovered the expected solution of one inner and one outer ring. Only one case resulted in three rings, with the inner power grasp split into two clusters, one higher up the handle than the other. Specific results for TPR, PRC, and CLL are shown in Fig. 3. In contrast, the fixed offset approach consistently performed poorly on the data set; in some cases, it identified only a single cluster with a wide variance. Mean CLL for using learned offsets was about 15% greater than for using a fixed offset. In addition, the TPR and PRC scores were more than twice that of the fixed offset case. According to a two-sample t-test, all three of these differences are statistically significant ($p < 0.0001$).

D. Spray Bottle

To cover a more complicated example, though still with different expected offsets, we demonstrated grasps around a spray bottle as seen in Fig. 5. The grasps included a power grasp of the neck with the finger on the trigger (shown in magenta) as well as a precision grasp of the neck (red). In addition to these two grasps, we also demonstrated grasping the top from both sides (gray and brown) including some placing of the fingertips under the head and also grasping the base from both the front and back (blue and green). Because of the different grasp types, we expected the use of variable offsets to outperform the use of a fixed offset despite the added complexity. In all, we demonstrated 6 grasps. In the full data set, we had 1000 samples for each grasp. We allowed a maximum of $B = 10$ clusters.

Fig. 6 shows typical solutions for both the learned and fixed offset approaches. Of particular note, the fixed offset case more often required two clusters (magenta and orange) in order to capture the case of holding the spray bottle with the finger on the trigger. Note also that the center of these clusters is offset by a few centimeters (across all trials, mean $x = -4.3cm$ for fixed as opposed to mean $x = -2.9cm$ for learned). Both approaches frequently allocated two clusters to one of the grasps from above (shown as gray). This happened because of the wide spatial distribution of the hand locations for this grasp.

Overall, the use of fixed offsets did not perform as poorly as for the hammer handle, despite different offsets having been demonstrated. The use of variable offsets usually resulted in 5 true positives, while using fixed offset usually resulted in finding all 6, but with an increased number of false positives (as reflected in the PRC score). The mean CLL for the fixed offset case was about 3.5% more than that for the variable offset, a difference that was significantly different (two-sample t-test, p $<$ 0.0001). However, the use of variable offsets had a PRC score of about 7.5% more than that for use of a fixed offset, again with a statistically significant difference (p $<$ 0.01)

The small difference in performance between the two algorithms was due largely to the fact that only Dimroth-Watson (single orientation) distributions were necessary to explain the data. Because there was very little variation in orientation between the samples in each cluster, there was little difference in the variance of the spatial distribution between the hand and grasp reference points. In contrast, with the hammer handle case, because the variation in rotation was substantial (i.e., from all possible approach directions), the variance in the spatial distribution between hand and grasp reference points was very different. Consequently, we see a significant advantage to the proposed approach for the hammer handle, but not the spray bottle.

E. Plate and Spoon

In addition to the hammer handle and spray bottle, we compared the techniques on two more objects: a paper plate and a plastic spoon. These additional cases provided an opportunity to see if the pattern of results would be consistent. Without going into as much detail as above, the expected grasps for the plate were both rotationally symmetrical, and the expected grasps for the spoon were unidirectional. We held the plate around the rim and also rested it on the palm of the hand, with 1250 sample poses

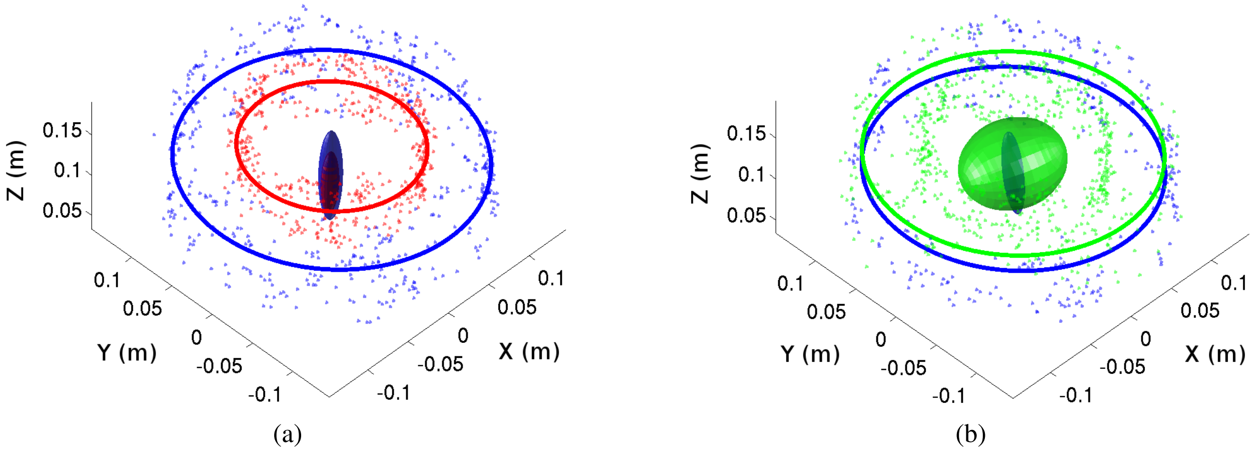


Fig. 4. Representative examples of hammer handle clusters for variable offset (a) and fixed offset (b). The wider Gaussian distribution in (b) was counted as a false positive.

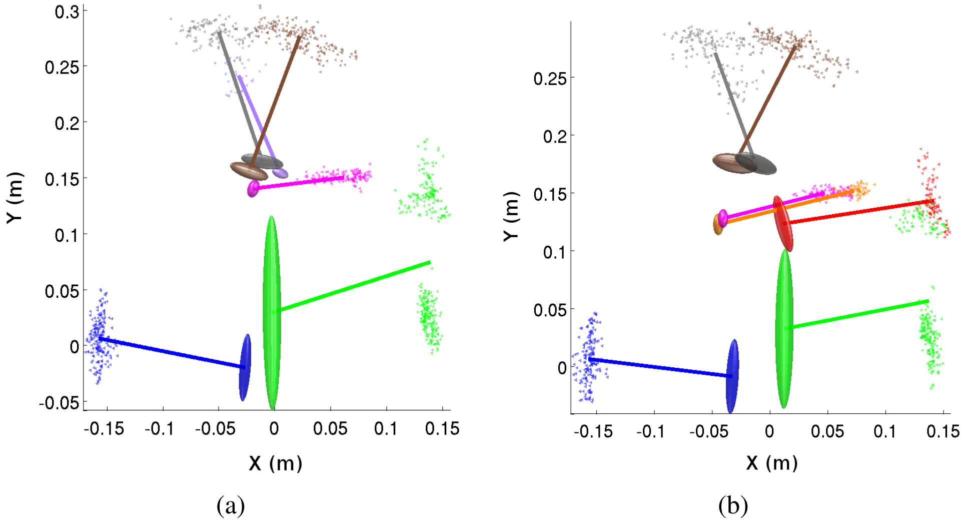


Fig. 6. Representative examples of spray bottle clusters for variable offset (a) and fixed offset (b). The variable offset results shown here were considered to have 5 true positives and 1 false positive (the additional grasp from above). The fixed offset results shown here were considered to have 6 true positives and 1 false positive (the additional grasp for trigger use).

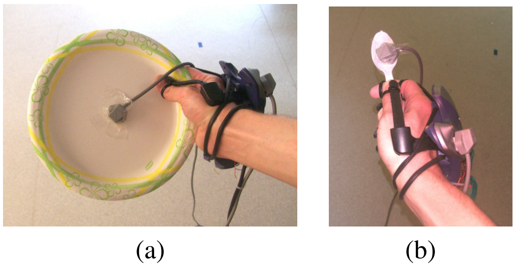
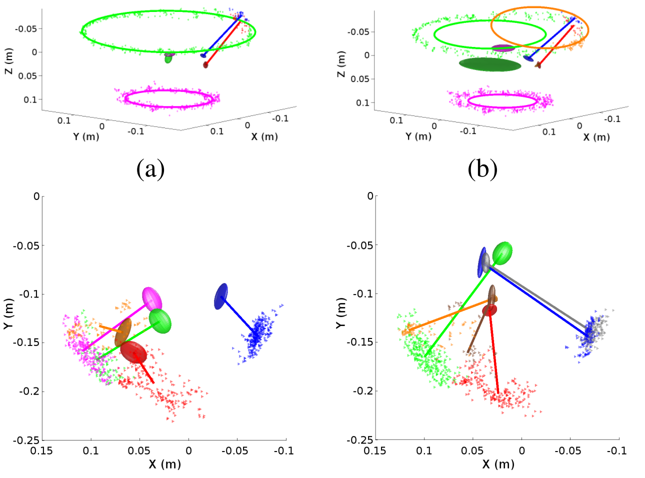


Fig. 7. Paper plate (a) and plastic spoon (b).

for each grasp type. We held the spoon in three fashions: from the side as if to feed oneself, from behind as if to feed another person, and from the tip of the handle as if to stir the contents of a tall container. We included 1500 samples of each grasp type for the spoon. When performing clustering for both the plate and the spoon, we allowed a maximum of $B = 10$ clusters. Figs. 7 and 8 show the objects and sample results.

As for the hammer handle, using a variable offset for the plate resulted in higher CLL and PRC than with a fixed offset, although by only about 11% and 5%, respectively. Still, both improvements were statistically significant (two-sample t-test, $p < 0.001$). Some differences exist from the hammer handle case; the two rings of hand positions were not concentric, and very different types of grasps were used. Also, the grasp centroid for the grasp around the rim was not enclosed by the fingers.

As for the spray bottle, using a variable offset for the spoon resulted in lower CLL (by about 2.5%) and higher PRC (by about 8.5%), again statistically significant results ($p < 0.015$). In addition, the clusters were in different places for each case. The variable offset placed the reference points nearer to the hand, while the fixed offset placed the reference points closer to the spoon handle. For both the plate and spoon, there were many false positives. For the plate, this seemed to be a side effect from the sensor cord being in the way of proper grasp demonstration. This yielded a substantially different distribution of samples for one approach orientation for the side of the plate. For the spoon, these false positives were likely due to the high



(a) (b)

(c) (d)

Fig. 8. Representative models learned for the plate (a and b) and spoon (c and d); for variable offset models (a and c) and fixed offset models (b and d). For the plate, the green upper ring corresponds to grasps around the rim, and the lower magenta ring corresponds to the plate resting on the palm of the hand. The extraneous clusters (usually unidirectional) seem to be a side effect from the sensor cord obstructing consistent grasp demonstration. For the spoon, only three clusters were expected. The side grasp (green and magenta) and tip grasp (orange, red, and brown) consisted of hand points more to the left of the handle. The grasp from behind (blue and gray), as demonstrated in panel c, consisted of hand points to the right of the handle. The sensor location, in the cup of the spoon, is at the origin.

rotational variance around a small object that was not large enough to qualify for a girdle distribution.

IV. DISCUSSION

In this paper, we propose an approach that allows an agent to observe a set of example grasps of an object made by a teacher and to construct a compact representation of the canonical grasps that may be made with the object. The object models are represented as mixture probability distributions defined in a hand posture space. In particular, by including a model parameter that describes the offset from hand to a center tool point, the algorithm is capable of distinguishing some functionally different grasps that involve different sets of contacts, even when there is not a dramatic difference in the pose of the hand across these grasps.

In using this approach in a complete system, several additional steps are necessary. First, although the learned affordance representation maps directly onto a reach controller that would enable a robot to move its hand into proximity with the object, we anticipate that haptic feedback would be used to further refine the grasp (e.g., [3]). Second, the proposed method is not limited to using data derived from a human teacher. Instead, a robot could produce experience that is specific to its own morphology (e.g., [8]).

Third, we are interested in making the connection between the visual representation of an object and these learned grasp affordances. Such a connection could be made in one of two ways. A learned visual representation could be used to recognize the identity and pose of a specific object. The pose would provide a coordinate frame onto which to hang the affordance representation, which, in turn, could provide reach goal locations. Alternatively, the learned visual representations could recognize more general components of objects. Each of these components would then be associated with their own affordance representation. Such an approach would enable a robot to approach a novel object, recognize its components and immediately have access to a set of candidate reach/grasp actions.

V. ACKNOWLEDGMENTS

This work is supported by a University of Oklahoma Foundation graduate fellowship.

REFERENCES

[1] G. A. Bekey, H. Liu, R. Tomovic, and W. J Karplus. Knowledge-based control of grasping in robot hands using heuristics from human motor skills. *IEEE Transactions on Robotics and Automation*, 9(6):709–722, 1993.

[2] C. Biernacki, G. Celeux, and G. Govaert. Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7):719–725, July 2000.

[3] J. A. Coelho, Jr., J. Piater, and R. A. Grupen. Developing haptic and visual perceptual categories for reaching and grasping with a humanoid robot. *Robotics and Autonomous Systems Journal, special issue on Humanoid Robots*, 37(2–3):195–219, November 2000.

[4] C. de Granville. Learning grasp affordances. Master's thesis, School of Computer Science, University of Oklahoma, Norman, OK, 2008.

[5] C. de Granville, J. Southerland, and A. H. Fagg. Learning grasp affordances through human demonstration. In *Proceedings of the International Conference on Development and Learning*, 2006. electronically published.

[6] C. de Granville, D. Wang, J. Southerland, and A. H. Fagg. Grasping affordances: Learning to connect vision to hand action. In Gaurav Sukhatme, editor, *The Path to Autonomous Robots; Essays in Honor of George A. Bekey*, pages 59–80. Springer, 2009.

[7] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood estimation from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

[8] R. Detry, E. Başeski, M. Popović, Y. Touati, N. Krüger, O. Kroemer, J. Peters, and J. Piater. Learning object-specific grasp affordance densities. In *Proceedings of the IEEE International Conference on Development and Learning*, 2009.

[9] J. J. Gibson. *The Senses Considered as Perceptual Systems*. Allen and Unwin, 1966.

[10] J. J. Gibson. The theory of affordances. In R. E. Shaw and J. Bransford, editors, *Perceiving, Acting, and Knowing*. Lawrence Erlbaum, Hillsdale, 1977.

[11] I. Kamon, T. Flash, and S. Edelman. Learning to grasp using visual information. In *Proceedings of the IEEE International Conference on Robotics and Automation*, volume 3, pages 2470–2476, 1996.

[12] J. Kuffner. Effective sampling and distance metrics for 3d rigid body path planning. In *Proceedings of the IEEE International Conference on Robotics and Automation*, 2004.

[13] A. T. Miller, S. Knoop, H. I. Christensen, and P. K. Allen. Automatic grasp planning using shape primitives. In *Proceedings of the IEEE International Conference on Robotics and Automation*, pages 1824–2829, 2003.

[14] A. Morales, P. J. Sanz, A. P. del Pobil, and A. H. Fagg. An experiment in constraining vision-based finger contact selection with gripper geometry. In *Proceedings of the International Conference on Intelligent Robots and Systems (IROS'02)*, 2002.

[15] J. H. Piater and R. A. Grupen. Learning appearance features to support robotic manipulation. In *Proceedings of the Cognitive Vision Workshop*, 2002. Electronically published.

[16] J. Steffen, R. Haschke, and H. J. Ritter. Experience-based and tactile-driven dynamic grasp control. In *Proceedings of the IEEE/RSJ International Conference on Robots and Systems (IROS)*, pages 2938–2943, 2007.