

# A Shape Matching Algorithm for Synthesizing Humanlike Enveloping Grasps

Ying Li and Nancy S. Pollard  
 School of Computer Science  
 Carnegie Mellon University  
 5000 Forbes Ave., Pittsburgh PA 15213, USA  
 {yingli, nsp}@cs.cmu.edu

**Abstract**—Humanoid robots must be capable of interacting with the world using their hands. A variety of humanlike robot hands have been constructed, but it remains difficult to control these hands in a dexterous way. One challenge is grasp synthesis, where we wish to place the hand and control its shape to successfully grasp a given object. In this paper, we present a data-driven approach to grasp synthesis that treats grasping as a shape matching problem. We begin with a database of grasp examples. Given a model of a new object to be grasped (the query), shape features of the object are compared to shape features of hand poses in these examples in order to identify candidate grasps. For effective retrieval, we develop a novel shape matching algorithm that can accommodate the sparse shape information associated with hand pose and that considers relative placements of contact points and normals, which are important for grasp function. We illustrate our approach with examples using a model of the human hand.

**Index Terms**—Grasp synthesis, hands, shape matching

## I. INTRODUCTION

Humanoids robots must be able to interact with the world using their hands. As robot hands become more humanlike, problems of grasping and manipulation become especially challenging: the hand has a large number of degrees of freedom, there are many constraints due to hand kinematics and object geometry, and we would like to achieve results that are high quality and also appear natural.

Many of these problems can be addressed by making use of a database of successful grasps. Even though the hand has a large number of degrees of freedom, many grasps have similar hand shapes, and it may be easy to find an example grasp that is a good match to the current problem. In addition, the grasp database can be designed to contain only desirable examples—for instance, we may wish to construct a database only from grasps that are humanlike in character.

This paper presents a data driven algorithm for selecting hand poses to grasp an object in a natural way. The key insight presented in this paper is that when a grasp database is available, grasping can be treated as a shape matching problem (Figure 1). As intuition for why shape matching may be an effective way to select a grasp from a database, consider Figure 2, which shows the hand pose for a grasp of a mouse. The shape of the inside surface of the hand gives us a great deal of information about the shape of the mouse. If we find a similar distribution of contact points and normals on the

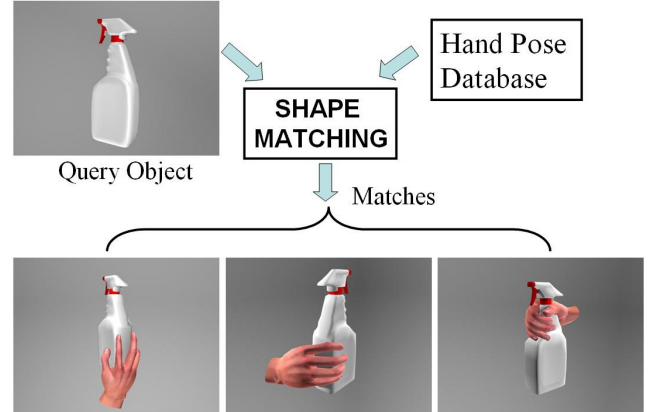


Fig. 1. Grasp synthesis as a shape matching problem. Offline, the user creates a database of hand poses. Online, a user or a program loads a query—a three-dimensional model of the object to be grasped. The shape matching system searches the database to find the hand poses that best match the query object. Representative hand poses for this example are shown at the bottom of the figure. The poses used are from a grasp of a mouse, a jelly jar, and a lightbulb (left to right).

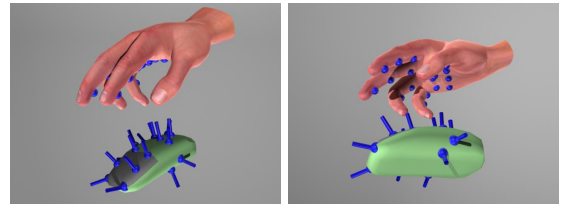


Fig. 2. Hand pose for the mouse grasp. The figure shows contact points on the hand and object, and contact normals on the object surface. Note that the inside surface of the hand contains a great deal of information about the shape of the mouse. If similar features can be found on a new object, it may be possible to use the same grasp for the new object.

surface of a different object, then we may be able to reuse the hand pose from the mouse grasp to acquire the new object. For example, this same grasp is used in the leftmost image on the bottom row of Figure 1 to hold the spray bottle.

Shape matching is commonly used in other fields—examples include object recognition in computer vision [14] and example-based retrieval of geometric models in computer graphics [21]. However, the use of shape matching for grasp synthesis presents some new challenges. The primary challenge is that the hand surface gives us information about only a portion of the object surface. Discriminative local features are typically used to compensate for this problem (e.g., [8]).

However, our local features are not discriminative; they are typically contact patches that are nearly planar and ellipsoidal in shape.

In this paper, we present a novel algorithm that allows us to overcome the problem of partial shape information that does not contain discriminative local features. The main idea behind our approach is to extract probabilistic samples of a global shape function from the hand shape. If similar samples can be located on an object surface, then the hand shape is a likely match to this object surface. We tailor our solution to the domain of grasp synthesis by making use of shape features that contain contact normal information, which is important for grasping.

Although our shape matching approach differs from previous work due to the characteristics of our problem, we believe that the primary contribution of this paper is to show that shape matching algorithms can be successfully used for grasp synthesis.

## II. BACKGROUND

Algorithms for grasp synthesis that consider complex hand kinematics are typically procedural or rule-based [25], [15], [9], [2], [3], [10]. These systems involve treating all or part of an object as a primitive shape (e.g. box, cylinder, sphere) for which a grasp synthesis strategy is available. This approach is supported by classifications and taxonomies of observed human grasps [26], [19], [4]. However, creating a rule-based grasping system based on these classifications requires a significant amount of skill on the part of the designer, and it may not always be clear how to fit irregularly shaped objects into a given classification system. Our goals in this paper are to reduce the knowledge and tuning required of the algorithm designer by (1) making use of a grasp database to capture variation in hand shape both across and within families of grasps, and (2) developing a shape matching algorithm that allows an appropriate hand shape to be identified directly from object geometry.

ElKoura and Singh [7] use a data driven approach to animate the hand for guitar playing. They use a database of human grasps to filter results produced using an inverse kinematics algorithm so that a natural coupling between joint angles is expressed. Yamane and colleagues [32] make use of a database to obtain full body posture for characters that are grasping and moving objects, and others have used data-driven approaches for tasks such as reaching, kicking, and locomotion (e.g., [31], [2], [12]). These systems work by identifying poses that match constraints such as end effector position. Extending this approach to work for grasps is challenging due to the large amount of contact between the hand and object. This difficulty motivated us to consider an alternative approach based on matching the shape of the inner surface of the hand to the geometry of the object.

Shape matching algorithms have been studied for related applications. (See [24], [14], [28] for surveys of techniques in these fields.) Most of these techniques, however, cannot be applied directly to our problem, because they assume

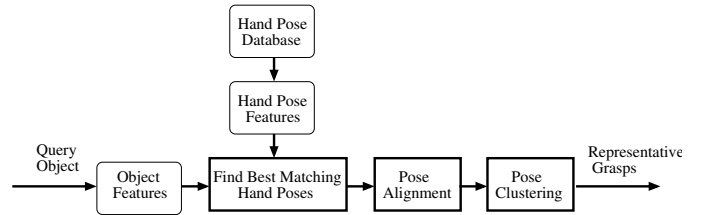


Fig. 3. Block diagram of the shape matching system. The online portion of this algorithm contains four main steps: compute the object features, find the hand poses that seem likely matches to this object, find good alignments of these poses to the object surface, and cluster the results.

problem characteristics that are not applicable here. Some techniques require information about the entire object, either to measure global features (e.g. [24]) or to compare probability distributions of features (e.g. [21]). We have only partial information about shape available from the hand pose. Some techniques require dense and discriminative information about local features, such as may be available from range images in cluttered environments (e.g., [11], [5]). We have very uninteresting local features: contact points are often flat patches on the object surface. Some techniques require registration / alignment of objects or feature correspondence, for example to build histograms characterizing the space occupied by the object [1]. We have no *a priori* way to align the query object to a hand pose.

In this paper, we combine and modify ideas from two bodies of work. First, we take inspiration from work on shape descriptors that probabilistically sample a global shape function (e.g. [21], [20]). This type of descriptor gives us information about how points on the surface are distributed with respect to one another. We adapt the descriptors used in previous work to provide more detailed information about contact normal distribution, which is important for grasping.

Second, we take inspiration from work on partial matching. In particular, we use a representative descriptor technique (i.e., a nearest neighbor search) to determine whether hand pose features can be found in the query object (e.g. [8], [17]). We modify this technique so that the descriptors are not based on local features but instead are based on a sample of a global shape function. We also introduce a weighting term to capture the intuition that some features are more important than others.

## III. SHAPE MATCHING FOR GRASP SYNTHESIS

Figure 3 shows a block diagram of our system. The input to the system is a geometric description of an object—the query object. The first step of the algorithm is to compute a collection of features that will be used to match the object against different hand poses. A hand pose database is available, and a similar feature set has been computed (offline) for each hand pose. There are then three main steps to the algorithm: (1) find the best hand poses to match the query object by comparing object features to hand pose features, (2) align these poses to the object geometry to obtain a set of possible grasps, and (3) cluster results into a few representative grasps. In this section we will talk about these three steps, beginning with our definition of the feature set we use here.

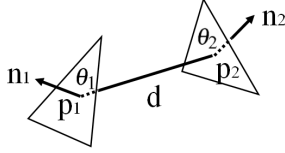


Fig. 4. We compute a three-dimensional feature value for pairs of points on the object or hand surface. This feature value consists of the distance  $d$  between points, and the angles  $\theta_1$  and  $\theta_2$  between the surface normals and the line passing through both points.

#### A. A Feature Set for Grasps

Our goal in choosing a feature set is to find a representation of object geometry that will allow us to quickly determine whether a given grasp can be matched to some portion of the object surface. In particular, we would like the feature set to capture important information about the relative configurations of contact positions and contact normals in the grasp.

We begin with an oriented point representation for both the hand pose and the object geometry. For the hand pose, the oriented point representation is just the set of contact points and normals taken from the grasp. For the query object, random sample points on the surface are identified using the sampling algorithm outlined in [21], and normals at those points are determined based on local surface geometry. The feature sets are computed from these oriented point representations.

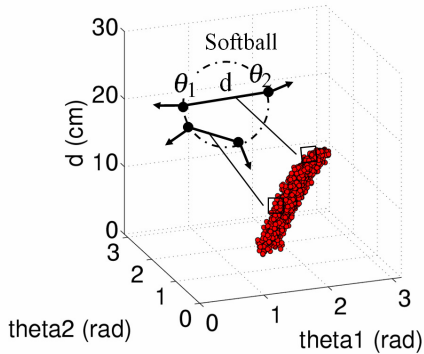


Fig. 5. Feature set for the softball. This figure shows 1% of the features actually used in our system.

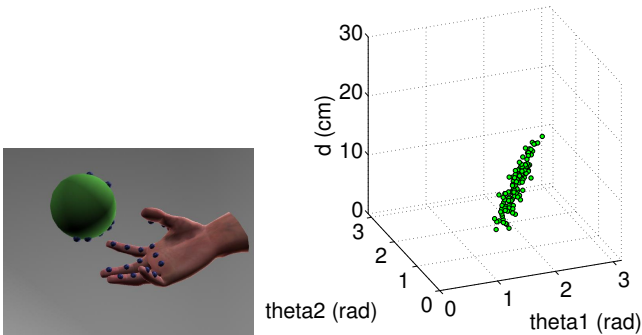


Fig. 6. (Left) Hand shape for the softball grasp. (Right) Feature set for the hand pose in this grasp. This figures shows half of the features available from the hand pose.

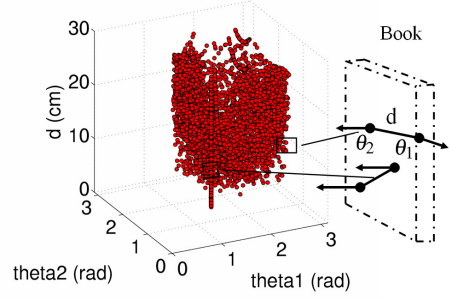


Fig. 7. Feature set for the book. This figure shows 1% of the features actually used in our system.

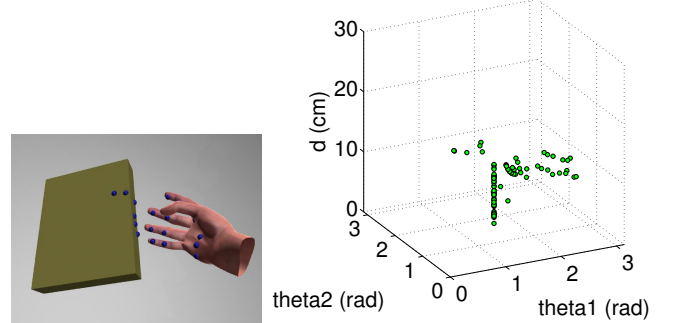


Fig. 8. (Left) Hand pose for the book grasp. (Right) Feature set for the hand pose in this grasp. This figure shows half of the features available from the hand pose.

A variety of considerations drove our particular choice of a feature set that differs from those already available in the literature. First, global features cannot be computed due to the sparse shape information available from hand pose. For example, aligning the “center of mass” of the hand pose with the center of mass of the object would not make sense. Second, local feature information is not helpful here, as regions of contact between the hand and object are typically devoid of discriminatory features such as edges, corners, and areas of high curvature. Third, features that are important are not necessarily those that are visually salient or visually discriminative as discussed by Shilane and colleagues, for example [27]. Instead, we want to capture how contact points and normals are distributed relative to one another in three-dimensional space, because this distribution of contacts is what allows us to create a solid grasp.

Because of these considerations, we develop a feature set based on random samples of pairs of contact points, and we include distance and normal information about each sampled pair. Specifically, we store for each pair of contact points a three-dimensional value that includes the distance between the points and the angles formed between the surface normals and the line passing through those points (Figure 4). This feature value is similar to that used by Ohbuchi and colleagues [20], but contains an extra dimension to locate the normals relative to the line between the contact points (not just relative to each other). Example feature sets are shown in Figures 5 and 6, which illustrate feature sets for a softball and for the hand pose when the hand is grasping the softball. Note that although the

$D$	database size = number of hand poses
$P_i$	set of features for hand pose $i$
$P_{i,k}$	feature $k$ of hand pose $i$ , represented $[d \ \theta_1 \ \theta_2]^T$
$Q_i$	set of features for object $i$
$Q_{i,k}$	feature $k$ of object $i$
$C(P_i)$	count of features for hand pose $i$
$C(Q_i)$	count of features for object $i$
$NN(A, b)$	nearest neighbor to $b$ in feature set $A$
$\alpha$	weights angular vs. linear terms in features

TABLE I  
DEFINITION OF TERMS.

hand pose feature set is more sparse than that of the softball, it captures its overall shape very well.

For good discriminatory power, we require that these feature sets differ substantially for different objects. As one example, Figures 7 and 8 show the feature sets for a book and for a grasp of that book. In this example, note that the feature set for the hand pose captures only a portion of the information contained in the feature set for the book. Yet this information is sufficient to distinguish this grasp from the softball grasp.

### B. Matching Hand Pose to Object

We compare a hand pose and a query object using their feature sets. For a matching hand pose, we expect all features present in the hand feature set to appear in the object feature set, while the converse is not generally true. We therefore use a representative descriptor approach for evaluating a match [17], [8].

The representative descriptor approach is designed for comparing a partial shape such as that obtained from a range image to a complete shape (a geometric model). The quality of the match is based on the average distance from features in the partial shape to their nearest neighbors in the complete shape. In this case, we compute distance from hand pose  $i$  to object  $j$ . This distance  $E$  is expressed as follows, where Table I contains notation for this section.

$$E = \frac{1}{C(P_i)} \sum_{k=1}^{C(P_i)} \text{Dist}(P_{i,k} - NN(Q_j, P_{i,k})) \quad (1)$$

$$\text{Dist}(x) = [d^2 + \alpha^2 \theta_1^2 + \alpha^2 \theta_2^2]^{\frac{1}{2}} \quad (2)$$

The first expression is just distance from each feature  $k$  in hand pose  $i$  to its nearest neighbor on object  $j$ , averaged over all hand pose features  $P_{i,k}$ . The distance metric  $\text{Dist}$  is designed to provide a reasonable weighting between linear and angular values. In our case, we use  $\alpha = 1\text{cm/rad}$ .

Although initial results were promising, this distance expression was not good enough. In particular, it would sometimes identify matches that were unacceptable because the object lacked certain key features (e.g., a place to put the thumb for the book grasp). To address this problem, we introduce a scheme to weight feature values more highly if they are relatively plentiful on the grasped object and relatively rare in

other hand poses. The weighted distance from hand pose  $i$  to object  $j$  is

$$E_w = \frac{1}{w_i C(P_i)} \sum_{k=1}^{C(P_i)} w_{i,k} \text{Dist}(P_{i,k} - NN(Q_j, P_{i,k})) \quad (3)$$

where the normalizing term  $w_i$  is the sum of all weights for hand pose  $i$ :

$$w_i = \sum_{k=1}^{C(P_i)} w_{i,k} \quad (4)$$

and the weight for feature  $k$  of hand pose  $i$  is computed as

$$w_{i,k} = \frac{\text{matchCount}(Q_i, P_{i,k})}{\sum_{l=1}^D (\text{matchCount}(P_l, P_{i,k}))} \quad (5)$$

The numerator of this equation is a count of the number of times a feature similar to  $P_{i,k}$  is found on object  $i$ , where object  $i$  is the object that was originally grasped using hand pose  $i$ . We check only a portion of the object surface, selected during construction of the database, as shown in Figure 11. The denominator contains a count of the number of times a feature similar to  $P_{i,k}$  is found in any hand pose. In other words,  $w_{i,k}$  is high when the feature is relatively common on the grasped object and/or when it is uncommon in other hand poses.

We can make a Bayesian argument for this weighting function. Let  $P(\text{pose}_i)$  indicate the probability of encountering the current hand pose  $i$ . Let  $P(\text{feature}_{i,k})$  indicate the probability of encountering the current feature  $P_{i,k}$ . Then the probability that hand pose  $i$  is a good match for object  $j$  based on finding the current feature  $P_{i,k}$  on object  $j$  can be expressed as

$$P(\text{pose}_i | \text{feature}_{i,k}) = \frac{P(\text{feature}_{i,k} | \text{pose}_i) P(\text{pose}_i)}{P(\text{feature}_{i,k})} \quad (6)$$

In weighting term  $w_{i,k}$ , we estimate  $P(\text{feature}_{i,k} | \text{pose}_i)$  by counting the number of times the feature appears on the object originally grasped using this hand pose.  $P(\text{pose}_i)$  can be considered uniform and we ignore it.  $P(\text{feature}_{i,k})$  is estimated by counting the occurrences of that feature over all hand poses.

The function  $\text{matchCount}$  used in computing weight  $w_{i,k}$  can be expressed as follows:

$$\text{matchCount}(A, b) = \sum_{k=1}^{C(A)} \text{isMatch}(A_k, b) \quad (7)$$

with

$$\text{isMatch}(A_k, b) = \begin{cases} 0 & \text{Dist}(A_k - b) > \epsilon \\ 1 & \text{Dist}(A_k - b) \leq \epsilon \end{cases} \quad (8)$$

A match is counted if the distance between corresponding features is less than a value  $\epsilon$ . In our experiments,  $\epsilon$  is set to 10% of the maximum distance between features in the hand pose under consideration.

In some cases, there may be many ways in which an object can be grasped. To narrow these options, we have implemented a very simple mechanism to allow the user to provide some

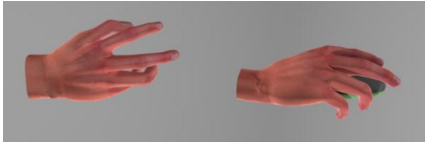


Fig. 9. It is not enough to identify matching hand poses. We must also find an alignment of the hand pose to the object surface.

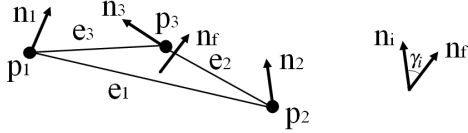


Fig. 10. A single triangle, called the Basic Partial Frame, is used to find a candidate set of alignments of hand pose to object surface. The Basic Partial Frame is created from three contact points selected from the hand pose.

guidance for the shape matching process. In particular, the user can select the portion of the object to be grasped, such as the object regions highlighted in Figure 11. Only surface points within the selected region will be considered during the shape matching process.

### C. Pose Alignment and Clustering

Once a hand pose is selected from the set of candidates, it must be aligned to the object shape. The shape matching algorithm described above identifies likely hand poses (those with minimum distance values  $E_w$  in Equation 3), but it does not compute an alignment of hand pose to object surface.

The goal of the alignment process is to find a transformation to be applied to the hand pose so that the contact points in the hand pose are brought into correspondence with points on the object having similar normals (Figure 9). Our alignment process has two parts. First, we narrow the set of transforms by identifying good alignments for a single triangle from the hand pose. Then, for each of these candidate transforms, we check all other contact points from the hand pose, discarding transforms that result in a poor match for any one of those contacts. Finally, we cluster the results. The next paragraphs provide more detail.

1) *Finding Candidate Transforms*: In the first step of the alignment process, coordinate transforms are collected by finding good matches between a triple of points from the hand pose and triples of points on the object. We indicate a triangle formed by three contact points on the hand pose  $p_1, p_2, p_3$  as a Basic Partial Frame, as shown in Figure 10. For the Basic Partial Frame, we compute edge lengths  $e_i$  and angles  $\gamma_i$  between the normals  $n_i$  and the triangle normal  $n_f$ . All triangles from the object surface are then tested for a match to this frame. Matching triangles must have edge lengths within value  $\epsilon_e$  and angles within value  $\epsilon_\gamma$ , where  $\epsilon_e = 0.5cm$  and  $\epsilon_\gamma = 0.5rad$  in our experiments. All matching triangles are recorded and passed on to the next step.

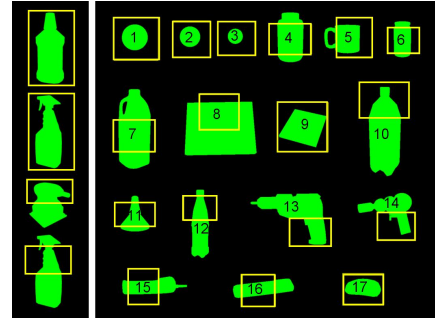


Fig. 11. (Left) New objects tested using our algorithm. (Right) Objects in our database. The highlighted “partial objects” on the right hand side show the regions of those objects that were grasped to form the hand pose database.

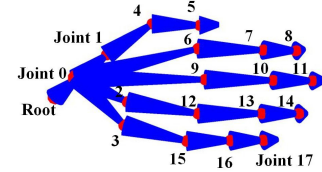


Fig. 12. Skeleton used to obtain poses from motion capture data. All joints were assumed to be ball joints for the purpose of pose capture.

2) *Pruning Poor Transforms*: Once a number of candidate transforms are generated using the Basic Partial Frame, each of these transforms is examined to test whether it represents a plausible grasp. All contact points are first transformed by aligning the corresponding triangles, and the pose is checked to see whether there are good matches for all contact points on the object surface. For each contact point, its nearest neighbor on the surface of the object is identified, and the entire transform is discarded if the distance from the contact point to the object surface is greater than value  $\epsilon_d$ , or the angle between the contact normal and the object surface normal is greater than value  $\epsilon_n$ , with  $\epsilon_d = 1cm$  and  $\epsilon_n = 1rad$  in our experiments.

3) *Clustering the Results*: At the end of the alignment process, there may be many possible grasps available for a single hand pose and object. These grasps are automatically clustered for effective presentation to the user. We use K-means clustering [6], with the number of clusters selected using the criterion proposed in Krzanowski and Lau [13]. Clustering was performed on the transformation used to align the hand pose to the object surface. We experimented with a variety of different subspaces that can be formed from this transformation and found that for our problem the most intuitive results were produced when clustering in the nine-dimensional space (the  $3 \times 3$  matrix) representing the rotation portion of this transformation.

## IV. THE GRASP DATABASE

We illustrate our results with a database consisting of 17 hand poses obtained from motion capture data. Objects that were grasped to form the database are shown in Figure 11. Grasps of these objects were captured using a Vicon optical

motion capture system, with motion capture data fit to the hand skeleton model shown in Figure 12 to determine the pose of the hand. Geometric models for the objects were created manually for each of the objects grasped. For each hand pose, contact regions were selected manually, and the normals at those contact points were obtained by taking the normal from the closest point on the object surface.

## V. RESULTS

We first tested whether our matching algorithm would select appropriate grasps for the objects used to form the database. The motivation behind this experiment is that if our algorithm does not work on the original data, then we must conclude that our feature set is not sufficiently discriminative.

Figures 13 and 14 show results from this initial experiment. Figure 13 illustrates results using the unweighted distance  $E$  from Equation 1, and Figure 14 illustrates weighted results using distance  $E_w$  from Equation 3. Hand poses 1 through 17 are shown on the x-axis, and query objects 1 through 17 are shown on the y-axis. The list of objects is:

- 1-Softball, 2-Baseball, 3-SmallBall, 4-SauceJar, 5-Mug, 6-JellyJar, 7-MilkJar, 8-LargeBook, 9-CD, 10-CokeBottle, 11-LightBulb, 12-WaterBottle, 13-Drill, 14-BubbleGun, 15-Phone, 16-RemoteControl, 17-Mouse

So, for example, column 15, row 15 represents the hand pose from the phone grasp compared to the object geometry of the phone, and column 10, row 15 represents the hand pose from the coke bottle grasp compared to the object geometry of the phone. All of these results are run with the user highlighting the relevant region of the object to consider. For example, for the drill geometry, the user selected the handle of the drill as shown in Figure 11 and other parts of the drill were not considered. In these two figures, a purely white cell would represent zero distance. The diagonal terms are not exactly zero due to errors in estimating contact positions from the motion capture data and due to discrete sampling of points on the object surface. This experiment shows that the weighting term used in Equation 3 is effective for pruning spurious matches.

Our second set of experiments was to find good matches for objects that were not in the database. We selected three objects of moderate complexity for this experiment: 1-MouthwashBottle, 2-Sander, and 3-SprayBottle. Figures 15 and 16 illustrate results for these three objects. Figure 15 contains results of searching for matches using the entire object geometry. Figure 16 contains results in a similar format when the user has highlighted the portion of the object to be grasped. Only results using weighted metric  $E_w$  are shown.

Table II shows the number of alignments and the number of clusters returned for the three new objects in each of the conditions tested. For example, the top section of the left column shows alignments for the mouthwash container when the entire object was considered. The five best matching hand poses based on their feature sets were the mug, phone, mouse, coke bottle, and softball poses. The mug pose resulted in

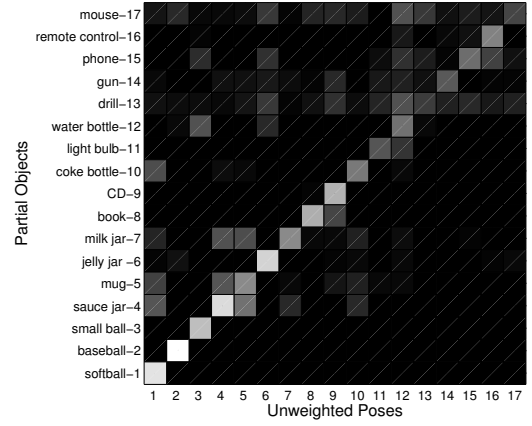


Fig. 13. Evaluation using the *unweighted* distance metric  $E$  (Equation 1). All hand poses in the database are compared to all of the objects for which grasps were collected. Lighter colors indicate lower values of  $E$ , or closer matches. A perfect discriminator would have white boxes on the diagonal and black everywhere else.

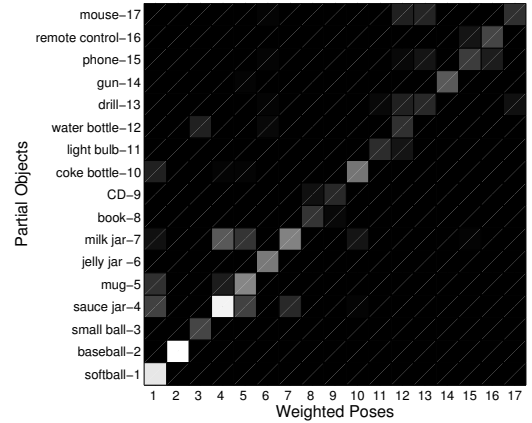


Fig. 14. Evaluation using the *weighted* distance metric  $E_w$  (Equation 3). All hand poses in the database are compared to all of the objects for which grasps were collected.

WHOLE OBJECT		PARTIAL OBJECT	
<b>Mouthwash</b>		<b>Mouthwash</b>	
5-Mug	399 (4)	17-Mouse	0 (-)
15-Phone	15 (4)	6-JellyJar	218 (4)
17-Mouse	2 (-)	15-Phone	8 (-)
10-CokeBottle	380 (4)	5-Mug	360 (4)
1-Softball	721 (2)	10-CokeBottle	11 (4)
<b>Sander</b>		<b>Sander</b>	
5-Mug	3 (-)	1-Softball	31 (4)
4-SauceJar	8 (-)	5-Mug	2 (-)
1-Softball	59 (2)	17-Mouse	0 (-)
<b>SprayBottle</b>		<b>SprayBottle</b>	
15-Phone	4 (-)	15-Phone	0 (-)
17-Mouse	13 (4)	13-Drill	0 (-)
6-JellyJar	163 (4)	12-WaterBottle	243 (7)
		11-Lightbulb	13 (3)

TABLE II

RESULTS FOR THREE NEW OBJECTS. (LEFT) THE ENTIRE OBJECT SURFACE WAS USED. (RIGHT) A PORTION OF THE OBJECT WAS SELECTED. BOTH ALIGNMENTS AND CLUSTERS (IN PARENTHESES) ARE SHOWN. HAND POSES WITH 8 OR FEWER ALIGNMENTS WERE NOT CLUSTERED.



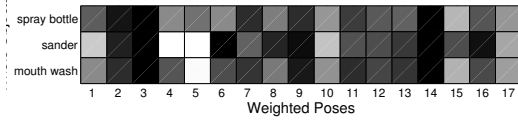


Fig. 15. Test of three new objects against the hand pose database. The *entire* object was considered. The figure shows weighted results, using  $E_w$ .

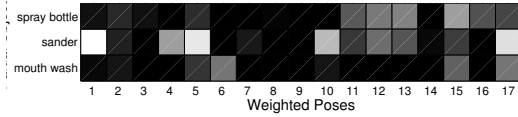


Fig. 16. Test of three new objects against the hand pose database. Only *a portion of the object* was considered. The figure shows weighted results, using  $E_w$ .

399 valid alignments with the object surface, which could be grouped into 4 clusters based on orientation of the hand pose. When only a portion of the mouthwash container was considered, the best matching hand poses were from the mouse, jelly jar, phone, mug, and coke bottle grasps. Although the mouse grasp was best in terms of its feature set, no valid alignments could be found on the object surface. In other words, even though similar features to all pairs of points in the grasp could be found on the object surface, none of these matching pairs led to a complete matching grasp.

Figure 17 shows some of the final results of our algorithm. Each row shows all cluster results where a single hand pose matches the given object. For example, the top row shows the median cluster representative for each cluster matching the coke bottle pose to the mouthwash object.

## VI. DISCUSSION

This paper describes a shape matching algorithm for synthesizing humanlike enveloping grasps that is inspired by work in geometric model matching. We have described an approach that handles the specific challenges of applying shape matching to grasp synthesis. The availability of partial information is handled using a representative descriptor approach, and the lack of good local features is handled by probabilistically sampling a global shape function. We introduce a novel global shape function that emphasizes the relations between relative positions and normals of pairs of oriented points and we introduce a mechanism to weight features relative to one another to improve hand pose selection.

Our results show that this algorithm can find very good matches between the hand and the object surface. In fact, we were surprised at the variety of good matches found using our small database of 17 grasps. As an interactive tool, this system is useful as-is, because the user can select the desired grasp from among the possibilities presented by the system. However, many of the grasps seem unintuitive—the hand is grasping the object “backwards” or “upside down” or is not grasping a tool in the manner necessary to complete a task. Objects such as tools may require customized shape matching

algorithms (e.g., see Miyata et al. [16]), and future work includes exploring the use of task based quality metrics (e.g., [22]).

The results shown in Figure 17 are not refined in any way. We have set the problem up as a database retrieval problem, and so the resulting grasps will typically not achieve exact contact with the object surface. However, in our results, the matches have hand configurations that follow the object surface quite closely, and the additional cleanup that is required should not be difficult. Automating this cleanup process is another area of future work.

To use this algorithm with a robot hand, we must have a grasp database specific to that hand. For hands that are nearly humanlike (e.g. the Gifu Hand [18], [29] and the Shadow Hand [30]), we have had good preliminary success mapping captured human grasps directly to the robot kinematics using the algorithm by Zordan and Horst [33], which works by using artificial forces to draw virtual markers on the hand toward measured marker positions at each frame of the motion capture sequence. To achieve the selected grasps, a suitable control algorithm must be designed. We have had success at extracting simple controllers for grasping based on human motion capture data [23] and plan to combine the controller from that work with the grasp synthesis algorithm described here to produce a more complete grasping system for humanlike robotic hands.

In general, we believe that the strength of our algorithm is in identifying a set of plausible grasps based only on object shape. Our primary emphasis for the future will be to provide more effective ways to prune these grasps, refine them, and present them to the user for consideration or provide a mechanism to automatically select the single most likely match for use by a robot.

## ACKNOWLEDGMENTS

We would like to thank: Moshe Mahler for help with the hand model and Jiaxin Fu and Ting Ting Wu for help in processing the motion capture data. We would also like to acknowledge NSF grants CCF-0343161, IIS-0326322, ECS-0325383, and CNS-0224419.

## REFERENCES

- [1] M. Ankerst, G. Kastentmiller, H.-P. Kriegel, and T. Seidl. Nearest neighbor classification in 3D protein databases. In *Proc. 7th Int. Conf. on Intelligent Systems for Molecular Biology*, 1999.
- [2] Y. Aydin and M. Nakajima. Database guided computer animation of human grasping using forward and inverse kinematics. *Computers and Graphics*, 23:145–154, 1999.
- [3] 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.
- [4] M. R. Cutkosky and R. D. Howe. Human grasp choice and robot grasp analysis. In S.T. Venkataraman and T. Iberall, editors, *Dextrous Robot Hands*. Springer-Verlag, New York, 1990.
- [5] C. Dorai and A. K. Jain. COSMOS—A representation scheme for 3D free-form objects. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(10):1115–1130, 1997.
- [6] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. John Wiley & Sons, Inc., New York, 2001.
- [7] G. ElKoura and K. Singh. Handrix: Animating the human hand. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 2003.



Fig. 17. Cluster results for grasps matching three query objects that were not in the database. The top two rows show matches to the complete object geometry. The bottom two rows show matches to a portion of the object geometry selected by the user. Top to bottom, these are cluster representatives for the coke bottle pose, the jelly jar pose, the softball pose, and the lightbulb pose.

- [8] A. Frome, D. Huber, R. Kolluri, T. Buelow, and J. Malik. Recognizing objects in range data using regional point descriptors. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2004.
- [9] Z. Huang, R. Boulic, N. Magnenat Thalmann, and D. Thalmann. A multi-sensor approach for grasping and 3D interaction. In *Proc. Computer Graphics International*, 1995.
- [10] T. Iberall. Human prehension and dextrous robot hands. *International Journal of Robotics Research*, 16(3), June 1997.
- [11] A. E. Johnson and M. Hebert. Using spin-images for efficient multiple model recognition in cluttered 3-D scenes. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 21(5):433–449, 1999.
- [12] L. Kovar and M. Gleicher. Automated extraction and parameterization of motions in large data sets. *ACM Transactions on Graphics*, 23(3), 2004.
- [13] W. J. Krzanowski and Y. T. Lai. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics*, 44:23–34, 1988.
- [14] G. Mamic and M. Bennamoun. Representation and recognition of 3D free-form objects. *Digital Signal Processing*, 12(1):47–76, 2002.
- [15] R. Mas Sanso and D. Thalmann. A hand control and automatic grasping system for synthetic actors. In *Eurographics '94*, 1994.
- [16] N. Miyata, M. Kouchi, T. Kurihara, and M. Mochimaru. Analysis of human hand posture to estimate mouse-operation. In *IEEE Int'l. Conf. on Systems, Man, and Cybernetics*, pages 3446–3451, 2003.
- [17] G. Mori, S. Belongie, and J. Malik. Shape contexts enable efficient retrieval of similar shapes. In *IEEE Computer Vision and Pattern Recognition (CVPR)*, 2001.
- [18] T. Mouri, H. Kawasaki, K. Yoshikawa, J. Takai, and S. Ito. Anthropomorphic robot hand: Gifu Hand III. In *Proc. Int'l Conference on Control, Automation, and Systems*, 2002.
- [19] J. Napier. *Hands*. Princeton Univ. Press, Princeton, NJ, 1980.
- [20] R. Ohbuchi, T. Minamitani, and T. Takei. Shape-similarity search of 3D models by using enhanced shape functions. In *Proceedings of the Theory and Practice of Computer Graphics*, 2003.
- [21] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Transactions on Graphics*, 21(4), 2002.
- [22] N. S. Pollard. Closure and quality equivalence for efficient synthesis of grasps from examples. *International Journal of Robotics Research*, 23(6):595–614, 2004.
- [23] N. S. Pollard and V. B. Zordan. Physically based grasping control from example. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 2005.
- [24] R. J. Prokop and A. P. Reeves. A survey of moment-based techniques for unoccluded object representation and recognition. *Graphical Models and Image Processing*, 54:438–460, 1992.
- [25] H. Rijpkema and M. Girard. Computer animation of knowledge-based human grasping. In *Computer Graphics (SIGGRAPH 91 Proceedings)*, volume 25, pages 339–348, 1991.
- [26] G. Schlesinger. Der mechanische Aufbau der künstlichen Glieder (the mechanical building of artificial limbs). In M. Borchardt, editor, *Ersatzglieder und Arbeitshilfen für Kriegsbeschädigte und Unfallverletzte (Prosthetic and Occupational Help for War Victims and Accident Victims)*. Springer, 1919.
- [27] P. Shilane, P. Min, M. Kazhdan, and T. Funkhouser. The Princeton shape benchmark. In *Proceedings Shape Modeling International*, 2004.
- [28] J. W. H. Tangelder and R. C. Veltkamp. A survey of content based 3D shape retrieval methods. In *Proceedings Shape Modeling International*, pages 145–156, 2004.
- [29] Gifu Hand III web page. <http://www.kk-dainichi.co.jp/e/gifuhand.html>.
- [30] Shadow Dexterous Hand web page. <http://www.shadow.org.uk/products/newhand.shtml>.
- [31] D. J. Wiley and J. K. Hahn. Interpolation synthesis of articulated figure motion. *IEEE Computer Graphics and Applications*, Nov/Dec:39–45, 1997.
- [32] K. Yamane, J. Kuffner, and J. K. Hodgins. Synthesizing animations of human manipulation tasks. *ACM Transactions on Graphics*, 23(3), 2004.
- [33] V. B. Zordan and N. V. D. Horst. Mapping optical motion capture data to skeletal motion using a physical model. In *ACM SIGGRAPH / Eurographics Symposium on Computer Animation*, 2003.