Towards a complete description of grasping kinematics: a framework for quantifying human grasping and manipulation

Qiushi Fu, Marco Santello

Abstract—**This paper proposes a framework for tracking both human hand kinematics and object contour during grasping task. The framework is based on modeling the object as point cloud and the use of marker-based tracking. We introduce how to estimate contact sites on both the hand and object, hand enclosing space, and graspable features from recorded data. Two experiments were performed to 1) verify the accuracy of contact site estimation (less than 5 mm), and 2) validate the feature extraction. Our approach can provide significant insight into how humans plan grasping and manipulation based on object recognition.**

I. Introduction

There has been extensive effort on tracking finger joint kinematics during human grasp behaviors to quantify the features of the high dimensional space of human hand [1, 2, 3]. Many studies have shown that grasping movements can be represented in a low dimensional space. However, there are other important features that characterize human grasping, in particular how humans select graspable features of objects. Surprisingly, systematic analyses of this issue have been largely overlooked in the neuroscience literature. The concept of grasping affordance has been studied for many years, which is defined as the quality of an object that allows a person to grasp and perform an action [4]. However, this concept has not been systematically modeled and analyzed, thus preventing further understanding of how object properties are represented and how grasping is planned in the central nervous system. These gaps stem from lack of systematic measurements and quantification of (a) where the object is grasped and (b) what parts of the hand make contact with the object. This is probably because traditional kinematic recording techniques used for human grasp studies make it difficult and tedious to track contact areas on the object and hand.

In the field of robotics, recent studies have attempted to quantify the mapping between cues derived from perception of object features and the interaction between robotic hands and objects. The robots can use data generated from human as a training set to generalize the knowledge of how and where to grasp. This can be attained through two main methods: one is to track position and orientation of the human hand as a demonstration for the robots [5, 6, 7], but the geometry of the object is not considered; the other is to label graspable features heuristically, such as grasping points [8] and graspable part [9] but the hand posture is generated computationally. It has been shown that a better understanding of human hand-object interaction and object recognition would help generate more robust and human-like grasps for robots [10]. It is then important to utilize a tracking framework that can help to build the complex mapping between object geometry, grasp affordance, and hand configuration.

The present work describes a framework that could bridge these gaps. Specifically, the key objective is to track the contour of the object and the hand at the same time. This is achieved by using our previously developed Kalman Filter for whole-hand tracking [11], combined with modeling the objects as real time point cloud. Point cloud has been one of the major approaches for object representation in robotics, therefore our approach can benefit from several well developed algorithms. Most importantly, this framework would extend our knowledge of human grasping quantitatively to the domain of human object recognition and grasp planning.

II. Method

A. Hand tracking

There are three main options for tracking whole-hand kinematics and they all have their advantages and drawbacks. While allowing the most natural movements, computer vision based tracking is the least accurate and is sensitive to environmental factors. Data gloves are easy to setup, but they are not very accurate if one wishes to reconstruct the spatial distribution of the digits. This is because they use joint angle sensors, and therefore errors that accumulate across many sensors lead to large errors in position estimation. We chose to use marker-based tracking because of its ability to provide the most accurate estimation of the positions of the tips and joints, both of which are essential for estimation of hand-object contacts. Specifically, we used a whole-hand tracking scheme based on Extended Kalman Filter that takes advantage of recursive estimation to reduce the effect of noise and marker occlusions. It consists of 24 markers and is capable of estimating 29 degrees of freedom (DoF) of the entire forearm, wrist, and hand sampled at a frequency of 200 Hz. The mean accuracy, measured in tip-to-tip tests, is 3 mm (see [11] for details).

Manuscript received March 26, 2011. This work was supported in part by a National Science Foundation Collaborative Research Grant IIS 0904504.

Q. Fu is with School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287 USA. (phone: 623-262-0769; e-mail: qiushifu@asu.edu).

M. Santello is with School of Biological and Health Systems Engineering, Arizona State University, Tempe, AZ 85287 USA. (e-mail: marco.santello@asu.edu).

B. Object modeling and tracking

The most important aspect of our kinematic tracking framework is to synchronize tracking of object geometry *and* the hand. We model the target object as point cloud $P_i(t)$ which represents the outer surface of the object (Fig. 1). There are multiple ways to obtain P_i in real time, for example, using depth sensors. For demonstration purpose, however, here we obtained point cloud p_i in an object frame of reference for each object prior to the experiment by manually creating CAD model of the objects. During the experiments, markers were attached to the objects to give a real time estimation of the 6 DoF pose of the object. The pose can be represented as a time dependent rotation matrix $R(t)$ and a translation vector $P(t)$ that transform the object center point cloud p_i into a global frame of reference.

$$P_i(t) = R(t)\, p_i + P(t) \qquad (1)$$

Note that this synthetic point cloud will be replaced with real time data from depth sensors for more intensive data collection to reduce the effort of manual construction of object models.

C. Contact site detection

With both hand and object being tracked, it is possible to estimate the points of contact when the hand is grasping the object by modeling it as a collision detection problem between the contour of the hand and the contour of the object. The phalanges can be modeled as cylinders and the joints can be modeled as spheres. It is also possible to generate a mesh that wraps around all joints and phalanges as the surface of the hand. In the present work, for simplicity, we only model the hand as a collection of 20 spheres that are located at the joint center as well as centers of the finger tips, J_k , with radius r_k equal to half of the measured joint thickness (Fig. 1).

To estimate contact points, we calculate the distance between P_i and J_k:

$$D_{ik} = \| P_i - J_k \| \qquad (2)$$

For each joint, if there's P_i satisfies

$$D_{ik} < r_k + h \qquad (3)$$

this joint is defined to be not in contact with the object, where h is a threshold that compensates for noise and tracking error. For each joint that is contacting the object, there would be a collection of points $Q_j \in \{P_i\}$ satisfying equation 3. We define the point of contact of joint k as $C_k \in \{Q_j\}$ which minimize the distance $D_{jk} = \| Q_j - J_k \|$. This equation essentially determines the point from the object point cloud that is the closest to each joint center that is in contact. Note that this simple collision detection algorithm can be replaced by more efficient algorithms if a real time application is necessary.

D. Hand enclosing space and graspable features

We define *hand enclosing space* as the space enclosed by the parts of hand that are used to grasp the object. The quantification of hand enclosing space is important because it can distinguish among grasp postures that might have similar hand kinematics but different contact areas, thus providing additional information about hand-object interactions. For instance, a side pinch (holding a key) is very similar to wrapping around a thin stick (holding a fork). However, the former grasp utilizes the side pad of the index finger for contact, whereas the latter utilizes all fingers for stability. Therefore, these two grasps are characterized by very different enclosing spaces. Most importantly, hand enclosing space could be further used to determine what parts of the object are enclosed by the hand.

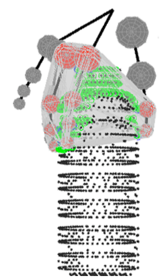

Figure 1. Grasping a bottle. Both the model of the hand and point cloud of the bottle are shown. The contacting joints are denoted by red spheres, the convex hull generated for hand enclosing space is shown in grey, and the green dots are graspable features.

The hand enclosing space is essentially a bounding region that is usually modeled by an ellipsoid or a polygon. After testing minimum bounding boxes, ellipsoids, and spheres, we eventually selected to use a convex hull that bounds all joint spheres that are in contact with the object (Fig.1). Each joint sphere is approximated by 64 vertices. Convex hull is more compact than other bounding geometries and the volume it encloses best fits the space occupied by the parts of hand that are in contact.

Once the object is grasped, the geometry that is enclosed by the hand can be defined as a graspable feature, or affordance. This is quantified by testing whether a point from P_i is inside the convex hull. Therefore the graspable feature that is associated with a grasp movement is a subset of the object point cloud (Fig. 1) and can be used for similarity testing between different objects and grasp configurations.

E. Experimental setup

In the implementation of our tracking framework, we used a motion tracking system with active markers (Marker diameter ~1mm, position accuracy <1mm; PhaseSpace Inc., San Leandro, CA, U.S.). The marker positions were sampled at 120 Hz and automatically labeled by the system. The joint thickness was measured using a digital caliper for initial calibration [11]. The EKF tracking scheme and contact detection was implemented in MATLAB as post processing.

Three subjects participated in the experiment and each of them performed 2 tasks. The first task was designed to determine the accuracy of digit contact detection, in which subjects were instructed to reach, grasp, and lift an inverted

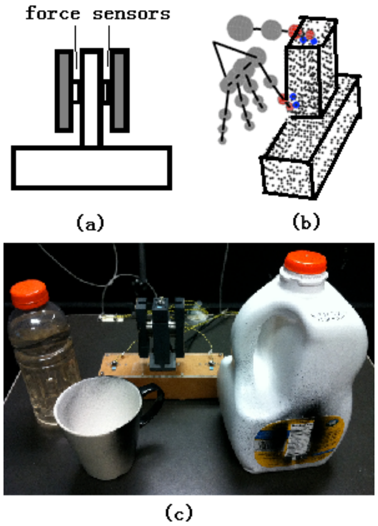

Figure 2. Objects used in tasks. (a) Front view of the inverted T-shaped object embedded with force sensors, (b) representative data from one trial in which subject used his thumb and index finger to grasp and lift the object. The red spheres denote joints that are in contact with the object whereas blue dots are estimated contact point on the two force panels. The objects used for the experiments are

T-shaped object. This object is equipped with two force/torque sensors (ATI Nano-25, ATI Industrial Automation, Apex, NC, U.S.) each mounted under long panels (Fig. 2a). This setup allows measurement of the net center of pressure applied by the finger in contact with the panel with an accuracy of ~ 1mm [12]. Subjects were asked to grasp only on the panels connected to the force sensors using thumb and one of the fingers (Fig. 2b). The net center of pressure was calculated in two ways: one was based on force/torque measurement and the other was based on the contact point detection algorithm we proposed (Section C). Each subject performed five trials.

In the second task, subjects were instructed to grasp and lift three objects (bottle, mug, milk jug; Fig. 2c) five times. Note that, as done in our previous studies of grasp planning [12], subjects were not given instructions on where to grasp the objects, and therefore they could plan and choose contacts based on individual preferences, object geometry and properties, e.g., mass, mass distribution, and frictional properties. This second task was designed to further validate the detection of contact sites as well as the hand enclosing space and graspable feature in a natural setup by using object with different geometric features. All four objects used in both tasks had a pre-defined point cloud and were tracked with the PhaseSpace system to generate a synthetic real time point cloud synchronized with the hand tracking.

III. Experimental Results

A. Spatial accuracy of contact detection

The contact sites detected by our algorithm could be more than the tips of the pair of fingers involved in the task of grasping inverted-T object. For instance, two distal joints of both thumb and index were in contact with the object in Fig. 2b. In order to compare the contact detection algorithm with the data measured by force sensors, we computed the mean of all contact points associated with each digit. The error was quantified as the distance between kinematic estimation of the contact point and the force estimation of the center of pressure. The result of mean error is shown in Table 1. The maximum mean error was less than 5 mm.

TABLE I
Errors of Contact Site Estimation (mm)

Finger in contact	Subject 1	Subject 2	Subject 3
Thumb/Index	3.9/4.5	2.3/3.2	4.1/4.3
Thumb/Middle	2.1/3.7	3.5/4.2	4.2/2.6
Thumb/Ring	3.6/4.4	3.3/4.1	3.4/2.3
Thumb/Little	3.2/4.9	4.4/3.8	5.0/4.8

B. Tracking hand-object interaction

The detection of contact sites, enclosing spaces and graspable features were captured at object lift onset by evaluating the velocity of the object. The overall results appeared to be reasonable, although there were cases of missing joint contacts due to marker occlusion (Fig. 3). An interesting observation is that, even though subjects were instructed to grasp each object as they wished, the five graspable features generated by different subjects were highly consistent. Specifically, the bottle was often grasped from top or from the side using a precision or power grasp, respectively; the cup was often grasped at the handle, over the top, on the side, or on the rim; the jug was often grasped at the handle, cap, or sides (Fig. 3 shows one representative subject). These results further demonstrate that geometric cues, together with familiarity with the object's intended use and properties, can significantly constrain the way humans grasp objects.

IV. Conclusion and Future Work

The proposed framework is capable of capturing the complete kinematic description of grasping including the hand and the objects, as well as their interactions. Although we only tested 3 objects, our preliminary results are promising. Specifically, we found that our algorithm can reconstruct, with reasonable accuracy, the distribution of unconstrained contacts on widely different objects (task 1). Furthermore, our approach allows us to describe not only the hand postures that subjects use to grasp different objects, but also *where* the objects are grasped. The estimated contact sites can be used to quantify the grasp quality of human grasps using robotics-based indexes [13, 14], such as stability.

A major limitation of our tracking framework is its sensitivity to missing markers, a problem that is particularly severe for objects that require wrapping of the digits around specific object features, e.g., handles. This issue would further limit application of our framework to experiments performed on non-human primates because they have smaller hands than humans [16, 17]. A possible improvement is to use a hybrid dataglove and marker

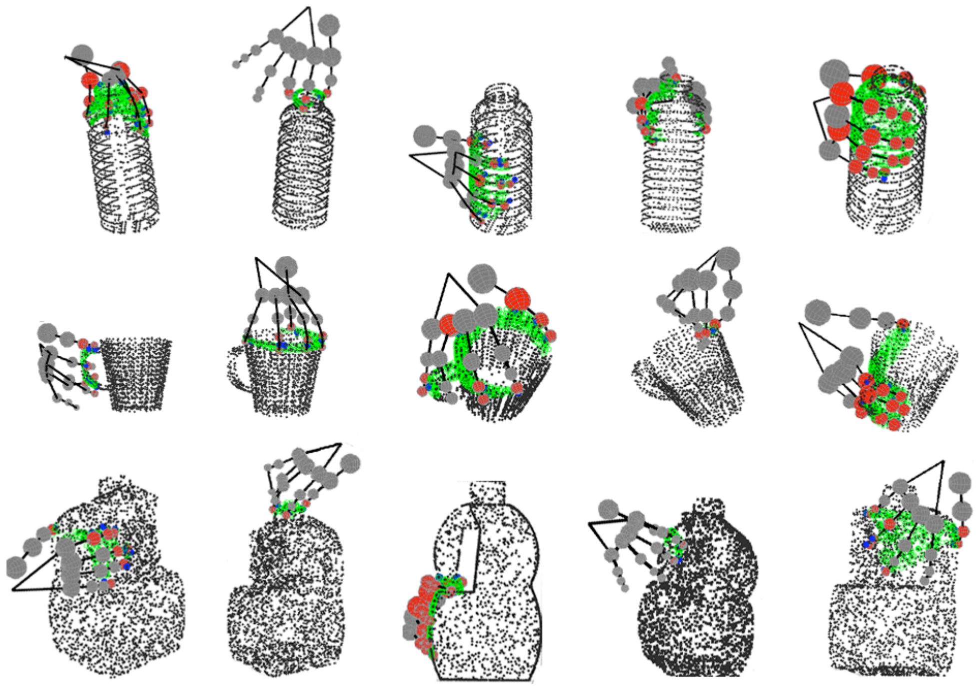

Fig. 3. Snapshots of one subject performing five different grasps on each of the three objects.

tracking scheme to minimize the effect of marker occlusion for specific grasp configuration. We are now working on optimizing our algorithms and generating a more automated data collection procedure, e.g., real-time depth sensors to eliminate the need of constructing object point cloud from CAD model.

We are aiming to generate a comprehensive human grasping database similar to Columbia Grasping Database used in robotics [15]. Our database would preserve the complex mapping between the hand and object geometry as well as its intended use, while allowing to quantify the sensory-to-motor interactions responsible for object representation and grasp planning. We believe that this work will also be a valuable tool as training data for robotic hands to generate 'human-like' grasps.

Acknowledgment

We thank Dr. Matei Ciocarlie for his valuable suggestions and Gabrielle Palermo for helping with data collection.

References

[1] M. Santello, M. Flanders, and J. F. Soechting, "Postural hand synergies for tool use." *J. Neurosci.*, vol.18, 10105-10115, Dec. 1998.

[2] E. Todorov, "Analysis of the synergies underlying complex hand manipulation." *Proc. IEEE Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*, pp. 4637-4640, San Francisco, USA, 2004.

[3] P. H. Thakur, A. J. Bastian, and S. S. Hsiao, "Multidigit Movement Synergies of the Human Hand in an Unconstrained Haptic Exploration Task." *J. Neurosci.*, vol. 28, 1271-1281, Feb. 2008.

[4] A. H. Fagg, M. A. Arbib, "Modeling parietal-premotor interactions in primate control of grasping." *Neural Netw*. vol.11, pp. 1277-1303, Oct. 1998.

[5] M. Hueser, T. Baier, and J. Zhang, "Learning of demonstrated Grasping Skills by stereoscopic tracking of human hand configuration." *Proc. IEEE Int. Conf. Robot. Autom*, pp. 2795-2800, Florida, USA, 2006.

[6] C. de Granville, J. Southerland, A. H. Fagg, "Learning Grasp Affordances Through Human Demonstration." *Proc. Int. Conf. on Development and Learning*, Bloomington, USA, 2006.

[7] Charusta, K, D. Dimitrov, A. J. Lilienthal, B., Iliev, "Extraction of grasp-related features by human dual-hand object exploration." *Proc. Int. Conf. Adv. Robot.*, pp.1-6, Munich, Germany, 2009.

[8] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic Grasping of Novel Objects using Vision." *Int. J. Robot. Research*, vol. 27, pp. 157-173, Feb. 2008.

[9] A. Sahbani, S. El-Khoury, "A hybrid approach for grasping 3D objects." *Proc IEEE/RSJ Conf Intelligent Robots and Systems*, pp.1272-1277, St. Louis, MO, USA, 2009.

[10] R. Balasubramanian, L. Xu, P. D. Brook, J. R. Smith, and Y. Matsuoka, "Human-guided grasp measures improve grasp robustness on physical robot." *Proc. IEEE Int. Conf. Robot. Autom*, pp. 2294–2301, Anchorage, USA, 2010.

[11] Q. Fu and M. Santello, "Tracking whole hand kinematics using extended Kalman filter." *Proc. IEEE Annu. Int. Conf. IEEE Eng. Med. Biol. Soc*, pp.4606-4609, Buenos Aires, Argentina, 2010.

[12] Q. Fu, W. Zhang, and M. Santello, "Anticipatory planning and control of grasp positions and forces for dexterous two-digit manipulation." *J. Neurosci.*, vol. 30, pp. 9117-9126, June 2010.

[13] M. Gabiccini and A. Bicchi. "On the Role of Hand Synergies in the Optimal Choice of Grasping Forces." *Proc. Robotics: Science and Systems*, Zaragoza, Spain, 2010.

[14] M. Ciocarlie, H. Dang, J. Lukos, M. Santello, and P. Allen, "Functional analysis of finger contact locations during grasping." *Proc. 3rd Joint Eurohaptics Conf. and Symp. Haptic Interfaces for Virtual Env. and Teleoperator Syst.*, pp. 401-405, Salt Lake City, USA, 2009.

[15] C. Goldfeder, M. Ciocarlie, H. Dang, and P. K. Allen,"The Columbia grasp database." *Proc. IEEE Int. Conf. Robot. Autom*, pp. 1710-1716, Kobe, Japan, 2009.

[16] C. E. Vargas-Irwin, G. Shakhnarovich, P. Yadollahpour, J. M. K. Mislow, M. J. Black, and J. P. Donoghue, "Decoding Complete Reach and Grasp Actions from Local Primary Motor Cortex Populations." *J. Neurosci.*, vol.30, 30(29): 9659-9669, July 2010.

[17] M. Velliste, S. Perel, M. C. Spalding, A. S. Whitford, and A. B. Schwartz, "Cortical control of a prosthetic arm for self-feeding." *Nature*, 453:1098-101, June 2008.