Tracking hand-object interaction: towards a database for human grasping and manipulation

Qiushi Fu, Marco Santello

I. Introduction

Extensive effort has been devoted on tracking finger joint kinematics to quantify the high dimensional space of the human hand during grasping and manipulation [1]–[3]. It has been shown that finger movements can be represented in a low dimensional space. However, the kinematics of the hand is only one component of grasping tasks. The other important component of grasping is the object that is being interacted with. Surprisingly, the systematic identification of object graspable features has been largely overlooked in the investigation of human grasping. 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 mostly been studied through simple observations [5] and task scenarios [6], but not quantitative analysis using kinematic data of hand-object interactions. This has prevented further understanding of how object properties are represented and how grasping is planned in the central nervous system. These gaps stem from lack of efficient measurement of (a) where the object is grasped and (b) what parts of the hand make contact with the object.

The present work proposes a framework that could bridge these gaps and advance our understanding of human grasp planning and control. Specifically, the key objective of our study is to track the contours of the object and hand kinematics at the same time. This is achieved by modeling the objects as real time point clouds and implementation of collision detection algorithms. Most importantly, this framework would allow us to establish a database of human grasping behaviors by considering not only the joint space of the hand, but also where the object is being grasped.

II. Methods & Results

A. *Proof of concept: using marker based tracking approach*

To demonstrate the concept of tracking human-object interactions, we first tested an approach that is based on optical markers [7]. 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 [8] for details). We then model the target objects as point clouds from CAD models and three markers were attached to the real object at known locations, thus allowing real time estimation of the 6 DoF pose of the object that is synchronized with the hand tracking.

The estimation of the points of contact was essentially a collision detection problem between the contours of the hand and the contour of the object. Since we did not have a point cloud or mesh representation of the hand, we simply modelled the hand as a collection of spheres that are located at the joint centers and fingertips. The contacts were then estimated by computing the distance between the hand spheres and the object point clouds. The goal of this approach was to improve the accuracy of contact estimation when precision grasps using fingertips only were performed. By comparing our estimated contacts with the contact locations (center of pressure) obtained from force/torque sensors, we showed that the position error was ≤ 5 mm.

Furthermore, this framework was designed to estimate the object parts that are being grasped. We defined the hand enclosing space as the convex hull that bounds all hand spheres that are in contact with the object. The geometry that is within the hand enclosing space geometry (as a subset of the point cloud) was defined as a graspable feature, or affordance.

This setup was tested by asking three subjects to grasp and lift three objects (bottle, mug, milk jug) five times with different self-chosen grasp locations. Interestingly, we found that the five graspable features generated by different subjects were highly consistent. For instance, 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. 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.

By tracking both hand and object, we can estimate not only which part of the object is being grasped, but also how we reach to the object under different task conditions. A recent study [9] using this tracking framework examined how human compensate for orientation uncertainty of a target object (a cylinder) without online visual guidance. It was found that subjects did not try to minimize post contact hand posture adjustments, but rather maximized the probability of

Manuscript received February 4, 2013. This work was supported in part by a National Science Foundation Grant IIS 0904504.

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

initial contact within the grasp aperture. Therefore, our results suggest that the reaching movement was adjusted to compensate sensorimotor noise for more efficient sensing of actual object orientation.

B. Work in progress: using Microsoft Kinect to improve efficiency

Although providing satisfactory results, the major issue with marker=based approach is that it requires sophisticated setup which is extremely time consuming if a large set of objects were tested. Additionally, as we are primarily interested in the general properties of the parts being grasped and the direction of hand movement before grasping, it is not necessary to guarantee high precision. Therefore, we tested the feasibility of using Microsoft Kinect as a low-cost and fast solution to implement our tracking framework.

There have been many Kinect-based hand tracking algorithms, but most of them focus on gesture recognition, which does not work well in hand-object interactions since occlusion often occurs. Recently, an advanced method has been developed using particle swarm optimization [10]. The occlusions were solved by modeling physical constrains between hand and the objects [11]. Moreover, hand-object interaction can be tracked by including the object model with hypothesis-and-test approach [12], which is used with Bayesian inference to predict human actions on a given object. For simplicity, we implemented a skin color based tracking algorithm for hand tracking (only the point cloud is tracked, but not the hand kinematics), a particle filter based object tracking algorithm, and an octree-based collision detection algorithm [13]. These implementations allowed us to measure how human subjects interact with a large set of common objects in their natural way efficiently without the interference from the markers or gloves. For further analysis of the graspable features from variety of objects, ongoing experiments are conducted to generate common graspable features for common object set.

III. Future directions

In the field of robotics, there have been attempts 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 learn how and where to grasp objects. 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 [14]–[16], but without considering the geometry of the object; the other is to label graspable features heuristically, such as grasping points [17] and graspable parts [18], but the hand posture is generated computationally. Additionally, the knowledge of human hand-object interaction would help generate more robust and human-like grasps for robots [19]. Therefore, a better understanding of where human make contact with the object in common scenarios could potentially advance grasp planning for robots. Furthermore, we propose that our tracking framework would provide a database of human grasping could be used as benchmark and/or training data for robotic grasping, in similar fashion as computationally generated grasp database [20].

References

[1] M. Santello, M. Flanders, and J. F. Soechting, “Postural hand synergies for tool use,” *J. Neurosci.*, vol. 18, no. 23, pp. 10105–10115, 1998.

[2] E. Todorov and Z. Ghahramani, “Analysis of the synergies underlying complex hand manipulation.,” in *Annual International Conference of the IEEE EMBS*, 2004, vol. 6, pp. 4637–40.

[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, no. 6, pp. 1271–1281, 2008.

[4] A. H. Fagg and M. A. Arbib, “Modeling parietal-premotor interactions in primate control of grasping,” *Neural Networks*, vol. 11, no. 7–8, pp. 1277–1303, 1998.

[5] N. Kamakura, M. Matsuo, H. Ishii, F. Mitsuboshi, and Y. Miura, “Patterns of static prehension in normal hands,” *J. Occup. Ther.*, vol. 34, no. 7, pp. 437–445, 1980.

[6] U. Kleinholdermann, V. H. Franz, and K. R. Gegenfurtner, “Human grasp point selection,” *J. Vis.*, vol. 13, no. 8, pp. 1–12, 2013.

[7] Q. Fu and M. Santello, “Towards a complete description of grasping kinematics: a framework for quantifying human grasping and manipulation,” in *IEEE Engineering in Medicine and Biology Society*, 2011, pp. 8247–50.

[8] Q. Fu and M. Santello, “Tracking whole hand kinematics using extended Kalman filter.,” in *IEEE Engineering in Medicine and Biology Society*, 2010, vol. 2010, pp. 4606–9.

[9] Q. Fu, A. Ushani, L. P. Jentoft, R. D. Howe, and M. Santello, “Human reach-to-grasp compensation with object pose uncertainty,” *IEEE Eng. Med. Biol. Soc.*, 2013.

[10] I. Oikonomidis, N. Kyriazis, and A. A. Argyros, “Efficient model-based 3D tracking of hand articulations using Kinect,” *Procedings Br. Mach. Vis. Conf. 2011*, pp. 101.1–101.11, 2011.

[11] I. Oikonomidis, N. Kyriazis, and A. A. Argyros, “Full dof tracking of a hand interacting with an object by modeling occlusions and physical constraints,” in *13th IEEE International Conference on Computer Vision*, 2011.

[12] D. Song, N. Kyriazis, I. Oikonomidis, C. Papazov, A. A. Argyros, D. Burschka, and D. Kragic, “Predicting Human Intention in Visual Observations of Hand / Object Interactions,” in *IEEE International Conference on Robotics and Automation*, 2013.

[13] R. B. Rusu and S. Cousins, “ICRA Communications 3D is here : Point Cloud Library (PCL),” in *IEEE International Conference on Robotics and Automation*, 2011, pp. 1–4.

[14] M. Hueser, T. Baier, and J. Zhang, “Learning of demonstrated grasping skills by stereoscopic tracking of human head configuration,” in *IEEE International Conference on Robotics and Automation*, 2006, no. May, pp. 2795–800.

[15] C. De Granville and A. H. Fagg, “Learning grasp affordances through human demonstration,” in *International Conference on Development and Learning*, 2006.

[16] K. Charusta, D. Dimitrov, A. J. Lilienthal, and B. Iliev, “Extraction of grasp-related features by human dual-hand object exploration,” in *International Conference on Advanced Robotics*, 2009, pp. 1–6.

[17] A. Saxena, J. Driemeyer, and A. Y. Ng, “Robotic grasping of novel objects using vision,” *Int. J. Rob. Res.*, vol. 27, no. 2, p. 157, 2008.

[18] S. El-Khoury and A. Sahbani, “Handling objects by their handles,” in *IEEE/RSJ International Conference on Intelligent Robots and Systems, Workshop on grasp and task learning by imitation*, 2008, pp. 58–64.

[19] R. Balasubramanian, L. Xu, P. D. Brook, J. R. Smith, and Y. Matsuoka, “Human-guided grasp measures improve grasp robustness on physical robot,” in *IEEE International Conference on Robotics and Automation*, 2010, pp. 2294–2301.

[20] C. Goldfeder, M. T. Ciocarlie, H. Dang, and P. K. Allen, “The Columbia grasp database,” in *IEEE International Conference on Robotics and Automation*, 2009, pp. 1710–1716.