A Comprehensive Method for Power Grasp on Incomplete Point Clouds

Ilaria Gori*, Ugo Pattacini*, Vadim Tikhanoff* and Giorgio Metta*
*iCub Facility - Istituto Italiano di Tecnologia

I. System Overview

Manipulation is a fertile research theme in robotics; we refer the reader to [5] for a complete overview of the topic. We present an effective method to grasp objects in real environments. The objects to grasp are unknown and perceived partially, causing uncertainty. We use only the visible part of the object. The target system is a humanoid robot equipped with stereo vision. The main contribution is the developing of a complete pipeline, which uses visual input to gain information about the object, but takes into account also the kinematic constraints of the robot's body, as well as simple tasks that can be requested to the robot. We show that, even though we use very simple information, we are able to obtain high accuracy in the grasping procedure. We focus specifically on power grasp actions.

1) Reconstructing and Segmenting the Point Cloud: We assume that the object to grasp lies on a table in front of the robot's cameras. We also assume that it is not too big to be grasped by the robot's hand. Each pixel of the object is projected in 3D via the Hirshmuller algorithm [1], which results in an incomplete point cloud of the visible portion of the object. We compute the minimum enclosing bounding box of the point cloud through a rotating calipers-based algorithm, estimating the size of the object. We then find smooth surfaces sufficiently large for the robot to place its hand before closing the fingers. To this end we apply the Region Growing Segmentation algorithm [4]. It does not need any initialization procedure and it works in real time. From the best regions we extract candidate grasping points.

2) Evaluating the Points: In order to choose a suitable position for the robot hand, we build an appropriate score function s to rank candidate points. This score function is composed of two contributions with respect to the point p to be evaluated: $s(p) = w_1 \cdot t(p) + w_2 \cdot v(p)$. The first fixed part $t(p)$ depends on the task that the robot has to fulfill when grasping the object. Since the objects we are dealing with are novel to the robot, there exist a few generic tasks that can be defined. In this case we can think of "pass the object to a person", or "explore the object". These tasks can be easily modeled giving higher scores to the points that lie in specific portions of the object. There are also more general conditions that can be considered, such as the fact that the hand is better placed far from edges or corners; this condition assures that points lying in the last visible part of the objects, whose normals are usually not accurate, are not taken into account. Fig. 2 shows the point clouds (one complete and one incomplete) in the first column, and two point extraction sessions with different $t(p)$ in the other two columns. The part $v(p)$ of the score function $s(p)$ is auto-adjustable and assigns higher scores to points with specific curvature values on the object surface. These curvatures are meant to produce better grasp performances as result of the experience the robot can continuously gain during his trials and errors. We learn a function that maps input curvature values to grasp success through regression using Least Square Support Vector Machine (LSSVM) [6], which is incrementally updated as the robot performs new actions. The grasp success measurement is based on the Euclidean distance between the configuration commanded by the system and the actual configuration normalized by the number of points belonging to the region that is taken into account: larger differences mean that the fingers encountered the object and deviated from the intended configuration. The weights of the function are chosen empirically, on the basis of the user's needs. The overall score function votes for the best n candidate grasping points.

3) Choosing the Best Configuration - End-effector Position and Orientation: We assess each candidate point out of the n best points. We first aim to have the robot hand aligned with the plane that locally approximates the object surface in the candidate point: given the robot kinematics (see Fig. 3, RGB color convention is used for representing the x, y, and z axes respectively), we require that the z axis in the hand reference frame is orthogonal to the tangent plane. We then sample the plane orthogonal to the z axis and containing the candidate point p by identifying m possible orientations for the axes x and y (see Fig. 3). For each point and each sampled orientation we run an inverse kinematics algorithm [3] to retrieve the corresponding arm configuration. We then examine all the possible joint configurations according to the manipulability measure [7]. The hand position and orientation that produce the best manipulability measure are finally selected.

II. Experiments

All the experiments were carried out on the iCub humanoid robot [2]. We build a training set of 100 data points, grasping the objects shown in Fig. 4, and changing randomly the curvature values in order to explore the function domain, still choosing the configuration with maximum manipulability. The learned map $v(p)$ between the local curvatures of the object and the grasp success is depicted in Fig. 4. From the image it is possible to deduce that points with too low or too high curvatures are likely to bring to unsuccessful grasps, and that there are two intermediate curvatures that seem to be suitable for the iCub's hand. In order to prove that the learned model can generalize over unknown values, we apply a k-fold cross validation procedure with $k = 5$ using 75 samples for training and 25 for testing; we then estimate on the test set the error between the predicted values of the function and the real grasp success measurement on the test set. The overall Mean Squared Error (MSE) is equal to 0.0028 radians. To verify whether the peaks of the function identify critical curvatures that would entail a gain in the grasp success rate, we run two grasping sessions on a sixth object shown in Fig. 4. In the first session, we let the robot perform 20 grasps choosing points with curvature close to the values that are likely to bring about unsuccessful grasps, obtaining a grasp success rate of 60%. In the second session, additional 20 grasps are executed, and points having curvature values close to the peaks are rewarded more; this second session brings to an accuracy of 85%. For a quantitative assessment of the **overall** grasping procedure, we executed 20 trials on each object depicted in Fig. 4. We refer to a grasp as successful if the robot is able to hold the object firmly without dropping it. We achieved a success rate of 84% over 100 grasping trials, which makes the framework suitable for robust manipulation. Notably, most of the failed grasps were due to a non accurate hand-eye calibration.

REFERENCES

[1] H. Hirschmuller. Stereo processing by semiglobal matching and mutual information. *TPAMI*, 2008.

[2] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori. The icub humanoid robot: an open platform for research in embodied cognition. In *8th Workshop on Performance Metrics for Intelligent Systems*, 2008.

[3] U. Pattacini, F. Nori, L. Natale, G. Metta, and Sandini G. An experimental evaluation of a novel minimum-jerk cartesian controller for humanoid robots. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1668–1674, 2010.

[4] T. Rabbani, F. van den Heuvel, and G. Vosselmann. Segmentation of point clouds using smoothness constraint. *ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*, 2006.

[5] A. Sahbani and S. El-Khouri. An overview of 3D object grasp synthesis algorithms. *Robotics and Autonomous Systems*, 2011.

[6] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, and J. Vandewalle. *Least Squares Support Vector Machines*. World Scientific Pub. Co., 2002.

[7] T. Yoshikawa. Manipulability of robotic mechanisms. *The International Journal of Robotics Research*, 4(2):3–9, 1985.