Robotics and Autonomous Systems 62 (2014) 1742–1754

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

journal homepage: www.elsevier.com/locate/robot

Grasping objects localized from uncertain point cloud data

Jean-Philippe Saut [a], Serena Ivaldi [a,b,*], Anis Sahbani [a], Philippe Bidaud [a]

[a] Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie, ISIR - CNRS UMR 7222, Boite courrier 173, 4 Place Jussieu, 75252 Paris cedex 05, France

[b] Intelligent Autonomous Systems Lab., TU Darmstadt, Germany

HIGHLIGHTS

- We propose a grasp planning approach for an object with a known shape, observed as a point cloud.
- We focus on uncertainties in the object pose distribution from a noisy point cloud.
- Experiments are conducted with the humanoid robot iCub and its stereo cameras.

ARTICLE INFO

Article history:
Received 18 September 2013
Received in revised form
20 July 2014
Accepted 28 July 2014
Available online 12 August 2014

Keywords:
Robotic grasping
Multi-fingered hand
Inverse kinematics

ABSTRACT

Robotic grasping is very sensitive to how accurate is the pose estimation of the object to grasp. Even a small error in the estimated pose may cause the planned grasp to fail. Several methods for robust grasp planning exploit the object geometry or tactile sensor feedback. However, object pose range estimation introduces specific uncertainties that can also be exploited to choose more robust grasps. We present a grasp planning method that explicitly considers the uncertainties on the visually-estimated object pose. We assume a known shape (e.g. primitive shape or triangle mesh), observed as a – possibly sparse – point cloud. The measured points are usually not uniformly distributed over the surface as the object is seen from a particular viewpoint; additionally this non-uniformity can be the result of heterogeneous textures over the object surface, when using stereo-vision algorithms based on robust feature-point matching. Consequently the pose estimation may be more accurate in some directions and contain unavoidable ambiguities.

The proposed grasp planner is based on a particle filter to estimate the object probability distribution as a discrete set. We show that, for grasping, some ambiguities are less unfavorable so the distribution can be used to select robust grasps. Some experiments are presented with the humanoid robot iCub and its stereo cameras.

1. Introduction

In this paper we focus on grasping under uncertainty: precisely, we propose a method that takes explicitly into account the uncertainty in the object observations to compute the best grasping configuration. Our objective in this study is to improve the grasping capabilities of the humanoid iCub that is equipped with two low-resolution stereo cameras and non-sensorized fingers (see Fig. 1).

We present a planner that computes grasp configurations from the observation, as a set of 3D points (point cloud), of an object with known shape (e.g. modeled as a primitive shape or a triangle mesh). The point set can be provided by means of any existing range scanning method. Conducted with a humanoid robot, the presented experiments are based on stereo-vision with robust 2D-features matching, but the approach would apply as well to data obtained with a device like Microsoft's Kinect or a time-of-flight camera. The point cloud is generally not dense enough and too locally-distributed to lead to a unique pose estimation, not talking about noise. The points sparsity can be the result of the small size of the object or of an insufficient number of points to match in case of stereo-vision. The points being not well distributed is because the object is seen from its front with respect to the robot sensing device and also to the non-uniform distribution of features to match in case of stereo-vision. To use the non-uniqueness of the solution of the pose estimation problem, we choose to reason with a probability distribution of the object pose instead of one pose

* Corresponding author at: Institut des Systèmes Intelligents et de Robotique, Université Pierre et Marie Curie, ISIR - CNRS UMR 7222, Boite courrier 173, 4 Place Jussieu, 75252 Paris cedex 05, France. Tel.: +33 0144279623.
E-mail address: serena.ivaldi@isir.upmc.fr (S. Ivaldi).

http://dx.doi.org/10.1016/j.robot.2014.07.011
0921-8890/

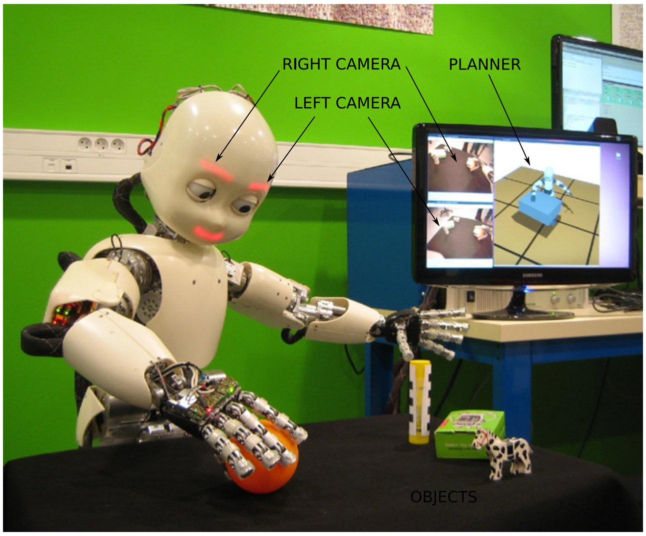

Fig. 1. The humanoid robot iCub grasping objects localized by its eyes' cameras.

estimation. The pose distribution is estimated through the use of an annealed particle filter. A discrete set of grasps is computed from the object geometry (known by hypothesis) and a set of uniformly-distributed hand poses (with respect to the object). A score is then computed for each grasp that depends on the previously computed pose distribution and is meant to evaluate how robust is the grasp with respect to the uncertainty in the object pose.

The individual components of the method are not new, however their combination is unique to our knowledge. In particular, addressing the uncertainties related to a sparse stereo point cloud together with grasp planning has not been addressed earlier in a similar fashion. Although projected light solutions such as Kinect have gained significant attention lately, stereo processing still has important application areas. For example, in multi-robot systems where the overlap of multiple projected light sensors often causes difficulties, and in humanoid robotics where most vision relies on the embedded robot's eyes/cameras.

Our method addresses a major problem in autonomous robotic grasping. The error in the object pose estimation is inevitable, since the cameras can have low resolution (it is the case for our robot) and their calibration does not compensate perfectly the distortion in the whole image. When an object is observed by matching the features in the two camera images, the resulting point set is small, sparse and the quality of observations is very noisy, especially if compared with the point cloud of an external Kinect. The localization error can be thus greater than the one we would have with such external sensor. In our robot, this error cannot be compensated reactively by tactile feedback, because the fingers are not equipped with tactile sensors, nor visual feedback, because of the visual occlusions of the robot's body during the grasp.

The paper's contribution is explained more in detail in Section 2 that also gives an overview of its principles. Next subsection presents related works proposed in the literature.

1.1. Related works

Considering uncertainty in robotic grasping strategies has known a revival these last years, due to the availability of many manipulation platforms usable in practical situations. As more and more robots are able to realize manipulation tasks, the goal is now to improve their robustness against all kinds of perturbations. The intrinsic sensor inaccuracy leads to uncertain world models, which frequently cause failure of robotic tasks involving interaction with the robot's environment; typically during a grasping task Weisz and Allen [1].

The first works dealing with robotic grasping and uncertainty were published during the 80's. Brost [2] presented a method to grasp a polygonal planar object with a parallel-jaw gripper, in case the object location is not known precisely. A sequence of squeezing operations is planned in order to progressively reduce the possible uncertainty on the pose of the object. Each time the grasp is fully closed on the object, the current aperture of the gripper is used to refine the knowledge on the possible object poses. Lozano-Pérez et al. [3] proposed a planning strategy for peg-in-hole insertion in the presence of uncertainty on the robot motion. An important idea of [3], which is useful in grasping in general, is the concept of pre-image. A pre-image is a set of points from which the goal can be attained in a single motion, like e.g. a simple translation. From the goal, a set of pre-images can be found, which represents the positions that can reach the goal in one motion. New sets of pre-images are iteratively developed until the start position belongs to a computed pre-image. Another way to deal with uncertainty is to design grasp quality scores that consider robustness against contact positioning error. In this purpose, the notion of *Independent Contact Regions* (ICR) was introduced by Nguyen [4]. ICRs are regions on the object surface – one per contact of the grasp – such that as long as each contact stays in its ICR, the grasp remains stable (i.e., verifies the force-closure property).

In the recent years, the interest in grasping and uncertainty has reappeared with a focus on real applications. Berenson et al. [5] represent the goal of grasping an object by one or several *Task Space Regions* (TSR). A TSR is a volume in SE(3), defined with bounds on the allowable translations and rotations. The uncertainness on the object position is represented as a set of possible poses. These different poses are used to duplicate and rotate the initial TSRs. The intersection of the TSRs defines then a new TSR that is the new goal region. Besides, a new collision model is built by duplicating the object model for each of the possible poses and merging the duplicates into one collision model.

The study of the relation between contact location uncertainty and grasp stability has also been carried on, with the design of more general and more practical algorithms to compute the ICRs of a grasp [6,7], working with any kind of shapes, simply described as points sets with neighborhood information. The size of the ICRs of a grasp can tell how precise must be the placement of the contacts in order to guarantee the grasp stability.

Other approaches, based on machine learning with a learning database obtained in realistic conditions in terms of variety and noise, were also introduced in the literature. Stulp et al. [8] propose to learn the motion primitives that maximize the chance to successfully reach and grasp an object uncertainly localized. The object is placed on a table according to a particular distribution. Several grasping trials are performed by the robot in order to improve its reaching trajectory via a reinforcement learning procedure. It is shown that the robot adapts its motion to perform better for the considered distribution. Other works based on machine learning have also been proposed to deal with partial information. Saxena et al. [9] use supervised learning to estimate the probability of a point, in an image taken by the robot's camera, to correspond to a good grasping point on the object. A set of visual features is computed for each image point. A logistic regression is used to model the probability of a point to lead to a successful grasp as a function of its visual features. An additional classifier considers a point cloud of the scene, provided by a depth sensor, in order to estimate the quality of a grasp [10].

If humans have some hidden strategy to grasp uncertainly-located objects, it might be interesting to learn from human demonstration. Faria et al. [11] proposed a grasping framework based on learning by demonstration from humans. Objects are