Robotics and Autonomous Systems 108 (2018) 140–152

Contents lists available at ScienceDirect

Robotics and Autonomous Systems

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

A learning framework for semantic reach-to-grasp tasks integrating machine learning and optimization

Zhen Deng [a], Xiaoxiang Zheng [a], Liwei Zhang [a,b,*], Jianwei Zhang [a]

[a] TAMS, Department of Informatics, University of Hamburg, D-22527, Hamburg, Germany
[b] School of Mechanical Engineering and Automation, Fuzhou University, Fujian, PR China

HIGHLIGHTS

- We proposed a learning framework for the implementation of semantic RTG tasks.
- A Bayesian-based search algorithm is designed for grasp planning.
- A model-based trajectory generation method is designed to generate reaching movement.

ARTICLE INFO

Article history:
Received 22 December 2017
Received in revised form 2 July 2018
Accepted 1 August 2018
Available online 14 August 2018

Keywords:
Semantic reach-to-grasp
Deep learning
Bayesian optimization
Model-based trajectory generation

ABSTRACT

The ability to implement semantic Reach-to-grasp (RTG) tasks successfully is a crucial skill for robots. Given unknown objects in an unstructured environment, finding an feasible grasp configuration and generating a constraint-satisfied trajectory to reach it are challenging. In this paper, a learning framework which combines semantic grasp planning with trajectory generation is presented to implement semantic RTG tasks. Firstly, the object of interest is detected by using an object detection model trained by deep learning. A Bayesian-based search algorithm is proposed to find the grasp configuration with highest probability of success from the segmented image of the object using a trained quality network. Secondly, for robotic reaching movements, a model-based trajectory generation method inspired by the human internal model theory is designed to generate a constraint-satisfied trajectory. Finally, the presented framework is validated both in comparative analysis and on real-world experiments. Experimental results demonstrated that the proposed learning framework enables the robots to implement semantic RTG tasks in unstructured environments.

1. Introduction

The rapid development in robotic research has enabled the robots to work autonomously in unstructured environments. One essential ability for operating autonomously is to implement semantic RTG tasks stably [1,2]. There are three main subtasks required to complete this task: detecting the object of interest from the table, determining an feasible grasp configuration and generating a constraint-satisfied trajectory to reach it. Therefore, the implementation of semantic RTG tasks contains typically three main aspects: object detection, grasp planning and trajectory generation. Due to the lack of object models and uncertainties from perceptual noise or kinematic errors of the robots, the stable implementation of semantic RTG tasks still remains an open problem.

In order to grasp the object of interest, an feasible grasp configuration is required. Typically, grasp planning is formulated as a constrained optimization problem. A number of grasp planning methods have been proposed [3–6]. Most of these methods employ numerical optimization [6] or search algorithms [5] to solve it. Most of these methods suffer from two drawbacks. First, these methods rely on accurate object models or good grasping examples. However, objects in unstructured environments are always partially perceived due to visibility constraints. Thus, these methods are limited to grasp known or familiar objects and difficult to generalize to unknown objects. Second, these methods use a handcrafted quality metric to evaluate the performance of grasps. Since the grasping process is intrinsically complicated, the quality metric is typically difficult to construct. Their performance decreases when implementing in unstructured environments. In

* Corresponding author at: School of Mechanical Engineering and Automation, Fuzhou University, Fujian, PR China.

E-mail addresses: deng@informatik.uni-hamburg.de (Z. Deng), zheng@informatik.uni-hamburg.de (X. Zheng), lw.zhang@fzu.edu.cn (L. Zhang), zhang@informatik.uni-hamburg.de (J. Zhang).

https://doi.org/10.1016/j.robot.2018.08.001
0921-8890/

this paper, the objective is to design a semantic grasp planning approach, especially considering unknown objects and uncertainties in an unstructured environment.

Recently, several interesting approaches have been introduced that use deep learning techniques for grasp planning [7–9]. These approaches formulate grasp planning as a detection problem, wh ere a Deep convolutional Neural Network (DCNN) is trained to build a mapping function from images to grasp configurations. However, most of these methods make a strong assumption that each object contains a single grasp configuration, which is not robust to uncertainties. Instead of directly predicting a grasp configuration from the image, the proposed grasp planning method first learns a quality function using DCNN and employ a Bayesian-based search algorithm to find the grasp configuration with highest quality. In this way, the uncertainties are taken into account and the grasp with highest quality is found. Most similar to us are [10–12]. However, there are two differences between the mentioned methods and the proposed method. First, the proposed method take advantage of the object semantic information, i.e., object identity and location, to assist the grasp planning. Second, we employ a Bayesian-based search algorithm instead of random searching algorithms [11] to determine the grasp with highest quality iteratively. That is important to grasp the object of interest stably.

Another key challenge is to generate a constraint-satisfied trajectory that drives a gripper to reach the object of interest. Although humans make a reaching movement effortlessly, it is difficult for the robots, due to the changing kinematics and the dynamic environment. In order to mimic human moving behaviour, a variety of methods has been presented, which use a linear dynamic model [13,14], a probabilistic model [15] or a cost function [16] to represent human demonstrated trajectories. Although these methods do well in some motion planning applications, these methods are limited to generate simple movements and have a poor generalization ability. However, in semantic RTG tasks, the trajectory generation method is required to generalize quickly to new scenes and adapt to the changing environment.

Studies of neuroscience demonstrated that human motor control is based on an internal model theory that the human maintains a forward dynamic model that predicts the consequences of motor commands and employs an inverse model to produce motor commands in order to achieve the desired movement [17,18]. Inspired by this theory, a model-based trajectory generation method is proposed in this paper, where a forward dynamic model in Cartesian space is learned from human multiple demonstrations in order to reproduce the state transition behaviour of the human arm and an inverse model constructed as a linear Gaussian model is online learned to produce robot control inputs. Compared with the previous trajectory generation methods [13,19,20], the proposed model-based trajectory generation approach is able to generalize quickly to new tasks.

This paper addresses the problem of the implementation of semantic RTG tasks in an unstructured situation. The objective is to reach to grasp the object of interest stably in a unstructured environment. The two key techniques, i.e., semantic grasp planning and trajectory generation, are studied in this paper. The following three contributions are presented.

(1) A learning framework which combines semantic grasp planning with trajectory generation is proposed for the implementation of semantic RTG tasks.

(2) A Bayesian-based search algorithm is proposed for grasp planning. Compared with random search algorithms, The proposed search algorithm has a chance to find more stable grasp configuration.

(3) Inspired by the human's internal model theory, a model-based trajectory generation method is designed to generate con straint-satisfied reaching movement for the robots.

The rest of the paper is organized as follows: Section 2 provides an overview of related works. Section 3 introduces the proposed learning framework that includes the semantic grasp planning and the model-based trajectory generation. Experiments and their results are presented in Section 4. Finally, the conclusion and future work are discussed in Section 5.

2. Related work

2.1. Grasp planning

Robot grasping is one of long-standing and important research topic in robotics. The overview introduced by Sahbani et al. [4] divides the existing grasp planning methods into two categories: analysis-based and data-driven methods. Analysis-based methods formulate grasp planning as a constrained optimization problem [21–23]. Since the grasping process is intrinsically complicated, the optimization objective and constraints are typically difficult to construct. Therefore, recent researches have mainly focused on data-driven methods [3]. For example, Kopicki et al. used kernel representation method to represent object shapes and employed the similarity measurement method based on kernel density to search an feasible grasp configuration. Detry et al. combined the dimensionality reduction with unsupervised clustering approach to search the grasp associated with the best-fitting part [5]. Herzog et al. learned grasp configurations for new objects from human demonstrations. The similarity is measured based on a template defined as a local shape descriptor [24]. These data-driven methods require a pre-defined a dataset of grasp examples that obtained from human demonstrations or simulations. These methods do well to grasp known or familiar objects, but always fail to grasp unknown objects [3].

Motivated by the recent success of deep learning, data-driven methods have attempted to incorporate deep learning into grasp planning. In this context, Redmon et al. trained DCNN to detect a grasp rectangle that represents the grasp configuration from an RGB image [8]. By defining different representation method of the grasp configuration, like a grasp region [25], a grasp rectangle [8] and a two-point grasp [26], some other similar detection approaches have also been presented. The similar ideas are also found in [7,27]. The strong assumption, i.e., each object contains a single grasp configuration, limits the practical application of these methods. Therefore, instead of predicting a grasp configuration directly from the object image, some studies first learned a quality function and then use it to determine the grasp configuration with highest quality [10–12]. In addition to establish an accuracy quality function, another key problem is how to search the grasp configuration with highest quality in a high-dimension grasping space. In this context, Mahler et al. randomly sampled a set of candidate grasps and rank them to find the grasp with highest quality [11]. Johns et al. learned a quality function considering gripper pose uncertainty and then use a random search algorithm to find the feasible grasp configuration [10] . In this paper, a Bayesian-based search algorithm is proposed to find the feasible grasp configuration. Compared with random search algorithms, the proposed search algorithm taken the uncertainties from perception noise and the search efficiency into account.

2.2. Semantic grasping

Semantic grasping is an important research direction in robotic grasping, where object semantic information is incorporated into grasp planning method [28]. The semantic information typically includes object identity [29], object affordances [30] or object shape knowledge [31]. A variety of research studies have demonstrated that object semantic information helps to improve the