Research Project Proposal: Physical perception from vision for Robotic Grasping

Luca Cavalli, luca3.cavalli@mail.polimi.it

1. Introduction

The problem of building a physical understanding of the environment around an agent is directly connected with the perception and planning of the real possibilities of actions with respect to the environment. Here we will focus on the grasping action, with the objective of framing a solid link between the robotic grasping itself and the more general field of affordances through physical modeling.

The Robotic Grasping research field tackles the problem of automating grasping actions on novel objects under different sensorimotor conditions. It follows the classical framing of sense, plan and act, thus the perception (intended as sensing and modeling the environment) is the basis on which the following steps must base.

The reason for connecting this field with affordance perception is that the real challenges and opportunities of grasping research can be better framed in a more general context. Since the first definition of affordances by Gibson in 1966 [3] a long discussion evolved, for a complete discussion refer to [6]. According to Michaels [4], affordances are emergent properties embodied in the relations between an animal and its environment directly connected with the possibility of action of the animal with the environment. Applied to robotics, affordance perception means understanding the possibility of action of the robot depending on the possible relations between its actuators and the environment to achieve high level tasks. In this context grasping represents an affordance for the control of some or all degrees of freedom of some object with a hand-like physical actuator.

The possibility of having control on some degrees of freedom of objects is fundamental for robotics applications as it is usually the main goal of actions. Moreover, in the wider context of affordances, task-oriented grasping enables the possibility of tool use, which in turns allows an enormous range of new possibilities of action: in the affordance community this goes under the name of *affordance chaining*. This synergy of grasping with its general context is not bounded to go in one single direction: also other kinds of affordance possibilities like pushing and pulling or the use of tools can derive different ways of grasping an object more effectively.

Providing a solid and general connection between high level planning of affordances and low level fine planning of grasps is still an open problem today, that goes under the name of *task oriented grasping*. Few works have been attempted, lacking a generalized framing of tasks either limiting task expressivity [7] or categorizing action possibilities [5]. Moreover, the modeling of physical quantities, which explain the connection between actions, tools and tasks, received almost no attention from the Robotic Grasping community and remains an unexplored opportunity.

We think that to achieve generality in this connection, a perceptual framework is needed to correctly model the key physical knowledge for grasping and the uncertainty associated to it.

2. Main related works

The idea of task oriented grasping is not new in the research community. A very common approach is to limit the generality of expressed tasks and focus on very specific tasks associated with some semantic. Dang et al. [1, 2] proposes a specification of task by means of semantic constraints which are extremely specific for each high-level task. Although effective, this approach is inherently unable to generalize to unknown or unconventional object uses, as the knowledge is hardcoded in the semantic constraints of each specific object-task couple.

A different approach is encoding the task as an objective in the physical domain, as proposed by Prats et al. [5]. They use this idea to effectively encode the tasks of interacting with common house objects (like doors,

windows, drawers), but they actually categorize grasp planning by using preshapes. This approach is effective when handling standard simple objects with handles designed on purpose, but cannot generalize to unknown objects: although the definition of the task is very general and synthetic, the grasping model associated to it is limited.

Another possibility is describing the task by demonstration, as proposed by Zhu et al. [7]. They propose a method to evaluate different physical dimensions associated to the usage of seen tools, and to extract the interesting features for the task from a video of a human performing the task. The physical dimension is involved as a significant feature for task understanding from the demonstration, which is a general, although not synthetic, representation. In this work, however, the authors concentrate much on task understanding and tool analysis and leave only a "grasping area" on the chosen tool as an indication of how to actualize the task with the tool. This representation is poor as it does not account to fundamental details such as hand direction, pose, and finger positioning for a solid hold.

3. Research plan

The final objective of this research is to provide a new method for task oriented grasping based on the physical understanding of the environment. The research plan is framed into three phases, where each phase has its own objectives, deliverables, and evaluation. Moreover, some intermediate milestones already represent novel results and thus can be object of publication. In the following we will detail all the phases and finally present possible future work.

3.1. Phase 0: World Representation

The objective of this preparatory phase is to build a model for the representation of arbitrary knowledge about space and individual objects and its related uncertainty. The output of this phase will be a working implementation of this model satisfying the following minimum requirements:

- the system will be integrated with at least one complete vision system (RGB-D camera or equivalent) and will be able to operate **online 3D reconstruction** of the environment and **estimate uncertainty** at any point.
- the system will be able to **identify separate objects** during the reconstruction.
- the system will allow the association of **arbitrary knowledge** and its uncertainty to both objects and spacial atomic points.
- the system will be able to **track rigid body motion** of objects and move all the information available accordingly.

The system will be based on a Bayesian framework to encode uncertainty while integrating new knowledge into the model via prior to posterior updates. Also, we plan to include trackable visual features like SIFT descriptors to effectively capture 3D rigid body motion and apply it onto the object model. The system will preferably process the input streaming in realtime, although this is not a requirement.
The implementation will be evaluated by the reconstruction accuracy of real world rigid objects whose surface is visible through motion and whose model is available as ground truth.

3.2. Phase 1: Physical Inference

The objective of this phase is the **dynamic estimation** of the **degrees of freedom** (DOFs) and **center of mass** of identified objects, which is a key feature to identify good grasps for a low level task. The output will be the integration of this functionality into the implementation provided by phase 0.

The expected path for this phase is to first build a model to infer prior estimations of interested quantities, and then build update rules to infer a posterior by observations of physical interactions in the world. In particular

for center of mass estimation we plan to employ pointcloud completion models to have a good shape prior and estimate the center of mass as the geometrical centroid as a first prior, assuming homogeneous density. Surface material classification models could be used to improve this prior estimation. Updates from observations will be done by detecting interesting physical events and by comparing the predictive distribution of observations as a function of the distribution of physical parameters with the actual observation. We plan to approach DOF estimation by finding translational and rotational invariants in the rigid movements that the object is observed to perform over time.

The evaluation of the quality of the output of this phase will consider DOFs and center of mass separately: DOFs estimation accuracy will be computed on a set of real objects displaying different DOFs, while center of mass estimation accuracy will be addressed on real objects quantifying also the improvement in accuracy after repeated demonstrations of physical interactions. Moreover, this phases already represents a milestone for a novel publication in the field of Computer Vision and Robotics.

3.3. Phase 2: Task-Oriented grasping

The third phase is aimed at demonstrating the effectiveness of a physical model for task-oriented grasping. The objective is to **define a limited set of elementary tasks** in terms of physical properties that can generalize well to express most high level tasks, and devising a search strategy which is able to **find appropriate grasps for these tasks**. The outcome will be an implementation of the search strategy on top of the physical inference framework. We plan to learn a model to infer probability distributions of successful grasps around the object to increase search efficiency, and to associate each task with soft constraints on the region and orientation of the grasp. These soft constraints can be used to bias the probability distribution of promising grasps to obtain a distribution of promising and task-coherent grasps.

With the availability of a real robotic arm, the results of this phase can be evaluated with real experiments on differently shaped objects, otherwise such experiments will be run on a simulator. This result represents a further milestone for a novel publication in the field of Robotic Grasping.

3.4. Future Work

The accomplishment of this research would open up many extension possibilities. The formulation of tasks as a limited but descrete set of atomic elements is compatible with an integration with logic inference to enable **chaining and reasoning** in high level planning. Moreover, the system can be integrated with different simple actions, like pushing, which can be used for **active perception** as they would cause physical interactions between robot and objects that can be observed to improve the robot knowledge of the environment.

3.5. Timeline

The whole research plan is expected to last seven and a half months and it is structured according to the three phases. Figure 1 shows the expected allocation of time, where each phase follows a cycle of detailed planning, actual implementation and evaluation assessment.

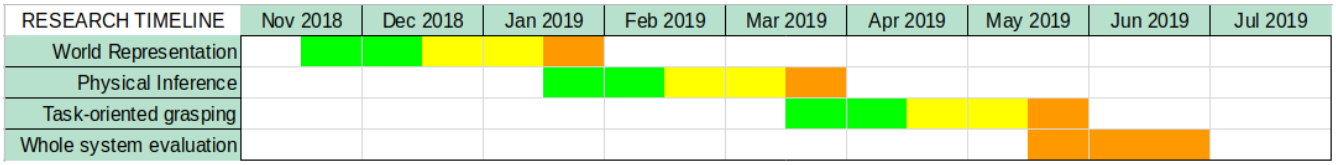

Figure 1: Research plan Gantt diagram. Green periods are dedicated to detailed planning and modeling, yellow periods to implementation, orange periods to evaluation assessment and documentation.

References

[1] Dang, H., and Allen, P. K. Semantic grasping: Planning robotic grasps functionally suitable for an object manipulation task. In *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on* (2012), IEEE, pp. 1311–1317.

[2] Dang, H., and Allen, P. K. Semantic grasping: planning task-specific stable robotic grasps. *Autonomous Robots 37*, 3 (2014), 301–316.

[3] Gibson, J. J. The senses considered as perceptual systems.

[4] Michaels, C. Affordances: Four points of debate. *ECOLOGICAL PSYCHOLOGY 15* (04 2003), 135–148.

[5] Prats, M., Sanz, P. J., and Del Pobil, A. P. Task-oriented grasping using hand preshapes and task frames. In *Robotics and Automation, 2007 IEEE International Conference on* (2007), IEEE, pp. 1794–1799.

[6] Zech, P., Haller, S., Lakani, S. R., Ridge, B., Ugur, E., and Piater, J. Computational models of affordance in robotics: a taxonomy and systematic classification. *Adaptive Behavior 25*, 5 (2017), 235–271.

[7] Zhu, Y., Zhao, Y., and Chun Zhu, S. Understanding tools: Task-oriented object modeling, learning and recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 2855–2864.