THE INSTITUTE OF INTELLIGENT SYSTEMS AND ROBOTICS*

INTERNSHIP OFFER

« Learning to grasp »

This thesis project is part of the ISIR federative project on agricultural robotics. It focuses on robotic input, particularly in an agricultural context (input of fruits or vegetables). This subject will be approached by a pluri-disciplinary approach between computer science (supervision S. Doncieux and A. Coninx), mechanics and automation (supervision F. Ben Amar).

Context: This thesis project deals with the grasping of objects by a robotic gripper. The goal is to design an automatic picking system for fruits and vegetables. It is thus a question of seizing fragile objects, of various forms and colors, objects which can moreover be at least partially occulted by the foliage of the vegetation. This thesis project will focus on a learning-based approach in which a significant expertise in mechanical modeling will be used at different levels of the methodological approach, from the design of the learning system to the analysis of its performance.

Object grasping is an emblematic task in robotics as it is a prerequisite for many other more advanced tasks such as object manipulation. From a learning point of view, it represents a particularly difficult challenge because it corresponds to a rare reward case: very few interactions are likely to lead to a correct grasp. Many movements will not even touch the object in question. The robot may therefore spend a lot of time exploring irrelevant behaviors. This problem is frequently addressed by providing robot demonstrations or motion primitives that, while not perfect, manage to generate correct grasps with sufficient probability that such motions are observed during the exploration phase [1]. This project aims at solving this problem through appropriate exploration methods that will complement existing deep learning methods.

Scientific objective: This thesis project aims at defining an autonomous approach to learning in robotics in the case of rare rewards and interactions. The developments will be tested on the problem of object capture, with a particular focus on applications in the agricultural domain. They will be based on the work of the supervisors on novelty search methods and quality-diversity algorithms [2, 3, 4] as well as on preliminary results on learning object grasping behavior with these methods. Within the framework defined in the European project DREAM, the approach developed will be iterative [5, 6]. It will first aim at generating input behaviors by a mixed approach between exploration in simulation and testing in reality. This step will be based on an open loop policy learning. Each seizure of an object at a particular position will therefore require a dedicated learning. The repetition of such learning will allow to generate a sufficient base of examples to train a deep learning system that will associate an adapted behavior to the visual perception of the robot. The main objective of the thesis will be to generate these bases of examples thanks to adapted exploration methods.

**The Institute of Intelligent Systems and Robotics (Isir) is a Joint Research Unit (UMR7222) under the supervision of Sorbonne University, The French National Centre for Scientific Research (CNRS) and Inserm (ERL-U1150). This multidisciplinary research laboratory brings together researchers and teacher-researchers from different disciplines of Engineering and Information Sciences as well as Life Sciences.*

This objective can be broken down into 3 sub-objectives:

1. **Definition of behavioral spaces adapted to the input.**

Exploration methods based on novelty or diversity quality search algorithms aim to cover, as exhaustively as possible, a behavioral space that it is relevant to explore. The diversity of the solutions thus found allows to identify the possible behaviors and to have many alternatives to reach the same goal. Since this learning is done in simulation, the availability of different alternatives increases the chances of finding transferable and efficient solutions on the real system [4]. Moreover, this multiplies the number of examples and thus the size of the learning base. Several approaches will be compared, from expert design based on mechanical models of the input to learning these descriptors [7].

2. **Definition of appropriate policies for input gestures.**

The mentioned exploration methods require defining parameterized policies. The choice of policy will determine the space of possibilities and ways to grasp objects. It is therefore necessary to choose a versatile representation allowing to make the most of the capabilities of the different robots used, whether they are classical (Franka Emika type robotic arms) or not (flexible elephant trunk type robot under development at ISIR). The learning method must be adapted to these different configurations in order to get the best out of them for the considered input task. This step will also rely on a strong expertise in mechanics and automation to define sufficiently generic parameterized motion primitives.

3. **Learning of input models. The exploration methods mentioned above require numerous policy tests.**

It is not possible to perform all of them on the real robot which would be damaged too quickly. It is therefore necessary to carry out most of them in simulation, which raises the question of a sufficiently precise model so that a significant part of the learned behaviors can be transferred to the real system without difficulty. However, grasping behaviors involve contacts that are often poorly mastered in the models and robotic simulators used for learning (mujoco or pybullet, for example). From an iterative perspective, this part of the thesis will use the acquired data to correct an existing model or learn a complete model of the underlying physics.

As for the previous sub-objectives, this subject will be approached in a multi-disciplinary perspective between computer science for learning and mechanics for modeling these dynamic systems in multiple contacts with a poorly controlled environment [12][13]. We will try to answer a fundamental question of current interest: how can we bring data-based modeling methods closer to physical modeling methods, in order to combine the advantages of the former with the explicability of the latter?

Justification of the approach: Reinforcement learning searches for a policy, i.e. a function associating a state with an action maximizing a reward. This allows to find the appropriate behavior to achieve a goal that is described only by the occurrence of rewards that the system seeks to maximize. The resounding success of deep reinforcement learning [8,9] has however few equivalents in robotics. This domain accumulates difficulties for learning [10], with very large state and action spaces, transitions between continuous and discrete, noisy perceptions and actions, rare rewards, ... Several approaches are possible to deal with these challenges [11]: (1) "step-by-step" approaches, in which a model is used to test several possible actions that a criterion will discriminate to choose the best one and (2) episode-based approaches in which the policy is described by a parameterized function that is evaluated over the course of a complete episode, i.e., a sequence of perception-actions that may lead to a reward.

Step-by-step methods are more efficient and generalize better, but have difficulty handling the case of rare rewards. Episodic approaches can cope with this problem, but at the cost of more exploration and poor generalization. This thesis project aims at combining the best of these two types of approaches by handling rare rewards with an episode-based approach to generate a training base allowing to start learning a "step by step" system.

Required profile: Computer science student with a "machine learning" profile.

Required skills: Programming in Python must be mastered. Skills in robotics are desirable.

- **Application deadline: April 23, 2021**
- **Thesis director: Stéphane Doncieux**
- **Possible co-directors**: Faiz Ben Amar, Alexandre Coninx
- **Location**: Isir (Institut des Systèmes Intelligents et de Robotique), 4 Place Jussieu 75005, Paris
- **Contact**: **Stéphane Doncieux** ; doncieux@isir.upmc.fr ; Send your application by mail, with [Thesis: Learning to capture objects] in the subject line, a CV and a cover letter.

References (supervisors' publications are marked with a '*') :

[1] Ibarz, J., Tan, J., Finn, C., Kalakrishnan, M., Pastor, P., & Levine, S. (2021). How to train your robot with deep reinforcement learning: lessons we have learned. The International Journal of Robotics Research, 0278364920987859.
[2]* Doncieux, S., Laflaquière, A., & Coninx, A. (2019, July). Novelty search: a theoretical perspective. In Proceedings of the Genetic and Evolutionary Computation Conference (pp. 99-106).
[3]* Paolo, G., Coninx, A., Doncieux, S., & Laflaquière, A. (2021). Sparse Reward Exploration via Novelty Search and Emitters. arXiv preprint arXiv:2102.03140.
[4]* Kim, S., Coninx, A., & Doncieux, S. (2021). From exploration to control: learning object manipulation skills through novelty search and local adaptation. Robotics and Autonomous Systems, 136, 103710.
[5]* Doncieux, S., Filliat, D., Díaz-Rodríguez, N., Hospedales, T., Duro, R., Coninx, A., ... & Sigaud, O. (2018). Open-ended learning: a conceptual framework based on representational redescription. Frontiers in neurorobotics, 12, 59.

[6]* Doncieux, S., Bredeche, N., Goff, L. L., Girard, B., Coninx, A., Sigaud, O., ... & Duro, R. (2020). Dream architecture: a developmental approach to open-ended learning in robotics. arXiv preprint arXiv:2005.06223.
[7]* Paolo, G., Laflaquiere, A., Coninx, A., & Doncieux, S. (2020, May). Unsupervised learning and exploration of reachable outcome space. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 2379-2385). IEEE.
[8] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Hassabis, D. (2017). Mastering the game of go without human knowledge. nature, 550(7676), 354-359.

[9] Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, Z. D., & Blundell, C. (2020, November). Agent57: Outperforming the atari human benchmark. In International Conference on Machine Learning (pp. 507-517). PMLR.
[10] Kober, J., Bagnell, J. A., & Peters, J. (2013). Reinforcement learning in robotics: A survey. The International Journal of Robotics Research, 32(11), 1238-1274.
[11] Sigaud, O., & Stulp, F. (2019). Policy search in continuous action domains: an overview. Neural Networks, 113, 28-40.
[12]* Bouton, A. and Grand, A. and Ben Amar, F. (2017). Obstacle Negotiation Learning for a Compliant Wheel-on-Leg Robot. Int. Conf. on Robotics and Automation, IEEE/ICRA. 2017.
[13]* Bouton, A. and Grand, Ch. and Ben Amar, F. (2020). Design and Control of a Compliant Wheel-on-Leg Rover which Conforms to Uneven Terrain. Transaction on Mechatronics. Volume: 25, Issue: 5, Oct. 2020...