

Is a unified theory possible for robotics in unstructured environments?

Pietro Falco¹ and Dongheui Lee¹

Abstract—The generation of motion for robots and mobile manipulators in unstructured, dynamic environments has been a key research topic in the last years. Nevertheless, due to the high complexity of the problem, a ultimate solution has not been defined and the issue is far to be solved. In order to tackle uncertainties and unexpected events, several methods and diverse scientific communities have contributed to the development of artificial cognitive systems that aim at working in unknown, dynamic environments. Such methods include motion planning, machine learning, and perception-based control. The paper wants to briefly recall those heterogenous techniques and to propose a concept that embeds learning methods into a kinematic control framework.

I. TAMING THE UNCERTAINTIES: A BRIEF SURVEY

This work provides a general idea of some important methodologies adopted to tackle uncertainties in unstructured environments and proposes a direction for embedding learning capabilities into a wider framework based on null-space-based multiple behaviors.

1) *Reactive planning*: When the goal of the robotic system is a known trajectory or a known pose in the task space but the environment is affected by unexpected events, local reactive planners can represent a feasible solution. Such methods are suitable when the preplanned trajectories may need local corrections around the nominal values in order to handle external, unexpected events like obstacles in the workspace or inaccurate position of scene elements. The first category of local planners is based on artificial potentials in the task and configuration space [1]. This method associates an attractive potential to the desired state and a repulsive potential to obstacles. Similar approaches applied to entire trajectories in the configuration space and in the task space are [2] and [3] respectively. More recent approaches are specifically designed in order to exploit in a straightforward way the redundancy of modern robots like mobile manipulators. A requirement for modern low-level planners is the capacity of combining several tasks potentially with different priorities. In [4], methods to combine different behaviors in the task space are presented. Another recent approach is presented in [5], which combines tasks with different priorities in the configuration space and adopts a smoothing techniques to avoid chattering during task transition. Model-based, local planner methods are adopted for self-collision avoidance in highly redundant manipulators [6]. The concept of flexible planning has been introduced also in force trajectories [7][8]. Local, reactive planners are subject to local

minima and they need an initial-guess trajectory to work effectively, since the correction will be performed around the nominal trajectory. As a consequence, to work in real environments, those methods need to be somehow combined with classical global planners or with machine learning (ML) techniques.

2) *Imitation Learning*: Imitation learning techniques, called also Programming by Demonstration (PbD) techniques, allow robots to define a mapping between the state of the world and the action to perform by observing humans. This association is called policy. The policy can be learned at different levels of abstraction: it can consist in low-level trajectories for motion control, basic high-level actions (called primitives), and complex high level actions. An excellent survey on imitation learning can be found in [9]. Even if the PbD community has accomplished significant results, there are several open issues when dealing with tasks in unstructured environments. It is difficult, in fact, to adapt the learned policy to unexpected events that can be counteracted only with high-rate perception and reaction capabilities. Also, when possible unexpected changes in the environment bring the robot in states completely uncovered by the teacher's examples, new demonstrations are required to complete the task. As a consequence, the possibility can be deeply investigated to combine learning process with classical global and local planning approaches [10].

3) *Reinforcement Learning and Adaptive Optimal Control*: In the Reinforcement Learning (RL) framework [11], the association between the robot state, the environment state and the action to perform is learned through the experience collected by the robotic system, called also agent, acting in the world. The robot explores possible strategies and it receives a feedback on the outcome of each action. The action is evaluated according to a reward function. Adopting this method, the robot can keep learning during its whole life cycle. Reinforcement Learning and Optimal Control are strictly interconnected. In [12] it is shown how RL can be seen as an adaptive optimal control problem. Nevertheless, in unstructured environments several problems occur that limit the effectiveness of RL and optimal control. First of all, the definition of the reward function, defined as reward shaping, requires a significant manual contribution and expertise in the application domain [13]. As remarked in [11], the problem of extracting reward functions from the data has not been solved yet. Moreover, the problem remains of integrating the RL methods with the perception and reaction to sudden, unexpected events.

¹The authors are with the Department of Electrical and Computer Engineering, Technische Universität München (TUM). E-mail: pietro.falco@tum.de (P. Falco), dhlee@tum.de (D. Lee)

II. IS A UNIFYING THEORY FEASIBLE?

The question arises [14] if it is possible to find a unifying theory in robotics motion planning and generation. It has the potential to combine the flexibility of machine learning and the fast reactivity of kinematic control methods. A first step can be to develop a framework that allows a robot to refine on-line the policy and, at the same time, to quickly react to unexpected events through predictable, local corrections of motion/force trajectories. For example, when a robot is learning the execution of a task in an industrial or a general anthropic environment, the reaction to unexpected obstacles, even in presence of stochastic ML methods, has to be deterministic and independent from the particular task and reward function. To discuss in more detail a possible direction to follow, let us consider the following discrete-time system that can model a robot provided with joint-level controllers:

$$x_{k+1} = f(x_k, u_k)$$

where $x \in X$ is the state of the system (e.g. the configuration of a robot), u is vector of input to the system (e.g. desired joint velocities), and f is the state-update mapping. We want to teach the system a given task. Adopting one of the methods described in Section 1.2, we can define a policy $\pi : (X, W) \rightarrow U$. For each time frame k , the input u_k to the system will depend on the current state and on the value assumed by the stochastic variable $w \in W$. Without losing generality, the policy can be expressed either in the configuration space and in the task space depending on the particular application. The policy will be refined by reinforcement learning methods described in Section 1.3. As described in [11], a stochastic policy allows improving dramatically the learning phase. In presence of unexpected events, however, the necessity arises to include and integrate deterministic methods to obtain a quick, predictable reaction. At the light of this objective, let us introduce the pseudo-energy index $E(x)$ as a function of the state. Generalizing the approach followed in [5], the pseudo-energy quantifies how close is an unexpected event to prevent the task execution. For example, for the navigation of a mobile manipulator, E increases when obstacles approach to the robot and a collision is about to happen. Also, a second source of pseudo-energy can be the slipping of a grasped object. The value of pseudo-energy is strongly based on high-rate perception. The algorithm handles unexpected events by locally correcting the trajectories computed by the current policy in order to minimize the index $E(x_k)$. Following this idea, the control input is computed according to three behaviors:

$$u_k = \begin{cases} \pi_k(x_k), & \text{if } E = 0 \\ -\alpha_E J_E(x_k) + N_J \pi_k(x_k), & \text{if } E > E_{th} \\ c_0(-\alpha_E J_E(x_k) + N_J \pi_k(x_k)) + c_1 \pi_k(x_k), & \text{otherwise} \end{cases}$$

where N_J is a projector into the null space of the pseudo-energy gradient J_E and E_{th} is an energy threshold which can be chosen empirically, $c_0 \in [0, 1]$ and $c_1 = 1 - c_0$. In absence of unexpected events, the system simply executes and refines

the policy. When an unexpected event occurs and the risk of task failure is high -i.e. the pseudo-energy is greater than the threshold E_{th} - the priority becomes to minimize E trying to execute the planned policy at the same time through a null-projection method. For example, if during the policy execution an obstacle enters the workspace, the robot avoids the obstacle trying to execute the learnt task at the same time. When unexpected events occur but the energy is smaller than E_{th} , a convex combination of the two described behaviors is computed. To exploit the knowledge of the pseudo-energy gained at the execution time, the reward function r_k that classically is a function of the state x_k and the action u_k , is defined as a function also of $\hat{E}(x)$, i.e. the statistical distribution of the pseudo-energy in the workspace. This way, the RL methods of policy improving will be inclined to prefer regions of the workspace where the probability to have high values of pseudo-energy is smaller. In conclusion, a possible concept is presented to embed adaptive optimal control/machine learning techniques into a kinematic control approach with multiple behaviors. The leaning process, this way, becomes only one of the potential multiple behaviors that the robot system carries out.

REFERENCES

- [1] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," *The international journal of robotics research*, vol. 5, no. 1, pp. 90–98, 1986.
- [2] S. Quinlan and O. Khatib, "Elastic bands: Connecting path planning and control," in *Robotics and Automation, 1993. Proceedings., 1993 IEEE International Conference on*. IEEE, 1993, pp. 802–807.
- [3] O. Brock and O. Khatib, "Elastic strips: A framework for motion generation in human environments," *The International Journal of Robotics Research*, vol. 21, no. 12, pp. 1031–1052, 2002.
- [4] J. Lee, N. Mansard, and J. Park, "Intermediate desired value approach for task transition of robots in kinematic control," *Robotics, IEEE Transactions on*, vol. 28, no. 6, pp. 1260–1277, 2012.
- [5] P. Falco and C. Natale, "Low-level flexible planning for mobile manipulators: a distributed perception approach," *Advanced Robotics*, vol. 28, no. 21, pp. 1431–1444, 2014.
- [6] A. De Santis, A. Albu-Schaffer, C. Ott, B. Siciliano, and G. Hirzinger, "The skeleton algorithm for self-collision avoidance of a humanoid manipulator," in *Advanced intelligent mechatronics, 2007 IEEE/ASME international conference on*. IEEE, 2007, pp. 1–6.
- [7] G. De Maria, P. Falco, C. Natale, and S. Pirozzi, "Integrated force/tactile sensing: The enabling technology for slipping detection and avoidance," in *International Conference on Robotics and Automation (ICRA) 2015, Seattle*, 2015.
- [8] X. Song, H. Liu, K. Althoefer, T. Nanayakkara, and L. D. Seneviratne, "Efficient break-away friction ratio and slip prediction based on haptic surface exploration," *Robotics, IEEE Transactions on*, vol. 30, no. 1, pp. 203–219, 2014.
- [9] B. D. Argall, S. Chernova, M. Veloso, and B. Browning, "A survey of robot learning from demonstration," *Robotics and autonomous systems*, vol. 57, no. 5, pp. 469–483, 2009.
- [10] M. Saveriano and D. Lee, "Point cloud based dynamical system modulation for reactive avoidance of convex and concave obstacles," in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, 2013, pp. 5380–5387.
- [11] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The Int. Jour. of Rob. Res.*, 2013.
- [12] R. S. Sutton, A. G. Barto, and R. J. Williams, "Reinforcement learning is direct adaptive optimal control," *Control Systems, IEEE*, vol. 12, no. 2, pp. 19–22, 1992.
- [13] A. D. Laud, "Theory and application of reward shaping in reinforcement learning," Ph.D. dissertation, University of Illinois, 2004.
- [14] *Researcher: Pietro Falco. Host Professor: Dongheui Lee. Learning Control tight interaction: a novel approach to execution fo mobile manipulation tasks(LEACON). Marie Curie Action, Horizon 2020.*