Chapter 9

Conclusion

Autonomous robotic manipulation in unstructured environments is still one of the grand challenges in robotics. In order to manipulate an object, the robot has to perform several steps, each of them a research area itself. First, the robot has to look for the object in the environment and determine its position. Second, it has to plan how to manipulate it considering the task constraints and the environment. Finally, it has to execute the planned action adapting to prediction mismatches and possible external interferences. In this thesis, we have made use of state of the art algorithms for the two fist steps and developed sensor-based approaches for the last step.

Nowadays, there are a lot of manipulator robots that have proven good manipulation skills. However, the solutions available usually deal with small subsets of the problem. Robots are still far from being close to the scene understanding and dexterity of humans in such scenarios. In this thesis we have taken inspiration from human grasping experiments to implement a system capable to perform manipulation tasks in unstructured environments. In the implemented reactive contact based manipulation system: sensory feedback, adaptive control, contact detection, contact prediction, object detection and object recognition are key. Although there are also assumptions and constraints on the algorithms and approaches presented, we believe that they will be slowly but steady removed in the future.

Regarding the limitations and possible extensions of the work presented, an important feature that a robot shall have is the ability to adapt to the environment and learn from its interaction with the real world. However, in the system detailed in this thesis, learning is not present and should be incorporated in the future. Learning can be incorporated at the different levels of abstraction, the guidelines to add learning capabilities to the robot are discussed in Chapters 3 and 5 and summarized in Section 9.2.

Even with the organization and structure that the presented architecture provides, it is difficult to prepare new experiments and execute new tasks. Although task definitions are easily created, the connections of the services, primitives, and the configuration files, force the user to have a deep knowledge of the existing modules. Learning approaches could be used to mitigate this problem enabling to program tasks by demonstration.

In this thesis we have implemented and validated a contact driven robotic manipulation system. First, in Chapter 2 we have taken inspiration from neuroscience studies about human sensorimotor control of manipulation and identified the key components of a contact driven manipulation system. Second, in Chapter 3 we have presented, implemented and validated the manipulation primitives paradigm, a vocabulary of simple sensor-based manipulation actions that are combined to perform complex tasks. Third, we have developed and validated the mechanisms for contact event detection and prediction in Chapters 4 and 5 respectively. Fourth, in Chapter 6 we have presented the software architecture created to integrate all the pieces of the system and some useful abilities such as grasp planning. Fifth, in Chapter 7 an abstraction mechanism that allows the same tasks to be used by different robotic platforms has been presented. Finally, to endow the contact-based manipulation system with visual perception of objects we have presented in Chapter 8 a hierarchical object recognition system based on primate brain mechanisms. Details about the robotic platforms used to develop the work presented in this thesis are provided in Appendix A.

9.1 Contributions

The main contributions of the work presented through this thesis are listed below:

- The main contribution is the development of the manipulation primitive paradigm. A framework to implement and specify atomic reactive actions that can be used as building blocks to define more complex tasks.
- The implementation of a reactive strategy for grasping objects. The value of adaptive sensor-based control strategies is validated through the implementation and testing of the robust grasp controller.
- A complete pipeline to unscrew bottle caps is presented as a use-case of the manipulation primitives framework, a reactive unscrew primitive is also implemented and validated.
- A novel method to detect contacts using the robot model, current arm motion information and RGBD images is provided.
- A sensor fusion framework focused on contact detection and localization is another of the contributions of this thesis. The implementation of several sensor, context and prediction cues into the framework is also provided.
- Study and implementation of robot dynamic simulation to predict the robot-object interaction and the contacts that arise from it.
- An action abstraction architecture using the manipulation primitives paradigm in order to transfer plans between different platforms with different embodiments.

- The control architecture that orchestrates all the system components in a four-layered fashion.

9.2 Open questions and future work

The development of this thesis, has produced several related publications and tried to tackle some of the frequent problems in the robotic manipulation. However, as written in the following paragraphs, each of the aspects presented in this thesis can be improved. Beyond the improvements, the work presented opens several opportunities for future work and propose several open questions that would enhance and expand the research conducted throughout this thesis.

Human grasping experiments

In the neuroscience studies reviewed in Chapter 2, we have highlighted that humans perform corrective movements when there is a mismatch between predicted and perceived sensory input. However, there are no experiments available in the literature that study in detail how those corrections are performed. In this thesis we have performed preliminary experiments to observe how humans perform corrections, nevertheless to better understand and determine how humans adapt to different unexpected situations, more experiments are necessary.

Manipulation primitives

Although in Chapter 3 we have proposed and implemented a set of manipulation primitives, there are still more primitives that should be identified and implemented in order to increase the range of tasks that can be described. Some of the missing primitives are already identified (push, pull) but there might be others related to more specific environments (e.g. cooking) that are still unknown.

An interesting open question that should be investigated is the suitability of learning techniques to replace the implementation of manipulation primitives. Action-phase controllers could be learned instead of programmed. However, as suggested by [Johansson and Flanagan, 2010], corrective movements are learned together with each action-phase controller. Therefore, using learning to acquire new skills would require to learn corrective movements as well. Corrections are triggered by prediction mismatches, hence a prediction mechanism should be introduced into the learning scheme. A possible solution to learn this kind of controllers was proposed by [Pastor et al., 2011], where a grasping manipulation primitive is learned together with sensorimotor memories and corrective movements are performed when prediction errors arise. Unfortunately it is not clear how the learned strategy would generalize for other grasping tasks with different environment, objects and hand configuration.

Perception

During manipulation tasks, humans detect contact events using multi-modal information from different sensory cues [Johansson and Flanagan, 2010]. In Chapter 4 we have shown a sensor fusion framework that gathers contact information from different sources and provides the estimates of the detected contact events.

To the best knowledge of the author, in the literature there are no experiments regarding how the sensor fusion is performed by humans. Whether there is precedence of a sensory cue (e.g. tactile) over other cues (e.g. vision or predictions) is still unknown. Intuitively, our implementation of the sensor fusion mechanism uses a probabilistic approach and the priorities of the sensory cues depend on the confidence of each contact hypotheses generator. Experiments focused on this aspect of human manipulation would be very valuable to improve the sensor fusion methods and the contact detection.

Prediction

The current set-up uses the simulation to foretell where and when contacts will arise. Thus, if the simulator is able to predict when and where contacts will happen, it can be effectively used in the proposed human-inspired contact-based manipulation system. However, the real-time accurate physical simulation required is still not ready. The continuous development of physics engines makes it difficult to select the best engine for our prediction engine. In this thesis we have selected ODE but in the future there might be other engines with better performance. Hence, the implementation of a physics abstraction layer would be very valuable for the prediction engine. However, to obtain accurate simulations, a lot of unknown parameters regarding the object material and inertial properties are required. Furthermore, accurate simulation is computationally too expensive.

The prediction problem can also be approached from different points of view, instead of using a dynamics simulator, the consequences of object interactions can be learned by the robot [Belter et al., 2014]. Another approach is to obtain sensorimotor memories from a successful task execution and use them to monitor and drive the task execution [Pastor et al., 2011]. Hence, an approach that combines sensorimotor memories, learning and dynamic simulation could be the hybrid solution to the prediction problem. First, physics simulation parameters can be tuned by the execution of exploratory interactions with the object. Second, the simulation can be used to provide training data. Finally, the learned sensorimotor memory can be used to drive the execution of tasks and trigger corrections when mismatches are detected.

High level task planning

The task definitions used for the experiments presented in this thesis were manually created. We think that the manipulation primitives vocabulary, together with the perceptual primitives can be used as the base symbols for a higher level task planner that can convert semantically meaningful orders such as “clear the table” or “mop the floor” into executable task definitions. As recently published by [Yang et al., 2015] such task representations can be automatically obtained from unlabelled video sequences. However, from a single example, it is difficult to obtain the parameters and constraints to tune the task for a specific scenario. Once the task descriptions can be automatically learned from examples, a method to adapt the learned task description to different scenarios will be required. With enough examples of the same task, imitation learning provides methods that could be used for that purpose.

9.3 Publications

Parts of this thesis have previously been published in the following journal and conference papers:

1. Felip, J. and Morales, A. (2009). Robust sensor-based grasp primitive for a three-finger robot hand. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 1811–1816
2. Laaksonen, J., Felip, J., Morales, A., and Kyrki, V. (2010). Embodiment independent manipulation through action abstraction. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 2113–2118
3. Chinellato, E., Felip, J., Grzyb, B. J., Morales, A., and del Pobil, A. P. (2011). Hierarchical object recognition inspired by primate brain mechanisms. In *Computational Intelligence for Visual Intelligence (CIVI), 2011 IEEE Workshop on*, pages 1–8
4. Bohg, J., Johnson-Roberson, M., Leon, B., Felip, J., Gratal, X., Bergstrom, N., Kragic, D., and Morales, A. (2011). Mind the gap - robotic grasping under incomplete observation. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 686–693
5. Felip, J., Bernabe, J., and Morales, A. (2012). Contact-based blind grasping of unknown objects. In *12th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS*, pages 396–401
6. Felip, J., Laaksonen, J., Morales, A., and Kyrki, V. (2013). Manipulation primitives: A paradigm for abstraction and execution of grasping and manipulation tasks. *Robotics and Autonomous Systems*, 61(3):283 – 296
7. Leon, B., Felip, J., Marti, H., and Morales, A. (2012). Simulation of robot dynamics for grasping and manipulation tasks. In *12th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS*, pages 291–296
8. del Pobil, A. P., Duran, A. J., Antonelli, M., Felip, J., Morales, A., Prats, M., and Chinellato, E. (2013). Integration of visuomotor learning, cognitive grasping and

sensor-based physical interaction in the uji humanoid torso. *Designing Intelligent Robots: Reintegrating AI*, pages pp. 6–11

9. Bernabe, J., Felip., J., del Pobil, A. P., and Morales, A. (2013). Contact localization through robot and object motion from point clouds. In *13th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS*, Atlanta, GA, USA. IEEE
10. Morales, A., Prats, M., and Felip, J. (2013). Sensors and methods for the evaluation of grasping. In Carbone, G., editor, *Grasping in Robotics*, volume 10 of *Mechanisms and Machine Science*, pages 77–104. Springer London
11. Felip, J. and Morales, A. (2014). Dual arm sensor-based controller for the cap unscrewing task. In *14th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS*, Madrid, Spain. IEEE
12. Felip, J., Morales, A., and Asfour, T. (2014). Multi-sensor and prediction fusion for contact detection and localization. In *14th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS*, Madrid, Spain. IEEE
13. Felip, J., Durán, A. J., Antonelli, M., Morales, A., and del Pobil, A. P. (2015). Tombatossals: A humanoid torso for autonomous sensor-based task execution research. In *15th IEEE-RAS International Conference on Humanoid Robots, HUMANOIDS*, Seoul, South Korea. IEEE