Chapter 6
Physical Interaction: Adding Vision

6.1 Brief Introduction

Whereas the previous chapter has been devoted to the study of how physical interaction tasks can be executed using exclusively force feedback, we consider now the advantages of adding vision sensors. Computer vision can provide a powerful way of sensing the environment and can potentially reduce or avoid the need of structuring it. The purpose of this chapter is to study the additional value that vision feedback can provide to force-based physical interaction.

Vision and force sensors provide complementary and very different information. Vision sensors can provide a rich knowledge about the position and orientation of objects in the environment, whereas force sensors can provide local information to eventually correct the 3D trajectory of the robot end-effector when it performs constrained motion. Each one is useful for a different phase of the physical interaction task: vision allows accurate part alignment in partially unknown and/or dynamic environments, before and after contact is performed, whereas force enables compliant motion during contact.

Several researchers have studied vision-force control techniques with the purpose of getting the benefits of both sensors. This chapter firstly presents a review of existing vision-based control methods, concretely in section 6.2 and pays special attention to object pose estimation strategies, necessary for estimating the position and orientation of the manipulated objects in the environment. In this line, we present a new active pose estimation method, suitable for large objects which do not fit in the robot’s field of view. This method is specially interesting in the service robotics context, where the robot frequently needs to manipulate some parts of large and fixed objects like doors, cupboards, etc. Next, a review of vision-force control approaches is presented in section 6.3 and a new control method is proposed: the external hybrid vision-force control. We then discuss in section 6.4 how the previously outlined vision-force control techniques can be applied in the basis of the physical interaction framework, for both aspects of a physical interaction task.
Finally, in section 6.5, we show a practical implementation of the theoretical aspects in a real robot, performing vision-force physical interaction in the real world.

Results show how the use of vision feedback allows to retrieve the pose of the objects in the environment, thus avoiding the need to work in structured scenarios. In addition, this pose can be used to directly estimate the grasp link state, thus detecting any misalignment between the hand and the manipulated object, and making the robot adapt to the particular constrained trajectory through a suitable vision-force control strategy.

### 6.2 Visual Control of Manipulators

Motivated by the desire to reduce or avoid the need of structuring the environment, the use of visual observations to control robot motion has been extensively studied in the last years [23, 24, 109]. Typical applications of vision-based methods include the positioning of a robot with respect to an object, known as visual servoing, using either an external or end-effector-mounted camera.

Several visual servoing methods have been proposed in the literature [76]. On the one hand, they can be classified into dynamic look-and-move versus direct visual servoing approaches. Dynamic look-and-move systems consider that a low-level joint controller exists, accepting position or velocity inputs in the cartesian space. In contrast, direct visual servoing completely replaces the robot controller by a vision-based control law which directly generates control signals at the low-level. All the vision-based experiments considered in this monograph are classified into the first group. On the other hand, the visual servoing systems can also be classified, according to the nature of visual features they use, into position-based, image-based and hybrid approaches. The first one is based on the computation of a 3D cartesian pose (position and orientation), and normally requires a 3D model of the object and a calibrated camera [113]. In contrast, image-based visual servoing does not need a full model of the target, because the control loop is directly closed in the image space [50], providing increased robustness to modeling errors and noise perturbations with respect to position-based approaches [22]. More recently, several researchers have explored hybrid approaches which combine euclidean and image information in order to design globally stabilizing control [30, 110]. Finally, a third classification exists, depending on whether visual observations of the end-effector are introduced in the control law, or, in contrast, only the target object is observed. The first approach, known as endpoint closed-loop, allows to work with relative measurements, and, thus, it is robust to hand-eye calibration errors. On the contrary, the second approach, denoted as endpoint open-loop, requires an accurate knowledge of the camera position with respect to a robot-related frame, although the tracking of the end-effector is avoided.