Chapter 1

Introduction

1.1 Introduction and motivation

In this thesis we address the *robotic bin picking* problem (or pick and place problem) in the context of which a number of *piled* objects should be unloaded by a robotic manipulator from a bin or a platform on which they reside, and placed at a target position. There is a variety of areas that systems dealing with the problem can be applied: Among them, automobile and food industry, agriculture, airports, distribution networks, and postal facilities.

The majority of the bin picking tasks in the areas above are even nowadays executed *manually*: Humans are engaged for this purpose, and in many cases this is a monotonous and strenuous occupation, which often constitutes a threat for the health of the people involved. As demonstrated in [124] the use of human labor for such tasks has no economical or social justification, while the development of robot technologies for addressing them will require technicians, programmers, maintenance workers, and operators, jobs that offer better work conditions on one hand, require intellectual rather than physical skills on the other. Besides, depending on technology, a robotic system is expected to perform better than the human workers. While for human labor the efficiency decreases as the time passes, a robotic system can theoretically work for twenty four hours a day, seven days a week, with higher accuracy and robustness, and with lower costs than the humans.

The majority of existing robotic systems employed for dealing with the bin picking problem in the industry so far, are hardly equipped with sensors to perceive their environment. These *blind* manipulators usually just play back a pre-programmed sequence of motions. Hence, they require a highly structured environment to be able to operate. This however is not easy to achieve, since in many cases the position of the platforms on which the object reside change, the objects vary in shape and dimensions, or they are jumbled in an arbitrary way on the platform. The *flexibility* of such robotic systems can considerably increase when they are equipped with vision sensors. Incorporation of vision sensors enables the systems to perceive whatever changes take place and operate accordingly, by acquiring and analyzing images of the environment.

However, image analysis for unloading operations is admittedly a difficult task [139]. The reason for this is that the image analysis task in the context of a bin picking application involves two subtasks, each of which is since years known to be very hard to deal with: Given an image of the object configuration, the system should be firstly able to isolate the areas in the image corresponding to a unique object, that is, to *segment* the object. Secondly, the system should be able to *recognize* the type of the particular object. The combination of these two tasks is often referred to as object *recovery*, or *localization*.

Ideally flexible systems should be able to satisfactory perform the recovery task, regardless of the type, dimensions or texture of objects placed on the platform, as well as the lighting conditions at the installation sites. Despite the research that since many years has been conducted on the issue, and the considerable progress that has been performed, the conception and implementation of a completely generic system does not seem to be feasible in the nearest feature. For this reason, the majority of approaches focus on a specific application, in the context of which configurations containing particular categories of objects in particular environments are dealt with.

Target of this thesis is the development of a vision guided robotic system for automating unloading of jumbled *box-like* objects. Box-like objects comprise the greatest majority of objects that can be found in distribution centers and warehouses [139]. The driving force for the implementation of the particular system, is mainly the industrial requirement for the reduction of costs in the logistic processes: In year 2002, 8.2%, 28%, and 13% of the product manufacturng costs of the automobile industries, food industries and industries involved in the production of various items of high demand respectively, are spent for logistics. Product handling within the distribution centers, which mainly involves depalletizing, is admittedly a considerably high percentage of these costs [139].

In distribution centers and warehouses, the objects are usually piled on rectangular platforms, the *pallets*, as in fig. 1.1, and the problem of automatic object unloading is widely known as the *depalletizing* problem. Fig. 1.1 illustrates representative objects encountered in distribution centers with which our system is able to deal. More specifically, fig. 1.1 (a) and (b) show configurations of bags (sacks) full of material. Fig. 1.1 (c) and (d) present two configurations of rigid (cardboard) boxes. Fig. 1.1 (e), (f) illustrate configurations of jumbled box-like objects wrapped in transparent foil. Fig. 1.1 (g) show a configuration of box-like pillows. Finally, fig. 1.1 (h) depicts a configuration containing a combination of the objects mentioned above. Note, that the application range of a vision guided depalletizing system for box-like objects is quite vast, since apart from distribution centers and warehouses, it may be used at airports to deal with suitcases, or at postal service facilities for parcel unloading and sorting.

This chapter is subsequently organized as follows: In section 1.2, we present in detail the requirements we want our system to meet, we touch upon the fact that existing systems fail to realize these requirements, and give a rough description of our approach for dealing with the problem. Section 1.3 highlights the contributions of our approach to the state of the art. Finally, section 1.4, presents an overview of the remaining contents of this thesis.

(a) Bags (1)

(b) Bags (2)

(c) Cardboard boxes (1)

(d) Cardboard boxes (2)

(e) Box-like objects (1)

(f) Box-like objects (2)

(g) Pillows

(h) Mixed configuration

Figure 1.1: Object configurations

1.2 Objectives

Our objective has been to design and implement a novel, vision-guided robotic system for automatic depalletizing of piled box-like objects, which meets the following five requirements:

- **Flexibility:** The system should be able to deal with objects of *arbitrary dimensions*. Besides, the fact that the objects could be both jumbled on the pallet, or neatly placed on layers should be taken into consideration. In addition, the system should be able to address the fact that many objects, a representative category of which is the sacks (see fig. 1.1 (a),(b)) may *deform*, since they are not entirely full of material, so that they shape becomes considerably different than this of the rigid boxes (see fig. 1.1 (c),(d)) for example. Finally, our system should be able to operate without problems regardless of the texture of the objects, or the lighting conditions at the installation sites, since both may greatly vary.

- **Robustness:** The system should be able to execute its task in a robust way. In simple terms, when the system thinks that an object resides in a particular position in space, this should actually be so. In other words, the system should be designed in such a way so that the number of false positive responses is kept as low as possible.

- **Accuracy:** The system should localize the objects with high accuracy. The more accurate the localization process is, the more accurate both the object grasping and placement operations will be, so that the danger of grasping the objects from the wrong positions or destroying them is minimized.

- **Computational efficiency:** Our system should be computationally efficient. Since our aim is to replace human labor, our system should even be faster than the human workers when performing its task.

- **Low cost:** Keeping the cost of the system, that is, the cost of the hardware parts of the system in low level, is very important for attracting possible future investors. Note that except for the hardware components, ease of system's installation is essential for preserving the costs at a reasonable level.

As we will see in chapter 2, up to our knowledge, no existing system satisfies *all* the requirements previously exposed. In the context of our system a six degrees of freedom industrial robotic manipulator is employed for object unloading. A laser sensor is mounted on the hand of the robot for acquiring images of the top side of the pallet. Besides, a vacuum gripper is as well mounted on the hand of the robot for object grasping and placement. Our system operates as follows: Firstly, a range image of the object configuration is acquired. Secondly, the image is analyzed and a number of objects on the top of the pile are recovered. Thirdly, the recovered objects are grasped by the robotic hand and placed at positions defined by the user. This process continues iteratively, until no objects lie on the pallet.

This thesis primarily focuses on the object recovery problem, that is on the way in which given a range image containing a configuration of piled box-like objects, the recovery of the graspable objects in the image can be robustly, efficiently and accurately performed. This purpose is achieved by a *twofold*, model-based strategy, where geometric parametric entities are employed for object modeling. According to the first aspect of our strategy, the target objects are modeled using globally deformable *Superquadric* entities [11], [148]. Object recovery is posed as a mathematical optimization problem, within the context of which the *posterior* probability of the objects' parameters given the acquired image is maximized. This is performed by means of a two stage process: Initially, a *hypothesis* about the values of these parameters is generated, so that a *rough* recovery is performed. Subsequently, the hypothesis is *refined* and a considerably more accurate recovery result is obtained. Noteworthy is, that within our approach for object recovery we *fuse* both range and object *boundary* information. The latter is acquired by an *edge image*, obtained by the input range image via a *boundary detection* operation. The second aspect for object recovery is based on the *Hough* transform [75], [96]. Modeling entity here is the boundary of the exposed surface of the piled objects. The recovery problem in this case is *decomposed* into two subproblems, the recovery of the pose of the objects followed by the recovery of their dimensions, and the Hough transform on the edge image is used for addressing each.

Both aspects, are standards in the field of computer vision for the recovery of *multiple* objects in images, but the way in which each one is used in the context of our application is novel. The reason that we worked with both aspects for object recovery, is that there exists a a trade-off between computational efficiency and flexibility: More specifically, the former aspect is more flexible than the latter, since it is able to effectively deal with object deformations, and is at the same time more accurate. It is however less computationally efficient than the latter. The latter has proved to be computationally efficient, but is able to accurately deal with *rigid boxes* only. In conclusion, given a configuration of objects their degree of rigidity suggests which of the two aspects should be employed for recovery: When rigid objects are only involved (as is the case in figures 1.1 (c),(d)) the latter strategy should be selected, and in any other case the former.

1.3 Main contributions

The *main* contributions of this work can be summarized in the following items:

- Our system contributes to the state of the art in vision guided applications for robotic grasping of box-like objects, since it meets all requirements mentioned in section 1.2, which, up to our knowledge, is not the case for any existing system.
- Our dual object recovery approach contributes to state of the art in the recovery of multiple geometric parametric models in range images in two ways:
    1. It extends the state-of-the-art in superquadric recovery, the *recover-and-select* framework [98], [99], [77], so that *boundary information* is taken into consideration. As a consequence, the recover-and-select framework is outperformed in terms of all robustness, accuracy and computational efficiency. Note that despite the fact that we mainly deal with box-like superquadrics in this thesis, the particular approach can be used for the recovery of superquadrics of arbitrary shape. In this respect, our robotic system is given the potential to unload a much wider range of shapes, that is, every kind of superellipsoids.
    2. It exemplifies how the Hough transform can be effectively used for three dimensional object recovery, by decomposing it into easier subproblems.

1.4 Thesis overview

This work is subsequently organized as follows:

- Chapter 2 attempts a categorization of existing systems dealing with recovery of multiple objects from images for robotic grasping purposes. The advantages and drawbacks of the systems are pointed out, and an introduction to the operation of our system is presented, along with a discussion of its advantages with respect to existing systems.
- Chapter 3 introduces superqadrics as modeling elements, and gives to the reader all necessary information about their properties required for comprehending the contens of the subsequent chapters, especially the contents of chapter 5.
- Chapter 4 focuses on the issue of edge detection in range images, and more specifically in the way in which model based information can be used for this purpose.
- Chapter 5 is the core of the thesis. It presents the way in which box-like superquadric models can be used for object recovery in range images, by fusing depth and boundary based information. Besides, a comparison of our approach with the recover-and-select framework for superquadric recovery is conducted. Finally, we touch upon the fact that our approach can be used for the recovery of superquadrics of arbitrary shapes and present initial experimental results.
- Chapter 6, demonstrates the results acquired using the approach of chapter 5 in a multitude of object configurations. Besides, a quantitative analysis performed by comparing the output of our approach with ground truth images prove that it is a robust, accurate and computationally efficient method for multiple object recovery. Additional experimental results are presented in appendix A.
- Chapter 7, focuses on an alternative means for the recovery of piled rigid boxes in a very efficient way, namely the decomposition of the Hough transform of the edge image of the configuration.
- Finally, chapter 8 concludes this work, points out that our system actually satisfies the requirements set in section 1.2, discusses ways in which the problems of our system can be addressed, and presents our future work.