Learning to predict grasp reliability with a multifinger robot hand by using visual features

Ph. D. Thesis

Antonio Morales i Escrig

Supervisors:
Angel Pasqual del Pobil
Pedro José Sanz

Castelló, 2003
Learning to predict grasp reliability with a multifinger robot hand by using visual features

Ph. D. Thesis
by
Antonio Morales i Escrig

Supervisors: Angel P. del Pobil and Pedro J. Sanz

Department of Engineering and Computer Science
Universitat Jaume I

Abstract

Grasping and manipulation skills are critical for a wide variety of robotic applications. As these applications become more ambitious, the uncertainty and unstructureness of the environment increases. As a consequence, additional flexibility and robustness is required from the components, mechanical and logical, of the robotic systems.

Extensive research on robotic grasp synthesis and analysis during the last two decades has established a strong theoretical framework. However, most of this research has been based on perfect models or ideal operation conditions. These assumption often become unrealistic in real world applications. A main goal of this thesis is to provide algorithms more suitable to practical conditions.

We face the problem of environmental uncertainty in the task of grasp determination and selection by following two separate but complementary approaches. First, we make use of sensor-based information, in particular from vision, to acquire a proper description of the environment and, thus, reducing the uncertainty. Second, we develop a learning framework that improves the predictive performance of the grasp stability prediction procedures by using the experience accumulated in previous grasp trials.

More in detail, we focus on the grasp determination and selection on unknown planar objects. We use vision to provide a description of the objects. A couple of algorithms are developed in order to compute two and three-finger grasps that meet necessary stability conditions. These algorithms require little computational effort.

An experimental grasping system that includes the synthesis algorithms is developed. This system is implemented over the UMass humanoid torso, that has two three-fingered Barrett hands. An additional algorithm to adapt unconstrained two and three-finger grasps to the geometric constraints of the hand is also presented. Extensive experimentation is carried out on this system.

The learning framework is based on a grasp characterization scheme that is developed in this thesis. It is based on a set of high-level vision-based descriptors that allows the abstraction and generalization of grasp physical properties. This characterization scheme is used to store the experience. Moreover, a practical test for providing a qualitative measurement of the reliability of a grasp is presented.

Finally, several predictive algorithms, based in pattern classification techniques, are presented. They are able to predict the reliability of a grasp based on its similarity to past experiences. An active learning procedure that guides the selection of grasps with the goal of more effectively increasing the accumulated experience is also described. Exhaustive experimental data is collected in order to test and validate the developed algorithms.

Castelló, December 2003
Chapter 1

Introduction

Motivation, goals, contributions, and structure of the present thesis.

1.1 Motivation

Humans show an extraordinary capability for manipulating and using objects. In fact, this is one of the features that most distinguishes humans from other species of living creatures. In the presence of such an outstanding skill roboticists, engineers and all kind of scientists have devoted a great effort in understanding and replicating this dexterity. Though the model has not been equaled, and will not in many years, this research activity has produced a vast amount of theoretical and practical knowledge about the principles of manipulation. The fields of research in robotics manipulation range from the design of robotic hands to the physics of contacts among objects, to studies on neurological and psychological foundations of manipulation.

Nonetheless, manipulation is one of the most useful skills in any robot system and constitutes a key component for many robotic applications on all kind of areas such as industry, medical, service, and space robotics. In general terms, manipulation encloses three different types of activities: exploration, restraining of objects, and dexterous manipulation (Bicchi, 2000). Contact exploration belongs to the most broad field of haptics and it is a complete research field in its own (Klatzky and Lederman, 1990; Hollerbach, 2000).

Restraining of objects, also called fixturing, refers to the task of immobilizing objects with the fingers. Finally, the term dexterous manipulation refers to the cooperation of multiple manipulators, or fingers, to grasp and manipulate an object (Okamura et al., 2000). Most grippers are used for fixturing, and not for dexterous manipulation, since mechanical and control features are necessary to make this possible. As an instance, Shimoga (1996) argues that the classical two-finger parallel jaw gripper is completely unable of doing any dexterous manipulation.

The kind of manipulation activities that are described in this thesis can be included in the fixturing group. We manipulate objects by caging them with the fingers and, then, use the robot arm to move and orient them. In order to simplify the notation, the term grasping will be used along this thesis in the sense of fixturing.

An important trend in robotics during the last years has focused on more flexible applications working in complex environments. An example of this is service robotics, that is, robotic systems collaborating with humans in human environments. In general, these applications need to be flexible enough in order to deal with unstructuredness and uncertainty in
their work-space. In the case of grasping tasks this uncertainty comes from several sources: variable location and pose of the objects, a priori unknown objects, and the diversity of tasks to perform with the objects. To overcome these difficulties, a wide range of different approaches have been used. Among them, two particularly relevant are the use of sensorial information, and the application of learning techniques.

In order to deal with unknown objects a grasping system must be able to use its sensors to detect and extract the necessary information about the object, as they are not already available as a model. This information is then used to decide on the most convenient way to grasp the object, possibly taking into account some constraints imposed by the task to perform with it.

The selection of the contact locations, or grasp synthesis, is not a trivial problem. Fortunately, extensive research in robotic grasping and dexterous manipulation over the last twenty years has established a theoretical framework for grasp analysis, simulation and synthesis. Though, often, as Shimoga (1996) remarks, there exists a lack of practical implementations that fulfill the theoretical promises. Among the reasons for this situation, prohibitive computational complexity, lack of precise sensing capabilities and, most importantly, the difficulty in modeling real-life objects are often highlighted.

Moreover, particular characteristics of the mechanic components involved are often not taken into account. This is particularly critical in the case of the grippers, the devices that make contact with the objects. They are often abstracted as a number of fingers that can reach any location in the object surface, capable of exerting forces and wrenches of any magnitude. Real grippers seldomly offer that flexibility and behave differently one from another.

However, the main problem of classical model-based approaches is that they rely on a critical assumption, that is, the availability of a complete geometric model of the object to be manipulated, either two-dimensional or three-dimensional. In many robotics application this is simply not possible.

One way to avoid the lack of an accurate model of the object, is relying on sensors, mostly vision, to obtain an appropriate description of the object. Later this description is used to compute the contact locations. However, an inherent problem of visual information is the inaccuracy of the descriptions, which affects any posterior processing. This problem has been faced by the use of sensors that reactively correct the possible deviations. Coelho Jr. et al. (1998) and Platt et al. (2002) use tactile sensors and finger-gaiting to find stable fixtures of an unknown object.

But even with the use of sensors, very often, the uncertainty and the lack of a sufficient a priori knowledge about the environment make it difficult, or simply impede, the design of a robust and reliable system. The only option is the design of systems able to adapt their functioning to the environment conditions and learn the relevant features of it from sensorial information. Bajcsy (1993), Kamon et al. (1996), and more recently, Coelho Jr. et al. (2000) present reference works in the application of such techniques in the fields of robotic grasping and manipulation.

1.2 Goals

The main goal of this thesis is to develop a learning approach to the grasp synthesis problem. More precisely, our goal is to learn how to predict the reliability of a given grasp candidate.

Since it is not possible to learn from scratch we intend to develop a practical grasp synthesis system that comprehends the apriori theoretical and practical knowledge about the
problem, thus reducing the search space that the learning techniques will explore. The goal of 
this system to produce a set of feasible candidate grasps that meet some necessary theoretical 
stability conditions. In theoretical terms we limit the grasp synthesis to the problem of the 
selection of the contact points on unmodeled planar objects. Planar restriction is imposed 
in order to reduce the complexity of the problem, and in particular of image processing, that 
otherwise could become the heaviest problem of the work.

Moreover, since we focus on a practical approach, this core grasp synthesis system will 
be enhanced by use of visual information that would be responsible of providing the required 
description of the objects. Other complementary practical goals are the adaptation of the 
produced candidate grasps to the the peculiar features of the used gripper, and the develop-
ment of computationally cheap algorithms. The final accomplishment of this part will be 
constituted by the implementation of these methodologies on a real robot able to execute 
any of the grasp computed. This will not only validate the usefullness of the grasp synthesis 
system but will serve as an experimental testbed for the learning procedures themselves.

The development of the learning framework is the main goal of this thesis. But this 
needs that several subgoals are met. A way of representing knowledge or experience has to 
be developed. It has to allow generalization among experiences in different kinds of objects. 
Moreover, a way to interact with the environment that allows the learning system to measure 
the performance of its decisions in real world has to be developed, too.

Once this component is available it will be possible to develop the core of the learning 
system, that is, the set of algorithms that have to allow not only to use the past experience to 
make a guess about the reliability of a given candidate grasp, but also to guide the acquisition 
of experience in order to obtain a better predictive model.

1.3 Contributions

This thesis deals with the above mentioned goals. In particular, we have developed a number 
of algorithms and novel techniques to accomplish this pursue. These contributions are divided 
in two main groups that are also the two main parts of this thesis.

First, a complete practical sensor-based grasping system has been developed and imple-
mented. The most relevant contributions in this part are:

I. A couple of heuristic vision-based grasp synthesis algorithms for planar two and three-
finger grasps. The main characteristics of these algorithms are that they make use of 
vision as the only input; they meet theoretical stability conditions; finally, they require 
a small amount of computational effort (Morales et al., 2001, 2002a).

II. A novel approach that adapts planar two and three-finger grasps to the particular 
kinematics of the three-fingered Barrett Hand\textsuperscript{1} is introduced. Again, this algorithm 
takes care of the required computation time (Morales et al., 2002b).

III. Finally, a real grasping system that uses the previously developed algorithms is imple-
mented. It is able to locate and grasp a previously unknown object lying in front of the 
robot. For this system we make use of the facilities of the Laboratory for Perceptual 
Robotics of the University of Massachusetts, in particular, the UMass humanoid torso.

\textsuperscript{1}Barrett Technology Inc. \url{http://www.barrett.com/}
Second, we have developed a learning framework that allows the acquisition, storing and retrieving of experience of the execution of grasps. In more detail it can be decomposed in four parts:

IV. A grasp characterization scheme is developed. This scheme represents a grasp in such a way that its description is invariant to the pose and location of the object, indeed, it is independent from the object itself. Moreover this characterization takes into account the peculiarities of the hand, has a physical meaning, and considers robustness and stability. Finally, and most important, this description is computed exclusively from visual information (Chinellato et al., 2003).

V. A practical test to obtain a qualitative measure of the reliability of a grasp is implemented. The results of this test are used to provide an experimental metric of the quality of a grasp. Moreover this test is used to carry out exhaustive experimentation, and, thus, build an extensive experimental dataset on which to validate any further procedure.

VI. Several algorithms that predict the reliability of a grasp based on previous experience are developed, too. The experience is represented as the characterization of the grasps previously executed plus the reliability measures obtained with the practical test. These algorithms are validated with the samples from the exhaustive data set and their usefulness is demonstrated (Morales et al., 2003a,b).

VII. Finally, a procedure that actively leads the acquisition of experience is implemented. The goal of this procedure is to obtain the necessary amount of experience to produce a good prediction performance with the minimum number of necessary trials. A secondary goal is to concentrate on most reliable trials. It guides its functioning by using the already obtained experience in a truly active learning.

1.4 Outline of the thesis

The thesis is divided in two main parts. The first one, composed by chapter 3 and 4, is devoted to explain the grasp synthesis strategy, while the second, chapters 5 to 7, develops in detail the whole learning framework. More precisely, the thesis is structured as follows.

Initially chapter 2 briefly describes the methodology of the work and presents an extended summary of the procedures and main results presented all through the thesis.

Chapter 3 introduces the grasp synthesis algorithms developed to compute planar two and three-finger grasps using visual information as input.

Chapter 4 describes the experimental setup, the UMass humanoid Torso, and the grasping system implemented on it. The necessary processing used to adapt the results from the previous chapter are described in detail. Finally, results and observations about the performance of the grasping system are presented.

The grasp characterization scheme used to describe the grasps independently from their original objects and including physical and robustness considerations is described in chapter 5. A practical test to evaluate the reliability of a grasp is also provided.

Chapter 6 and 7 develop the main contributions of this thesis. The former describes the classification methods that are used to predict the reliability of a grasp from a set of past executed and tested grasps. The protocol followed to obtain a sample test database is also
provided. Finally, analytical results that demonstrate the appropriateness of the introduced
methods are shown and discussed.

Chapter 7 describes the active learning technique proposed and the simulation environ-
ment used to test and validate its performance.

Finally, chapter 8 shows some general conclusions and briefly describes some future lines.
Chapter 2

Methodology and overview

Extended summary of the procedures and main results presented all through the thesis.

This chapter presents an overview of the methodology followed for developing the work in this thesis. We also try to give a high-level description of the algorithms and results that are described through the whole thesis. At this point, few justifications are given, and as an advice, the reader is referred to the corresponding chapters where the full details, references, and results are explained.

2.1 Methodology

The final goal of this thesis is to develop a learning framework for predicting grasp reliability. But learning from scratch is not possible, and, in general, the problem and the search space must be narrowed before learning is possible. Moreover, there exists a fairly extensive knowledge about the physics, synthesis and analysis of manipulations, which is sensible to exploit, at least partially. The question is, then, for what learning is necessary.

What do we want to learn ?

In theory, within a perfectly modeled problem, it is possible to determine the best grasp for a given object. There exist several algorithmic methods for computing these grasps, and analyze their kinematic and dynamic properties, under a variety of contact models. However, in many real applications it is not possible to provide a complete model of the systems, or the assumptions can not be guaranteed. Even with the use of sensors the information provided by them is affected by inaccuracies or errors. In those cases theoretical-based approaches are insufficient, though able to provide an estimation of the solutions. Learning is useful, then, to adapt and apprehend the still unknown properties of the environment, thus completing the a priori insights.

More precisely, in our approach a grasp synthesis system provides a set of feasible grasps for a given object. These computations are based on a description of the objects obtained from images of them. It makes use of some stability conditions to ensure a minimum quality of the computed grasps. The main consequence of this preprocessing is that it reduces the infinite combinations of possible contact locations to a few dozens of feasible candidates.

However, only one grasp candidate can be executed. So, a procedure for the selection must exist. In this thesis we propose a learning-based experimental approach. The system
learns the relation between the features that characterizes every grasp and its reliability. It does so by using the recorded results from previous trials and using them to assess the reliability of every candidate. Finally, an exploration procedure that guides the selection with the purpose of acquiring the necessary experience with the minimum number of trials is also presented.

Once the subject of learning has been established, we propose a framework that is based on four main pillars.

- A grasp characterization scheme that provides a unique description of any grasp.

- An experimental test, by means of which the robot can determine the reliability of a given grasp.

- A set of techniques for estimating the reliability of a grasp from its similarity to other grasps.

- An active learning scheme to select the next grasp to execute with the purpose of improving prediction performance of the accumulated experience.

General assumptions

We make some assumptions in our approach in order to focus our problem and limit the reach of our research. These assumptions can be summarized in these points:

- **Planar objects** and **2D grasp synthesis**. We consider the objects to be planar, and, thus, our approach is based on 2D grasp planning techniques. In practical cases with real 3D objects we consider the projection of their silhouette over a plane (i.e.: the plane of the table that supports them).

- **Visually-guided grasp synthesis**. Our grasp synthesis approach is based in information about the objects obtained form a vision system.

- **Two and three-finger grippers**. We focus on the computation of planar grasps with two and three-finger hands.

Experimental setup and system architecture

The grasping system we are considering is composed of four modules, which also correspond to the steps of the processing steps. These are: vision processing, synthesis of grasp candidates, candidate selection, and candidate execution. This architecture is considered as a reference for the design of the different modules. In chapter 4 we give a more detailed description of the implementation of this model on a real system, the UMass humanoid torso. This system is composed of a pan-tilt head with a stereo-camera system, two 7 d.o.f. arms and two three-finger Barrett Hands. The image-processing module has to be adapted to deal with stereo images. It is used not only to obtain a description of the object but also to locate the object in the workspace.
2.2 Grasp synthesis

We define as *grasp synthesis* the procedures aimed to the computation of the location of the contact points on the surface of the objects where the forces are going to be exerted. These locations have to be such that the forces exerted must meet several stability conditions:

1. **Non-sliding condition.** Since we assume the model of contact with friction, and according to Coulomb friction model (Coulomb, 1781), the directions of forces exerted on the surface by the fingers can not differ from the normal direction for an angle larger than \( \beta = \arctan \hat{\mu} \), where \( \hat{\mu} \) is the friction coefficient. The value of coefficient depends of the materials of both the finger and the object surface. Otherwise, a finger will slip. The forces can be of any magnitude, but their directions must meet this condition.

2. **Curvature condition.** As stated by Montana (1991) the curvature of the surfaces in contact is also a relevant characteristic for stability. Planar surfaces provide a better support surface that very curved ones. According to this condition, contacts on planar surfaces are desirable, and locations in points of the surface with an excessive curvature are discarded.

3. **Force-closure condition.** Nguyen (1988) defines force-closure as the property of a set of contact points, a *grasp*, such that the net force composed by the forces exerted in these points can counteract any external force applied to the object.

All these are necessary conditions for a set of contact points to be considered a stable grasp. Our approach implements geometrical interpretations of all these conditions.

Image processing

Ours is a practical approach that obtains the object shape description from visual input instead of using precomputed models. The image processing system is responsible of providing an appropriate description of the object. In our case this description consists of the contour of the object.

We have developed a simple contour extraction algorithm. It assumes gray-level images where a single object appears. It also assumes that the object and the background are well contrasted. The histogram of this kind of images is bimodal. An automatic thresholding algorithm is used to find the separation between both modes. This threshold is used to binarize the image and then apply a trivial contour following algorithm.
Figure 2.2: On the left, the two-finger grasps computed for this object. On the center and on the right, two of the three-finger grasps computed

Grasp region extraction

A grasp region is defined as a sequence of points in the contour that meet the curvature condition. The computation of grasp regions discards all the points with an excessive curvature, and groups the remaining points in an extremely useful abstraction. This processing allows to reduce the complexity of the problem from a set of potentially hundreds of points in the contour to a couple of dozens of grasp regions, which are abstracted as straight segments.

Two-finger grasps

As stated by Nguyen (1988) force-closure with two friction contact points is achieved when the grasping line lies inside both friction cones. The procedure for computing two-finger grasps from grasp regions is based on this principle. It is composed of two steps.

1. Finding pairs of compatible regions. Two grasp regions must meet two conditions in order to be compatible

   (a) The angle between the normal vectors to each region is in the range $180.0^\circ \pm 2\hat{\beta}$.

   (b) The projection of each region, in the direction of its normal, intersects with the other region, i.e.: the regions are confronted.

2. Refining of the pairs of compatible regions in order to find a couple of points, one per region, that meet the force-closure conditions.

Three-finger grasps

In the case of three-finger grasps, both Park and Starr (1992) and Ponce and Faverjon (1995) demonstrate two necessary conditions that three-finger grasps must meet assure force-closure:

I) The intersection of the semiplanes defined by the friction cones of the forces at the three grasp points is not empty. (see fig. 3.7 on page 29).

II) The unit normal vectors to the surfaces defined by the grasp regions positively span the plane. That is, the three vectors are not contained in the same half space.

As in the two-finger case, the procedure to compute three-finger grasps is composed of two steps:
1. Finding triplets of compatible regions. These are characterized by two conditions.
   
   (a) The normal vectors of the three regions positively span the plane.
   (b) The intersection of the interior projections of each of the grasp regions is not empty.

2. The center of the intersection is, then, projected on the grasp region segments, and these three points are the contact points that conform the grasp.

The procedures described in the above two last sections have been implemented and tested with images of different kinds of objects. The results obtained can be summarized in two ideas. First, these algorithms are able to compute sets of force-closure two and three-finger grasps. And second, they do it very fast, making these algorithms suitable to be embedded in interactive grasping systems. These results are more detailed in section 3.6 on page 30.

### 2.3 A whole grasping system

The grasp synthesis algorithms do not constitute a whole grasping system, though they are the core of such. Indeed, a full sensor-based grasping system is implemented over the UMass Torso and several additions must be included to make it fully operational. These extensions refer basically to three aspects: the vision system, the constraining to the hand kinematics, and the execution of a grasp.

#### Stereo-vision system

The assumptions about the grasping system consider only one camera. Indeed, the object description can be obtained from a single image. As we mentioned before, the visual facilities of the UMass torso are composed by a pan-tilt head and a stereo-vision system. In this grasping system the cameras are fully calibrated and the images are used to locate the object in the work-space.

A secondary problem appears due to the fact that the view of the objects is oblique. To overcome this difficulty we restrain the shapes of the object to be planar with a constant height. Moreover, a predefined color scheme of the parts of the object is applied to make the top of the object, its sides and the background clearly distinguishable. In this case we can obtain the top face silhouette by a simple adaptation of the image processing explained above. The calibrated model of the camera is also used to estimate the size of the object.

#### Constraining to hand kinematics

The Barrett hand is a three-finger gripper with only four degrees of freedom. This limits the number of three-finger grasps that it can execute with respect to a “theoretical” three-finger hand. This means that the grasps computed by the grasp synthesis algorithms are not directly applicable to the hand. They must be adapted to the particular kinematic constraints of the hand, and often this is not possible.

In this thesis we develop a novel approach for constraining two and three-finger planar grasps to the mechanics of the Barrett hand. For two-finger grasps this procedure is based on the concept of virtual finger (Mackenzie and Iberall, 1994). It conceptually groups a pair of real fingers in the role of one theoretical finger.
In the case of three-finger grasps, we demonstrate that it is possible to find an infinite number of hand configurations that match the three contact locations of the original grasp (see section 4.4.3 on page 45). In order to choose one configuration we adopt the criterion of selecting those having forces exerted in directions as similar as possible to those in the original grasp. This criterion permits to minimize the search space.

In addition to these procedures, filters that discard hand configurations that are not mechanically feasible or reachable are included.

**Execution**

The final module refers to the control software that executes a particular hand configuration. A simple open-loop process that uses the location information provided by the vision system is implemented. The main feature is that it uses the force-sensors located in the finger-tips to feel the contact with the object. Once a hand configuration has been selected, the robot approaches the object from above and closes the hand using the computed configuration. When contact is perceived on all three fingers, the object is lifted.

**Experimental results**

A set of objects is built and a number of grasp trials are attempted. This experiments reveal a series of deficiencies and problems in the grasp execution, although many successful executions are observed, too. Most of the problems are a consequence of deficient hand configurations, that are computed without taking into account unforeseen hazards and problems. Moreover, the selection problem is still to be addressed. The learning approach we propose tries to address these questions by learning which features of a grasp indicate potential risks, making it more unreliable, and thus, less appropriate to be executed.

### 2.4 Grasp characterization

We introduce a characterization scheme to provide a way to describe grasps so that they can be used by the learning procedures. In our approach we have opted for a scheme that measures a set of properties of each grasp. In this way a grasp will be represented by $n$ measurements becoming a point in an $n$-dimensional space.

We have designed nine of these high-level features with the next requirements:

**Vision-based computation.** The features are computed from visually-extracted information.

**Hand constraining.** Features take into account particular characteristics of the hand.
Section 2.6 Experimental grasp reliability

Location and orientation invariance. Displacements and rotations of the object do not affect the values of the features.

Object independence. Grasps with the same physical properties have the same characterization independently of the object for which they are computed.

Physical meaning. Features are computed to measure physical properties relevant to grasping.

Stability and reliability. Features consider stability and reliability hazards of a grasp.

Since the description of each of the features must be detailed enough to understand their meaning and purpose the reader is referred to section 5.2 on page 58 for a full explanation of the nine features that compose the grasp characterization.

2.5 Experimental grasp reliability

We have implemented a test by means of which the robot can obtain an experimental measurement of the reliability of a grasp. Roughly explained, this test consists in executing the grasp which reliability is desired to know. The object is lifted and, then, three sequences of shaking movements are applied with increasing acceleration. After that, the object is returned to its initial location. During the whole process the tactile sensing capabilities of the hand are used to check whether the object is dropped or not. Depending on the phase during which the contact with the object is lost, if so, a qualitative category of reliability is assigned. Five different categories are defined.

Experimental database

In order to acquire a sample database large enough to validate the learning methods, a series of exhaustive experiments are carried out. For these experimental series up to thirty-six different configurations on four objects are selected. Each of them was executed up to twelve times, varying location and pose of the object, in order to obtain statistically significant results. Moreover, some physical properties of the environment—weight of the objects and friction of the materials in contact—were changed in order to have different sets of data.

After these series of experiments more than nine hundred samples were aggregated in four datasets of different physical conditions.

2.6 Prediction techniques: off-line learning

Each point in the sample datasets is characterized by a nine-element tuple and a reliability label. With these premises, we approach prediction as a classification problem. We propose three techniques that try to classify a query grasp (i.e: assign a reliability category). These techniques are based on well-known classification techniques.

Density estimation

This is a statistical approach. It assumes that the samples that belong to every reliability category are distributed in the feature-space according to a particular density function. In our implementation this is a multivariate normal density. We use the existent datasets to
estimate the parameters of this density functions, in our case, the mean and the covariance matrix.

**Voting K-nearest neighbors rule**

This is also a statistical method, but no assumptions are made about density distributions. Given a query grasp, it is assigned the category that more often appears among its $k$ nearest neighbors. The weight of each neighbor in the voting function is weighted by its distance to the query point. This method requires the application of a distance metric, in our case the Euclidean distance.

**Artificial neural networks**

Finally, a method based in the use of feed-forward back-propagation neural networks has been implemented, too. These neural networks have an input layer composed of nine neurons, and a final layer with five outputs, corresponding to the five categories.

**Validation and comparison of methods**

Several validation experiments are executed with the help of the sample datasets. The main purpose of these is to find out whether the proposed techniques are useful for classification within the framework of our problem. A secondary but important goal is to determine the number of necessary samples to reach the optimal classification performance.

In order to have a reference for comparisons we compute the theoretical performance obtained by a classification that would select randomly the category for a sample.

For the validation of the prediction functions we have developed several adaptations of the classical cross validation technique. This consists in splitting randomly the whole sample dataset in a training and a test set. The first is used to train the method, which is used to predict the samples in the test set. The performance is computed from the classification and misclassification rates. We use two variants called one-grasp-out and one-object-out validations, that consist in using repeatedly as test sets the points that belong to a given configuration or object, respectively.

Each dataset is used to train the three different prediction procedures and both validation procedures are used. A second analytical experiment consisting in varying exhaustively the size of the test dataset is also carried out. Extensive results and graphs are presented in section 6.5.2 on page 90 and 6.5.4 on page 94. The main conclusion that can be inferred from these analyses is our approach is able predict the reliability a grasp with a reasonable confidence. Another important conclusion is that the voting $k$-nearest neighbor approach is the most stable and obtains the best classifications performances. Moreover, it proofs to improve its results as the amount of experience available increases.

### 2.7 Active learning

Up to this point we have developed an offline, classical, learning approach. The classification techniques used make their prediction from a dataset of already collected points. This dataset has been collected exhaustively, and no considerations have been made about the best grasps to execute in order to improve the descriptive properties of the dataset.

We propose an active procedure that selects the next grasp to execute from a set of available candidates with the aim of improving the predictive performance of the accumulated
dataset. This is designed having in mind the normal operation of the robot, that at any given
time would have to select among a set of feasible candidates. During the exploration phase it
would experiment with the grasp that best can increase its knowledge about the task. While
during the exploitation phase it will simply choose the most reliable grasp candidate.

Since it has not been possible to run the active learning procedures on the real system in
Massachusetts we have designed a framework for testing the active learning procedures. This
makes use of the real sample datasets that have been already used in previous analyses.

The active selection procedure proposed consists in selecting the candidate with the
smaller prediction confidence. Since for each candidate a category is predicted, and this
prediction is associated to estimated probability, this approach selects the most doubtful
prediction. This is based in the intuition that these points are located in the proximity of
decision boundaries. As in previous experiments a reference methods which selects randomly
among the candidates is also defined for comparison.

Results and graphs presented in chapter 7 indicate that the proposed procedure improves
clearly the random selection function, and is able to reach minimum performance levels with
less than a hundred trials.
Chapter 8

Conclusions

This thesis has fully developed a complete grasping system able to grasp previously un-modelled objects using embedded theoretical knowledge about the stability conditions, and knowledge accumulated from the execution of past trials.

In the first part of the thesis we focus on the grasp synthesis procedures. We follow a practical approach. First, we use visual information to provide a description of the object shape. We have also developed a couple of algorithms that compute planar two and three-finger grasps using that description of the object. These algorithms take into account stability conditions to compute feasible grasps, and does this with a small computational effort.

A procedure to adapt the results of these grasp synthesis algorithms to the particular geometry of the three-fingered Barrett Hand is also presented. Again, this procedure needs low computational requirements. The combination of this set of algorithms leads to a grasp determination tool able to produce a number of feasible grasp well-suited to the characteristics of the hand. This tool has been integrated in a complete sensor-based grasping system, the UMass humanoid torso, that have been used to test and validate the usefulness of the above mentioned grasping tool.

The existence of a real system able to grasp an object following the directions provided by the grasp system procedures serves as experimental demonstration of their appropriateness, but also reveals the unforeseen limitations and problems. A number of experiments carried out with this system has demonstrated that an additional reliability checking of the computed grasps is necessary. We have developed a framework to learn how to predict the relevant reliability and stability of a grasp.

This framework is based on a novel characterization of a grasp, that consists of a set of high-level vision-based features or descriptors. These features are independent from the objects where the grasps were initially computed. This characterization is a solid base for the representation of the historical experience of the robot in grasping tasks.

In parallel to the development of the characterization features, a practical test of the reliability of the grasp has been also developed. The goal of this test is to obtain qualitative measurements of the stability of a grasp. Once this test has been developed, an exhaustive experimentation in order to obtain an extensive sample dataset has been carried out. This set has been extremely useful to develop and validate the algorithms introduced the last two chapters of the thesis.

Several methods for predicting the stability of a non-tested grasp based on previous experience has been used. This experience is stored by using the grasp characterization schema. To develop these methods we have used and adapted classical methods from the pattern classification field. In particular, we have develop a couple of prediction procedures
based on the use of artificial neural networks and non-parametric statistical techniques. Both methods have been validated using the sample data set previously collected, and they have proof their adequateness to the prediction task.

Finally a procedure that selects the next grasp to execute with the goal of improving the accumulated experience has been developed. The validation experiments have demonstrated that by using this method is possible to reach an optimal level of predictive performance with an smaller number of trials.

This last procedure, the predictive technique together with a grasp characterization scheme suitable for generalization constitute a framework for active learning in the field of grasp determination.

8.1 Publications

The work developed in this thesis has produced a number of publications. The most relevant of them are listed here:

**International journals**


**International conference proceedings**


National conference proceedings


8.2 Future work

Robot grasping is one of the most complex and sophisticated fields in robotics. The basics of manipulation are ruled by complex dynamic models of force, contacts and object properties. Moreover, the amazing versatility, and usefulness of the human model causes this field to be one of the most interesting and promising within robotic research.

In our opinion, in order to advance in this field are necessary approaches based in two main pillars: intensive use of sensor information, and smart application of artificial intelligence techniques. This thesis has tried to follow these trends. Several lines can be followed from the current state of the work

- **Complete validation of the learning method in a the working system.** We only have shown simulations of the active learning system working over the sample datasets. The final accomplishment of this work wold be to validate the learning system working on the real robot. First, the learning procedures must be improved to solve the problems that were observed. Ans second, the trade off between exploration and exploitation has to be determined.

- **Use of learning techniques on other modules.** Machine learning techniques can be used in many ways. They have been used widely in control problems. In our case, an interesting problem where they can be applied is the identification and recognizing of objects. In the same way that grasps are stored, objects can also be stored along with the grasps computed on them. Marin et al. (2003) describes a educational system that makes use of this principle for two-finger grasps on planar objects.

- **Grasp synthesis of three-dimensional object by means of vision.** On this thesis we have focus on planar objects. However, this is a serious limitation for practical applications in service robotics. It is necessary to face the problem of three dimensional objects. Traditionally, the main obstacle for this approache is represented by the lack of versatile visual processing methods that provide the information required by the
grasp synthesis algorithms. Usually, the solution for this consists in simplifying the scene to make visual processing easier. Our approach is a good example.

A more ambitious approach is to adapt the grasp synthesis algorithms to the limited information provided by vision. This can consist in simplifying the object description to a few shape primitives in the same way that some knowledge approaches do. Another approach could consist in searching in the image for local features of the objects particularly appropriate for grasping actions, e.g. planar surfaces, hooks, bars, holes, etcetera.

The uncertainty derived from these approaches has to be managed by the use of sensorial feedback and by the design of robust grasp procedures.

- **Sensorial feedback of grasping actions.** The use of sensorial information is the most effective way of reducing the uncertainty and improving robustness. Thus, the immediate consequence of this is an intensive use of sensors in the control of manipulation actions. Vision plays an important role in the grasp synthesis problem, but it can be applied to other related fields. Visual servoing of manipulation actions is the most obvious, and there already exist a vast literature on these approaches. Recatala (2003) presents a good example on how visual servoing can be applied to the grasping problem.

A much interesting problem is the fusion of visual and tactile information for dexterous manipulation tasks. Little work has been done in this line, and it would be worthy to advance in this path.
Appendix A

Vertex enumeration problem

The vertex enumeration problem\(^1\) is one of the subproblems that arises the set of LP problems named polyhedral computation. These are general problems for multidimensional (mostly Euclidean) spaces.

A convex polyhedron is a subset \(P\) of \(\mathbb{R}^d\) such that it is the set of solutions to a finite system of inequalities, if its bounded it is called a convex polytope.

Given these definitions two basic problems arise. The first one is the convex hull problem. For a subset \(S\) of \(\mathbb{R}^d\), the convex hull \(\text{conv}(S)\) is defined as the smallest convex set in \(\mathbb{R}^d\) containing \(S\). The convex hull computation means the determination of \(\text{conv}(S)\) for a given set of \(n\) points \(S = \{p^1, p^2, \ldots, p^n\} \subset \mathbb{R}^d\), which are the vertexes of a polyhedron.

The usual way to determine \(\text{conv}(S)\) is to represent it as the intersection of half-spaces, or more precisely, as a set of solutions to a minimal system of linear inequalities. This amounts to output a matrix \(A \in \mathbb{R}^{m \times d}\) and a vector \(b \in \mathbb{R}^d\) for some \(m\) such that \(\text{conv}(S) = \{A|Ax \leq b\}\).

A related problem is the vertex enumeration problem. Given a matrix \(A \in \mathbb{R}^{m \times d}\) and a vector \(b \in \mathbb{R}^d\), such that \(Ax \leq b\) is a set of finite inequalities or half-spaces. The vertex enumeration problem is the determination of whether that finite system defines a convex polyhedron and if it does which are the vertexes of that polyhedron.

There is a large number of formal algorithms and implementations (Fukuda, 2000) that solve these problems efficiently for a large number of dimensions and inequalities, and is beyond the goal of this appendix the explanation of such algorithms, but the above reference and Chávtal (1983) are good starting points for further interest on these topics.

For our implementation, we have proposed a case of the vertex enumeration problem for inequality systems in \(\mathbb{R}^2\). We have also used the cddlib library Fukuda (1999) which is an implementation in C of several solving algorithms for this problem.

\(^1\)The content of this section has been extracted mainly from (Fukuda, 2000)
Bibliography


