

# Two-Fingered Grasp Planning for Randomized Bin-Picking: Determining the Best Pick

by

Donna C. Dupuis

B.A.Sc., Queen's University, 2006

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

Master of Applied Science

in

The Faculty of Graduate Studies

(Electrical and Computer Engineering)

The University of British Columbia

Vancouver

August, 2009

© Donna C. Dupuis 2009

# Abstract

For many years, the manufacturing industry has pursued a commercially viable, vision-guided, robotic bin-picking system. The goal of such a system is to select a target part and a corresponding grasp from a pile of jumbled parts. Strategic planning of this selection to reduce the risk of a failed grasp attempt would increase the system’s reliability, and, thus, its commercial viability, and is the focus of this thesis. Specifically, this work aims to find the best *pick*; namely, the best combination of a target part and corresponding grasp.

The primary contribution of this work is a novel method for generating many high-quality, rated, pick options for a given vision-guided robotic bin-picking cycle, enabling the selection of the best pick. The method is tailored for a two-fingered (antipodal) gripper, typically used in industry; however, it may be extended to other gripper types (i.e., three-fingered). The method is broken down into two stages: (1) offline generation of many high-quality two-fingered grasps for a given part, and (2) online evaluation of these grasps in the context of the pile to determine a collision-free set of rated picks, and, ultimately, the most desirable pick. In evaluating grasps online, the effect of gripper finger clearance is considered to further minimize the risk of collision when executing the selected pick.

Subsidiary contributions of this work include: (1) an automatic grasp-generation method to sample the space of all two-fingered grasps for the target part, (2) a metric function for evaluating grasps, and (3) a measure of the robustness of a grasp.

The proposed method for pick selection is validated using stereo data of a real pile of parts. We compare the use of a small set of nominal grasps for pick selection (an approach typical in industry) to the use of an extensive evaluated grasp set generated using the proposed method. Our experimental results show that, in the majority of cases, the use of our method results in more valid and higher quality picks.

# Table of Contents

<b>Abstract</b> . . . . .	ii
<b>Table of Contents</b> . . . . .	iii
<b>List of Tables</b> . . . . .	v
<b>List of Figures</b> . . . . .	vi
<b>Acronyms</b> . . . . .	vii
<b>Notation</b> . . . . .	viii
<b>Acknowledgements</b> . . . . .	x
<b>Dedication</b> . . . . .	xi
<b>1 Introduction</b> . . . . .	1
1.1 Motivation . . . . .	1
1.2 Context, Goals, and Objectives . . . . .	2
1.3 Contributions . . . . .	4
1.4 Outline of Thesis . . . . .	7
<b>2 Background and Related Work</b> . . . . .	8
2.1 Bin-Picking: The General Problem . . . . .	8
2.2 Increasing the Number of Picks . . . . .	9
2.3 Automatic Grasp Generation . . . . .	10
2.4 Grasp Evaluation . . . . .	11
2.4.1 Grasp Stability . . . . .	11
2.4.2 Grasp Robustness . . . . .	13
2.5 Randomly Stored Objects: Selecting a Good Pick . . . . .	14
2.6 Summary . . . . .	15

*Table of Contents*

---

<b>3</b>	<b>Methodology</b>	16
3.1	Offline: High-Quality Grasp Generation	16
3.1.1	Densely Sampling the Grasp Space	19
3.1.2	Grasp Evaluation	24
3.2	Online: Determining the Best Pick	26
3.2.1	Generating a Model of the Pile Surface	27
3.2.2	Computing Clear Picks	32
3.3	Summary	35
<b>4</b>	<b>Experiments and Results</b>	36
4.1	Creating a Densely-Sampled Set of Evaluated Grasps	36
4.2	Determining the Best Pick: Validating the Proposed Method	39
4.2.1	Determining the Best Pick in Simulation: Earlier Work	42
4.2.2	Determining the Best Pick: Using Stereo Data of a Real Pile	45
4.3	Summary	51
<b>5</b>	<b>Conclusions and Future Work</b>	52
5.1	Conclusions	52
5.2	Recommendations for Future Work	53
	<b>Bibliography</b>	54
	<b>Appendices</b>	57
<b>A</b>	<b>Gamma-Type Probability Distributions</b>	57
<b>B</b>	<b>Histogram Data: Number of Picks</b>	58

# List of Tables

4.1	Input parameters used for grasp generation and evaluation . .	37
4.2	Grasp generation results . . . . .	37
4.3	Input parameters used for simulated experiment . . . . .	43
4.4	Results from simulated experiment . . . . .	43
4.5	Summary of input parameters for pick evaluation experiment	46
4.6	Scaling parameters used for gripper finger clearance . . . . .	46
4.7	Results of evaluating picks for two grasp sets . . . . .	47
4.8	Statistical analysis summary for experiment using real stereo data . . . . .	48

# List of Figures

1.1	Sample grasp using standard industrial gripper . . . . .	4
1.2	Flowchart illustration of proposed method . . . . .	5
2.1	Illustration of contact friction cone [24]. . . . .	13
3.1	Illustration of grasp set hierarchy . . . . .	18
3.2	Part model and corresponding wire-frame . . . . .	20
3.3	Illustration of gripper and sampling directions . . . . .	21
3.4	Grasp Generator Algorithm . . . . .	22
3.5	Generating grasp from 2-D cross-section . . . . .	23
3.6	Procedure for generating many high-quality picks (performed online at each cycle). . . . .	27
3.7	Procedure for generating pile surface mesh . . . . .	28
3.8	Iterative Averaging Algorithm . . . . .	30
3.9	Illustration of generating pile surface model . . . . .	31
3.10	Procedure for computing clear picks . . . . .	33
3.11	Scaling directions of gripper . . . . .	34
4.1	Visualization of generated, evaluated grasp set . . . . .	38
4.2	Nominal grasps used for experiments . . . . .	41
4.3	Comparison of valid candidates using method from earlier work	44
4.4	Comparison of number of clear picks for two grasp sets at three levels of clearance . . . . .	48
4.5	Comparison of best pick generated from two grasp sets . . . . .	50
B.1	Histogram data: Number of picks . . . . .	58

# Acronyms

**VGRBP** Vision-Guided Robotic Bin-Picking

**GWS** Grasp Wrench Space

# Notation

$S_p$  - a collection of line segments that comprise a wire-frame approximating the skeleton of the part (see Figure 3.2)

$N_{sp}$  - number of line segments comprising  $S_p$

$s_i$  - single line segment within  $S_p$

$L_i$  - length of  $s_i$

$\Psi_g$  - 2-D region of space between the fully-opened gripper fingers, located at the gripper fingertips (see Figure 3.3)

$d$  - linear translation parameter along a line segment

$\theta$  - axial rotation parameter about the z-axis of the current part wire-frame line segment

$\phi$  - current-frame rotation parameter about the pinching (or sliding) direction of the gripper, defined in the plane of  $\Psi_g$ ; the pinching direction is always perpendicular to the line segment

$\Delta d$  - translational step-size for  $d$

$\Delta\theta$  - rotational step-size for  $\theta$

$\Delta\phi$  - rotational step-size for  $\phi$

$\varepsilon$  - tolerance parameter to model compliance of soft gripper finger contacts

$i$  - index variable for accessing or enumerating grasps from a set of grasps

$g$  - a grasp

$g_i$  -  $i^{th}$  grasp

$r$  - grasp robustness measure

$r_i$  - robustness of the  $i^{th}$  grasp

## Notation

---

$q$  - grasp stability measure

$q_i$  - stability of the  $i^{th}$  grasp

$\gamma$  - parameter used to emphasize grasp stability over robustness ( $\gamma > 1$ )

$Q$  - overall grasp quality measure

$Q_i$  - overall quality of the  $i^{th}$  grasp

$\{G_{FULL}\}$  - set of all generated robust grasps; generated by sampling the grasp space

$\{G\}$  - subset of  $\eta$  highest quality grasps from the generated grasp set,  $\{G_{FULL}\}$

$\eta$  - number of highest quality generated grasps from  $\{G_{FULL}\}$  comprising  $\{G\}$

$\{N\}$  - set of  $k$  nominal, “intuitive” grasps

$k$  - number of nominal grasps comprising  $\{N\}$

$D$  - initial disparity map

$D'$  - final filled-in disparity map (has same dimensions as  $\{D\}$ )

$h_D$  - height of the disparity map, in pixels

$w_D$  - width of the disparity map, in pixels

$D(m, n)$  - disparity value at row  $m$  and column  $n$  in the map,  $D$

$m$  - row coordinate (non-negative integer)

$n$  - column coordinate (non-negative integer)

$B_{init}$  - user-defined disparity initialization parameter

$\Delta x$  - user-defined step-size parameter ( $0 < \Delta x < 1$ )

$DIFF_{max}$  - maximum difference between a disparity value  $D(m, n)$  and the average disparity of the four nearest neighbours of pixel  $(m, n)$

$DIFF_{init}$  - user-defined initialization parameter for  $DIFF_{max}$

$THRES_{stop}$  - non-negative user-defined threshold parameter that determines when the algorithm stops

# Acknowledgements

First and foremost, I would like to thank my supervisors, Dr. Elizabeth A. Croft and Dr. James J. Little, for their guidance and support that has made this thesis come to fruition. As well, I would like to thank fellow CARIS lab members Simon Léonard and Matthew Baumann for their input and advice that has been invaluable along the way.

I am fortunate to have enjoyed such a positive and memorable working environment in the CARIS lab, and thank my colleagues who contributed to that. We learned, and we had fun. I am also very thankful for the moral support provided by my friends and family.

I gratefully acknowledge the financial support of Dr. Croft and Dr. Little, the National Sciences and Engineering Research Council of Canada, Precarn Inc., Braintech Inc., and the Department of Electrical and Computer Engineering at UBC, as well as the facilitative support of Dr. Sid Fels.

*To my loving parents, whose unfailing support and encouragement were only ever a phone call away, regardless of the time of day. And to Nonna - your courage and strength are my inspiration.*

# Chapter 1

## Introduction

### 1.1 Motivation

The problem of robotic bin-picking is well-known in manufacturing. In this problem, a storage bin contains many randomly oriented industrial parts, and a robot must repeatedly recognize a part within the bin, grasp and manipulate the part, and deliver it to some target location. Typically, the parts are rigid and comprise complex surface geometry (e.g., with beveled edges or curved lips), and CAD model data or analytical geometric descriptions of the parts are often unavailable or difficult to obtain. For over thirty years, there have been efforts to successfully commercialize this process using intelligent robots equipped with vision systems to quickly and reliably identify and locate a part, move toward and grasp the part, and then remove it from the bin safely. Along with the grasping problem, a number of other problems must also be addressed, including collision-free motion planning, pose estimation, occlusion avoidance, visual tracking, and servoing, and are considered in recent works by Chan [8], Baumann [2], Léonard [20], and others.

However, over the last decade, the bin-picking problem has not received as much attention. This is not because the problem was solved, but rather, at the time, the manufacturing industry was still at the stage where specialized automation for singulation, fixturing, and palletizing of parts were more efficient and effective, due to sufficiently large enough line volumes, than moving forward with a vision-based robotic bin-picking solution. Today, the manufacturing industry is facing a major paradigm shift. For example, product lines that used to run for years are now being updated and revised in shorter and shorter cycles, approaching numbers of months rather than years. This change puts a great deal of pressure on fixed manufacturing lines. Furthermore, only recently has the technology in the fields of computers, robotics, vision, and lighting systems become sufficiently advanced to make an industrially feasible robotic bin-picking system possible. For these reasons, there is a renewed interest towards developing a commercially viable vision-guided robotic bin-picking (VGRBP) solution.

Such a VGRBP system must be highly reliable; that is, it must meet the following requirements:

1. Continual picking of parts out of a bin while avoiding collisions with obstacles.
2. Average cycle time not exceeding a predetermined length per successful pick-and-place operation.

It is estimated that a VGRBP solution should have an average cycle time of 10 seconds or less per operation. Because parts are randomly-situated within a bin, meeting this second requirement is challenging, and is one of the main reasons why randomized bin-picking systems have yet to be widely adopted by industry.

One of the key challenges in bin-picking is that a single part can take on virtually any pose within the jumbled pile inside the bin, and can be easily obstructed by the other parts. Even when a part is recognized and localized, a pre-selected “nominal” grasp for retrieving the part does not guarantee successful picking. Thus, in order to meet the requirement of reliability, a commercially viable VGRBP system must be able to provide many good picking options at each cycle. The work presented herein addresses this need. This work also finds relevance in the applications of automated luggage handling and fruit-picking.

## 1.2 Context, Goals, and Objectives

In VGRBP, a cycle is described as a single pick-and-place operation. During each cycle, the system must determine (1) a target part in the pile (selected from a set of candidates that have been identified by the computer vision system) and (2) a collision-free, stable grasp for that target part, where a grasp is defined by the gripper’s pose (position and orientation) relative to the part and a corresponding set of contact points on the surface of the part. Herein, these two items shall collectively be referred to as a *pick*. Thus, two distinct grasps for the same target part represent two distinct picks. Likewise, the same grasp used for two distinct parts also represents two distinct picks. Ideally, the system should be able to determine and successfully execute a viable pick at every cycle.

If there are multiple pick options, the system must attempt to select the best one. This choice is relevant because it can impact system reliability. Assuming the VGRBP system can recognize and localize parts within the

pile, a “naive” method for pick selection would select a target part at random from the set of localized parts, and then attempt to retrieve it using a standard “nominal” grasp. However, due to the random nature of parts jumbled in a bin, some parts are more accessible and easier to retrieve from the pile than others, and it might not be possible for the robot to successfully extract the target part from the pile. Furthermore, it is possible that, under this approach, no stable, collision-free grasps exist for the target; in this scenario, time-consuming intervention is required (e.g., stirring or shaking the pile, or even manual operator intervention).

Therefore, it is desirable to have a large number of viable pick options at a given cycle to reduce the likelihood of an unsuccessful, or excessively time-consuming, part retrieval attempt. Having many pick options necessitates a method for evaluating picks in order to select the best one. The “best” pick would ideally be that which increases system reliability by (1) minimizing the likelihood of a collision between the robot arm and environmental obstacles (i.e., the bin, or other parts in the pile), and (2) minimizing the likelihood of the part slipping out of the grasp (i.e., prematurely dropping the part). Thus, the broad goals of this work are:

1. To reduce the probability of an unsuccessful part retrieval attempt by increasing the likelihood of finding a high-quality, viable pick at every cycle.
2. To develop a way to evaluate candidate picks.

We limit the scope of this problem to a two-fingered (antipodal) gripper like that shown in Figure 1.1, as this type of gripper is commonly used in industry [30]. Throughout this work, we use a connecting rod, or con-rod as our exemplar part. Con-rods are a common engine part and typical in size and shape of many parts delivered in bins to the assembly process.

The achievement of the aforementioned goals is realized through (1) a formally-expressed, novel, metric function, developed for the purpose of evaluating candidate grasps (presented in Section 3.1.2), and (2) a novel method for generating multiple high-quality pick options at each cycle that uses grasp evaluation to evaluate picks (presented in Section 3). An overview of the method described in (2) is illustrated in Figure 1.2. We experimentally validate the proposed method in (2) using stereo data of real piles of parts (Section 4), since the use of a stereo camera sensor to obtain 3-D information is currently being developed for a commercial bin-picking system [6]. In our experiments, we compare the use of a small set of nominal grasps for



Figure 1.1: Standard industrial two-fingered (antipodal) gripper using a nominal grasp to grip a connecting rod, or con-rod [17].

pick selection (an approach typical in industry) to the use of an extensive evaluated grasp set generated within the proposed method.

### 1.3 Contributions

The primary contribution of this work is a method for generating multiple high-quality (rated) pick options, enabling selection of best pick, for a given cycle in VGRBP. Our method is broken down into two stages: (1) offline generation of many high-quality, two-fingered grasps for a given part, and (2) online evaluation of these grasps in the context of the pile to determine a collision-free set of rated picks, and, thus, the most desirable pick. This method is illustrated in Figure 1.2. Inputs to the method include the part model and corresponding wire-frame, the gripper model (that must be an antipodal, two-fingered gripper), stereo images (taken at each cycle), and system parameters (not shown in figure). The output of the offline computa-

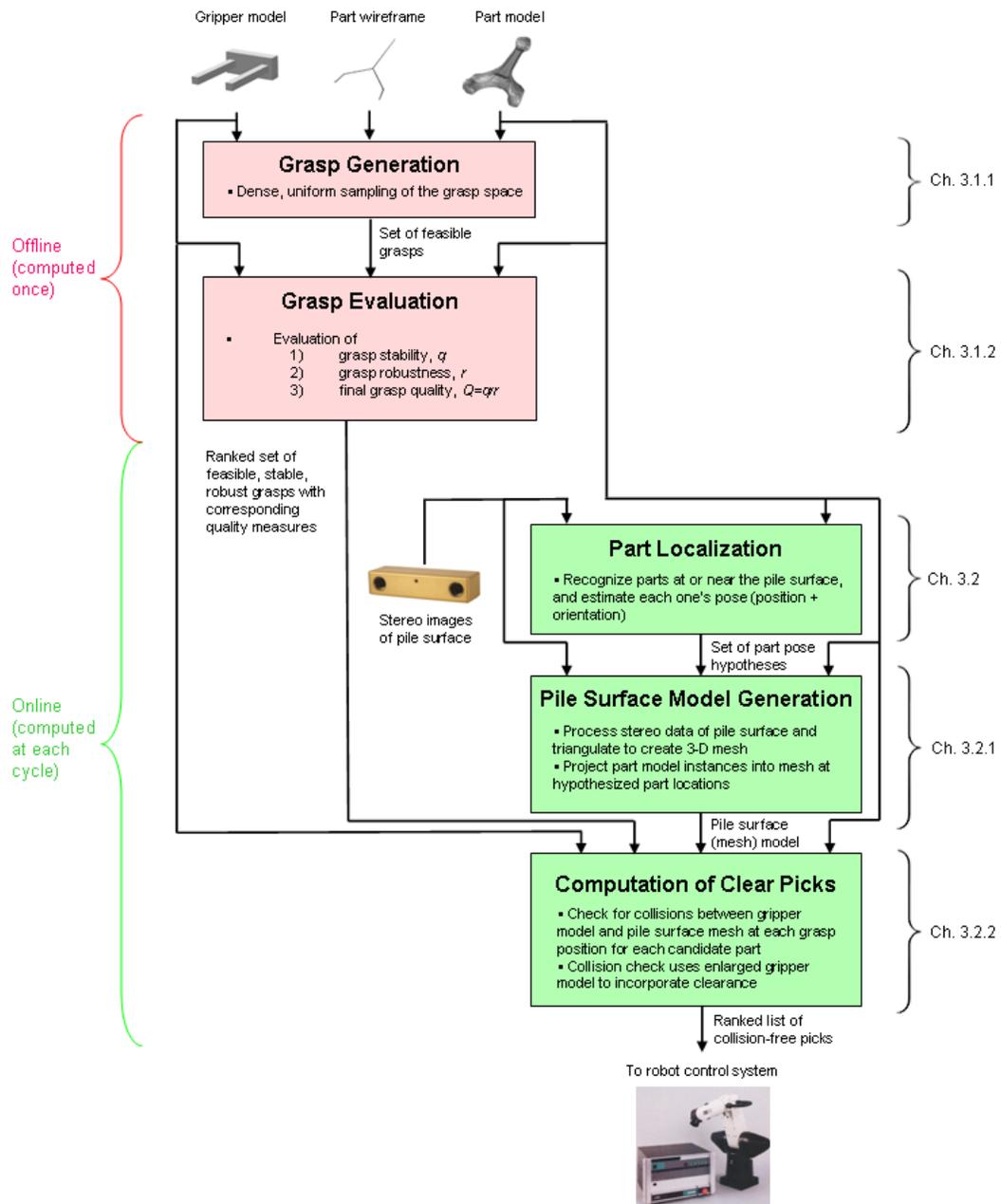


Figure 1.2: Illustration of the proposed method for generating many high-quality pick options. See Section 1.3 for details.

tion is a ranked set of feasible, stable, robust grasps. A subset comprising the highest-quality grasps is used, online, to compute high-quality, collision-free picks. The resulting set of rated picks is passed to the robot control system to select the highest-quality pick that is feasible given the robot workspace, collision, and joint limit constraints. In evaluating grasps online, the concept of gripper finger clearance is taken into account to further minimize the risk of collision when executing the selected pick; providing gripper finger clearance creates a buffer for the various sources of error in the system, such as robot positional error, object pose estimation error, and noisy or uncertain stereo data.

Other contributions of this work include: (1) a method for automatic grasp-generation, to sample the space of all two-fingered (antipodal) grasps for the choice part, (2) a metric function for evaluating grasps, and (3) a measure of the robustness of a grasp. All three are necessary to the aforementioned method for generating multiple high-quality pick options.

The automatic grasp-generation method enables sampling of the space of all two-fingered grasps, and it may be used on any object for which a representative wire-frame model can be generated. Thus, one input to this method is the wire-frame model of the object to be picked. Herein, we define the wire-frame manually; automatic wire-frame model generation is possible; however, is beyond the scope of this work. Mesh surface models of both the object to be picked and the gripper are also required as inputs. The user specifies rotational and translational step-sizes to determine the density of the sampling. To simplify the complexity of the problem, only planar grasps are considered, i.e., the grasp contact points on the object's surface lie in a plane. The output is a set of feasible (i.e., collision-free), stable, robust grasps for the specified object, whereby a grasp is defined by the pose (position and orientation) of the gripper relative to the part, and a corresponding set of contact points on the surface of the part.

The metric function contribution enables quantitative evaluation of grasps for the specified part. It considers grasp stability (evaluated using a pre-existing approach that is described in Section 2.4.1) and grasp robustness, where the measure of robustness is also a contribution of this work, and is a measure of a grasp's insensitivity to slight positional changes.

The proposed method for generating many high-quality pick options (enabling selection of the best viable pick) has the potential to increase the reliability of a VGRBP system by decreasing the likelihood of a failed pick attempt on a given cycle. Increased reliability in turn leads to increased commercial viability of VGRBP systems employing the proposed method, making it more likely that such systems will be widely adopted by indus-

try. Widespread adoption of reliable VGRBP systems is desirable because these systems reduce expensive fixturing costs and labour costs, leading to increased profit margins for industry.

In addition, these contributions expand the body of grasp planning research, especially grasp planning in the presence of environmental obstacles, a topic not commonly addressed in the literature; most research in grasp planning assumes the target object is isolated or clear to be grasped. This is relevant to any industry that involves robot manipulation tasks, e.g., in space exploration, manufacturing, and assistive robotics.

## **1.4 Outline of Thesis**

This thesis is organized as follows: Section 2 provides background information and a summary of related work, Section 3 describes the proposed methodology for evaluating picks (including the offline process of grasp generation and evaluation and the online process of generating a surface mesh of the pile and determining the best pick), Section 4 describes the experiments and results, and Section 5 concludes the thesis and discusses future work.

## Chapter 2

# Background and Related Work

Although there is little existing literature that directly addresses the problem of finding the best viable pick in the context of bin-picking specifically, this problem involves various aspects of computer vision and robotic grasping. In this chapter, the bin-picking problem is reviewed, and these aspects relevant to the “best pick” problem are discussed.

### 2.1 Bin-Picking: The General Problem

The general task of transferring a collection of like objects, one at a time, from one location to another as part of an assembly line process is prevalent in manufacturing. In order to avoid expensive fixturing, tooling, and component feeders, as well as labour costs, there has been a strong desire in industry to have an automated bin-picking system, in which a robot efficiently picks randomly-oriented parts out of a bin with the aid of a computer vision system, and delivers them to another location within the assembly line.

However, a commercially viable system to perform such a task has proven elusive due to a number of factors. For one, it is difficult and computationally expensive to automatically recognize and localize parts within a bin, as the parts can take on any random orientation, and can obstruct each other. Also, the path that the robot must execute to retrieve a part is variable, as it is dependent on the selected target part, and must be recomputed at each cycle; this is not a trivial task, since the robot can potentially collide with many obstacles. Furthermore, an isolated object may be grasped in any number of ways, but, due to the random nature of a pile of parts, only certain grasps are possible for a given part within the pile; thus, planning a good picking option (i.e., one that yields a collision-free, stable, robust grasp that is within the robot’s reachable workspace) must also be computed at each cycle. These are only some of the many challenges involved in developing a fully-automated robotic bin-picking system.

Because of these challenges, the kinds of systems currently in place in assembly lines tend to be semi-structured. For example, parts may be singulated on a moving conveyer belt [18], so the task of picking becomes a 2-D vision robot-guidance task, which is much less computationally taxing.

Recently, however, technology has advanced enough in the areas of computer vision and robotics to make it possible to realize a fully-automated robotic bin-picking solution for an unstructured bin of parts. Currently, there is a great deal of variety in the proposed systems (i.e., sensor types, sensor placements, computer vision algorithms, and grasp planning approaches), and substantial research has been done on sub-problems related to the task of bin-picking. Some systems use range and laser sensors to obtain the required depth information to locate objects within a randomized pile ([1], [5]), although the use of stereo cameras is more common, as they are faster and cheaper. We base this work on a robotic system that uses stereo vision. Variety also exists with the sensor placement (i.e., fixed, eye-in-hand, or a hybrid of the two) and lighting techniques for improved part localization (i.e., structured lighting, or switched lighting) [18]. This work assumes a fixed (stereo) camera mounted above the bin, and does not consider lighting techniques.

For a proposed bin-picking system to become widely adopted in industry, it must be reliable, i.e., it must be able to continually pick parts out of one or more bins while avoiding collisions with obstacles, and without the average cycle time exceeding a predetermined length per successful pick-and-place operation. (If the system is unreliable and fails on a given cycle, time-consuming intervention is required, which increases the average cycle time.) However, achieving a reliable system is difficult due to the inherent variability of a randomized pile of parts, and is one of the reasons why there exists so much variety in systems that have been proposed. Thus, improving reliability is the primary motivation for the work herein.

## 2.2 Increasing the Number of Picks

As previously discussed, meeting the requirement of reliability is challenging due to the nature of randomly-situated parts within the bin, and is one of the main reasons why VGRBP systems have yet to be widely adopted by industry. For example, given a set of pre-defined grasping points on a particular part, and a set of candidate parts within the context of a randomized bin, in many cases the pre-defined grasps are obstructed by neighbouring parts or by the walls of the bin. In such cases, the grasps are not feasible

since they would result in collisions with the gripper. If there is a limited number of pre-defined grasps, it is possible that all grasps for all candidate parts are infeasible, resulting in no viable options for picking.

In some systems, if no valid pick exists, a second attempt is made at locating a viable candidate; for example, by taking a closer look at the pile, or by mechanically stirring the parts [28], and then re-examining the pile. However, these solutions increase the cycle time.

Thus, one way to enhance reliability in a VGRBP system is to increase the number of pick options available to the robotic system at each cycle. This decreases the probability of having no viable pick options, and thus, decreases the probability of an unsuccessful part retrieval attempt. By definition, the number of possible picks depends on (1) the number of candidate parts localized by the vision system, and (2) the number of feasible grasps that are available for each candidate. Much computer vision research has been done for the purpose of enhancing recognition and localization of objects for VGRBP systems (i.e., addressing item (1)); if more objects can be found by the vision system, there are more candidates to choose from at each cycle. For example, the authors of [13] present a bin-picking system based on a technique called *active depth from defocus*, in which better recognition and localization of objects is achieved by improving 3-D range data for model-matching. Similarly, a hybrid coarse-to-fine stereo-matching algorithm geared towards the task of VGRBP is presented in [32]; the purpose is to improve the stereo data since accurate 3D information of the pile surface is required in order to localize parts. In [19], Harmonic Shape Contexts (HSC) features are used to improve part localization using 3-D data, for the application of bin-picking. Other examples are presented in [1] and [16].

However, there is little existing research that addresses item (2), the number of feasible grasps, which is part of the motivation for this work. Herein, we attempt to increase the number of picks by increasing the number of feasible grasps that are available for each candidate.

## 2.3 Automatic Grasp Generation

One way to increase the number of picks for a given VGRBP cycle is to increase the number of potential high-quality grasps for a given part by sampling the grasp space. Grasp sampling of a particular object for a given gripper is not a trivial task, and has been addressed by the authors of [25] and [14]. In [25], objects are represented as a collection of primitive 3-D shapes

(i.e., cones, spheres, and cubes), each of which is manually assigned a set of grasp starting positions and pre-grasp shapes. Each grasp starting position and pre-grasp shape collectively define a grasp (when the gripper fingers are closed at that configuration). The union of all such grasps from each primitive shape comprising the object results in a set of grasp samples. In [14], objects are represented by superquadratic decomposition trees in order to reduce the space of possible grasps, and the surface of each superquadratic is sampled at a uniform interval. Both methods can accommodate a variety of object types and gripper types (including two-, three- and five-fingered grippers), and involve abstracting the surface of the target object to simplify the problem. In our work, we present a method for densely sampling the grasp space that is tailored specifically for a two-fingered gripper (as this is commonly used in industry); this allows us to make the simplification of considering only planar, antipodal grasps. Instead of abstracting the target object’s surface, we abstract the target object’s volume using a wire-frame.

## 2.4 Grasp Evaluation

In order to determine the best pick, an evaluation method must be established. Since a pick is defined as, collectively, a target part and a corresponding grasp for that part, it follows that the quality of a pick will depend on the quality of its constituent grasp. Typically, grasps are evaluated with regards to stability. Much research has been done on quantifying the quality of a grasp in this way (e.g., see [11], [7], and [21]). Little research has been done on the topic of grasp robustness, which becomes important when evaluating grasps in a real system where there is error and uncertainty in the data and components of the system. The concepts of grasp stability and grasp robustness are discussed in Sections 2.4.1 and 2.4.2, respectively.

### 2.4.1 Grasp Stability

Analyzing grasps and evaluating grasp quality has been a research topic for many years. Typically, grasp quality measures describe the stability of a grasp. Since it is desirable in VGRBP to select picks that are highly stable, we incorporate grasp stability into our method for generating many high-quality picks. We use the open-source software *GraspIt!* [24] for evaluating grasps with regards to stability.

Before describing the grasp evaluation approach, we describe the following relevant terms:

*Friction cone* - a cone-shaped volume aligned with the normal component of a contact force applied to an object's surface. It defines the set of all forces that can be applied at the contact point without slippage. The size of the friction cone depends on the coefficient of friction between the contacting materials.

*Wrench* - a 6-D vector comprising a 3-D force and a 3-D moment.

*Force-closure* - a grasp is considered to be in force-closure if and only if it is in equilibrium for any arbitrary disturbance wrench [27]. This implies that if a grasp is force-closed, it can withstand any disturbance wrench (with enough strength).

*Grasp wrench space (GWS)* - the union of all wrenches that can be applied to an object by a grasp [24].

The method used by *GraspIt!* [24] for grasp evaluation is based on the widely-accepted grasp analysis method proposed by Ferrari and Canny [11]. In [11], grasp quality is described as the magnitude of the largest, worst-case, disturbance wrench that can be resisted by the grasp with a grip of unit strength. The process involves computing the GWS of the grasp under consideration, and is summarized by the following steps:

1. Determine the grasp contact points on the object surface.
2. Approximate the friction cone at each contact point as the convex sum of a set of finite force vectors around the cone boundary (see Figure 2.1).
3. Compute the corresponding object wrench for each force vector.
4. Compute the GWS as the convex hull of all wrenches.
5. Compute the grasp quality as the distance from the wrench space origin to the nearest point on the convex hull.

If the GWS does not contain the origin, the grasp is not considered to be force-closed, i.e., there exists some disturbance wrench that the grasp cannot resist (regardless of the forces applied by the grasping fingers). For a grasp to be considered stable, it must be force-closed. It should be noted that the quality measure provided by *GraspIt!* [24] does not take into account robustness.

*GraspIt!* [24] is the main tool used to support the quality analysis done in this work; for more in-depth background on grasp analysis, we direct the reader to [4].

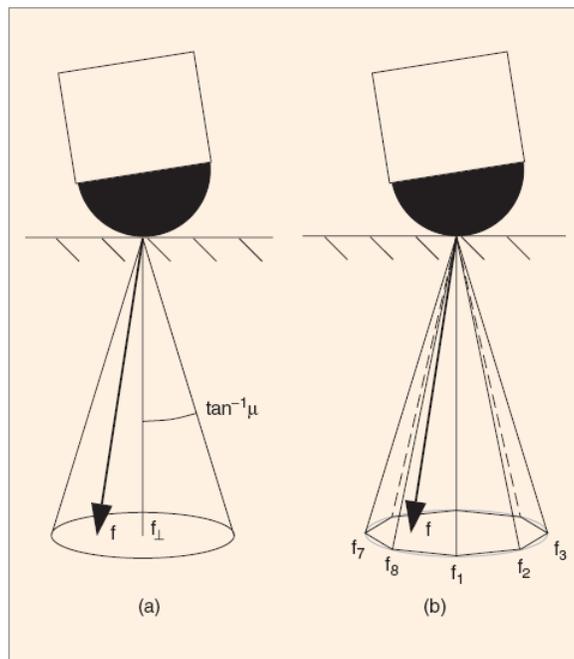


Figure 2.1: The figure, reproduced from [24], illustrates a contact point between the gripper and the object surface, and the corresponding friction cone. (a) The size of the friction cone depends on the coefficient of friction between the contacting materials,  $\mu$ . The total contact force,  $f$ , must lie within the friction cone to prevent slippage. (b) The friction cone is represented by the convex hull of a finite set of force vectors around the cone boundary.

#### 2.4.2 Grasp Robustness

In VGRBP systems, candidate parts must be recognized and localized using computer vision techniques. However, even with state-of-the-art computer vision algorithms, there is always some error in the pose (position and orientation) estimation of each recognized part. Furthermore, robot calibration is likely to have some error, resulting in errors in the gripper pose (position and orientation). These sources of error are compounded in the process of localizing a part and then moving the robot gripper to some desired grasp location on the surface of that part. Therefore, it is likely that the resulting grasp will be offset from the desired grasp, leading to an actual grasp that is

less stable than the planned grasp, or that results in an undesired collision between the gripper and other nearby parts, or the part itself.

One can expect that, by avoiding such scenarios, the VGRBP system will be more reliable. Therefore, to improve reliability, we desire to incorporate a preference for grasps that are “robust”, where we define “robustness” as the insensitivity of a grasp to small pose errors. There is precedent for this approach: the authors of [10] address the concept of robustness by examining the effect of rotational variation of a grasp on the quality of the grasp. However, the work in [10] does not take into account translational variation. Both translational and rotational variations are considered in the method proposed herein. In fact, in general, the concept of grasp robustness has had very little attention in the grasping literature. This may change as more autonomous robotic systems appear that are required to perform unstructured grasping tasks.

## 2.5 Randomly Stored Objects: Selecting a Good Pick

As just indicated, very little literature addresses the specific problem of selecting a good pick. The problem is not formalized well, and, usually, the focus is on selecting the best *part* to pick up, rather than the best *pick*. A common approach is to select the top-most object in the pile and attempt to apply a nominal grasp, as is done in [1], [13], and [16]. Locating the top-most object can be accomplished using image segmentation methods, such as those described in [31] and [12]. Although this would likely produce feasible picks in many cases, it is not clear which part to select when multiple parts are considered to be on top, or when parts are entangled such that no part can be clearly distinguished as being on top. Furthermore, selecting the top-most object does not guarantee either gripper finger clearance, or selection of the best available grasp.

The work in [3] presents a cost function for evaluating a set of 6-D *HPOs*, where an HPO is a Hand Position and Orientation, in the context of grasping an object in a cluttered environment. Factors considered are (1) whether or not the fixed part of the hand will be in collision, (2) the error of the fit between a preselected hand preshape and a potential target, and (3) the likelihood of the fingers being able to reach desired contact points without collision. Force-closure and grasp quality are not checked until after a set of HPOs are generated, so even if a given HPO has a low cost, and thus, a high probability of resulting in a valid grasp, it may not result in the most

stable grasp available. Also, the concept of grasp robustness is not explicitly addressed.

## 2.6 Summary

In this chapter, we described the bin-picking problem, and reviewed various aspects of computer vision and robotic grasping research relevant to the problem of determining the best pick for VGRBP. We described how increasing the number of picking options can enhance system reliability, and outlined two approaches for achieving this: (1) increasing the number of recognized candidate parts, and (2) increasing the number of feasible grasps for a given part. Much computer vision research has addressed (1), but little work has addressed (2) since densely sampling the grasp space of an object is a complex, non-trivial problem.

Once a list of potential picks is generated, there is a need to evaluate these picks to determine the best one. Since grasp quality affects pick quality, we summarized previous work that addresses grasp evaluation. We described the simulator *GraspIt!* [24] used herein to evaluate grasp stability, and the widely-accepted grasp analysis approach that *GraspIt!* is based on. We found that previous work focused on evaluating grasps with regards to stability only, and little work exists that considers grasp robustness. In general, there is little research addressing the problem of pick evaluation in the context of a randomized pile.

In the following chapter, we will address both of these issues, first looking at grasp robustness in an offline stage, and then pick selection in an online mode.

## Chapter 3

# Methodology

In this chapter, we detail the main contribution of this work: a method for generating multiple high-quality (rated) pick options, enabling selection of best pick, for a given cycle in VGRBP. Our method is broken down into two stages: an offline module (described in Section 3.1), followed by an online module (described in Section 3.2). In the offline portion, we generate many high-quality two-fingered grasps for a given part. First, we densely sample the space of all two-fingered grasps for the selected part, discarding any infeasible grasps. We then evaluate each feasible grasp using our proposed quality metric, which considers grasp stability and robustness. In the online portion, we evaluate a subset of the highest-quality grasps (that were generated offline) in the context of the pile, using stereo data of the pile surface, to determine a collision-free set of rated picks. Gripper finger clearance is taken into account to create a buffer for the various sources of error in the system, such as robot positional error, object pose estimation error, and noisy or uncertain stereo data. In the context of a VGRBP system, the resulting set of picks is then passed to the robot control system to select the highest-quality pick that is feasible, according to robot workspace and joint limit constraints. An overview of the method is illustrated in Figure 1.2 in Section 1.3.

### 3.1 Offline: High-Quality Grasp Generation

In the case of a manufactured part, ideally, information of all possible ways in which a given object can be robustly grasped could be generated from CAD (computer-aided design) data, since this information is useful for pick selection. In this work, we define a grasp as the 6-D pose, or configuration, of the gripper relative to the part, and the corresponding set of contact points on the surface of the part (defined by the pose). The 6-D configuration space of all robust grasps for an object and gripper is highly dependent on the object and gripper geometries, and is, typically, non-discrete and non-uniform; there are infinitely many ways to grasp the object, and grasp quality will likely vary widely over the space. Therefore, formulating a

closed-form, mathematical description of the robust grasp space is challenging, and, since it would vary from part to part, impractical. Furthermore, CAD data of a part is often unavailable, and analytical descriptions of part geometries is, generally, quite difficult to obtain. Alternatively, one could uniformly sample the grasp configuration space and evaluate each sampled grasp to generate a set of grasps that is a good representation of the grasp space, using empirical data describing the part geometry. Such empirical data may be obtained by laser-scanning the part. This approach has the benefits of being general and practical, as well as potentially extendable to new parts without CAD data, and is the approach we use.

Generating an extensive list of grasps for a given part can be computationally expensive, and is very difficult to compute online within the required time constraints. Typically, in the context of industrial bin-picking, a-priori knowledge of the part to be picked is available. This allows for offline generation and evaluation of grasps with little concern for computation time. Thus, the output of the offline module is a rated set of robust grasps. We summarize the offline process as follows:

1. Densely and uniformly sample the space of all two-fingered, antipodal grasps, discarding infeasible grasps.
2. Evaluate feasible grasps for stability, discarding unstable grasps.
3. Evaluate stable grasps for robustness, discarding non-robust grasps.
4. Evaluate robust grasps using proposed quality metric function.

Here, “feasible” means that the gripper does not collide with the part. Note that, if a grasp is evaluated as robust, it is also feasible and stable by definition (see Section 3.1.2). This grasp set hierarchy is illustrated in Figure 3.1.

This section is broken up into two parts. In Section 3.1.1, the approach for generating an extensive list of feasible grasps for an exemplar part - a connecting rod, or con-rod (a common automotive part) - is detailed. Section 3.1.2 describes how the quality of each grasp is evaluated.

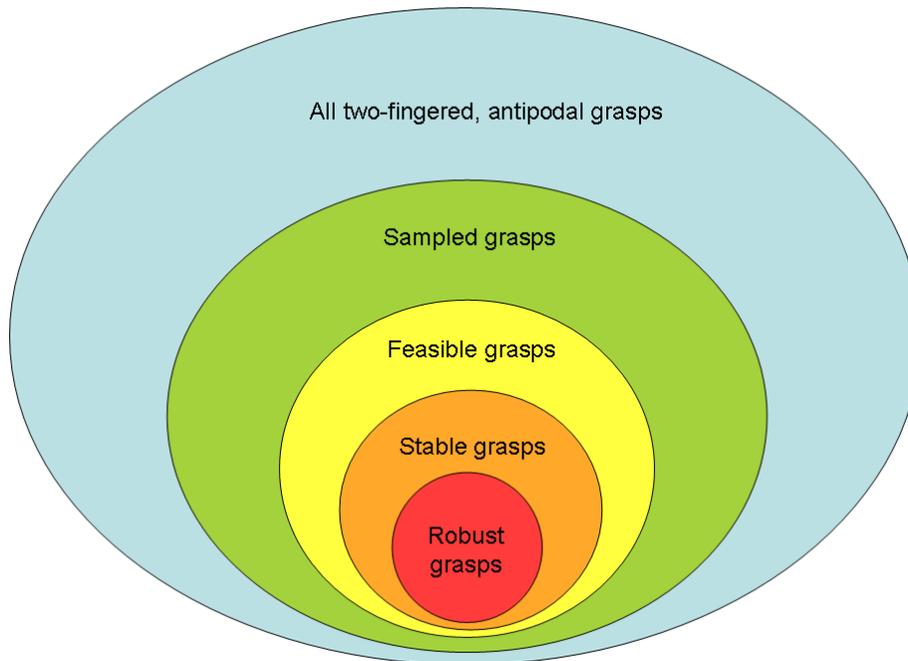


Figure 3.1: Illustration of grasp set hierarchy. If a grasp is evaluated as robust, it is also, by definition, feasible and stable.

### 3.1.1 Densely Sampling the Grasp Space

In order to densely sample the grasp space, the following system inputs are required: a surface model of the part, a corresponding wire-frame, and a surface model of the gripper. An example of a part model and corresponding wire-frame is shown in Figure 3.2.

For a standard industrial two-fingered gripper, we generate grasps at multiple positions and orientations by intersecting the space between the gripper fingers with the part at uniform intervals. To reduce the complexity of grasp generation, only planar grasps are considered; i.e., the grasp contact points lie in a plane. This choice enables us to model the space between the gripper fingers as a bounded 2-D (planar) region located at the gripper fingertips (see Figure 3.3). To formally describe this intersection process, we present the following definitions:

$S_p$  - a collection of line segments comprising a wire-frame that approximates the geometry of the part (see Figure 3.2)

$N_{sp}$  - the number of line segments comprising  $S_p$

$s_i$  - a single line segment within  $S_p$

$L_i$  - the length of  $s_i$

$\Psi_g$  - the 2-D region of space between the fully-opened gripper fingers, located at the gripper fingertips (see Figure 3.3)

$d$  - linear translation parameter along a line segment

$\theta$  - axial rotation parameter about the z-axis of the current part wire-frame line segment

$\phi$  - current-frame rotation parameter about the pinching (or sliding) direction of the gripper, defined in the plane of  $\Psi_g$ ; the pinching direction is always perpendicular to the current line segment

$\Delta d$  - translational step-size for  $d$

$\Delta \theta$  - rotational step-size for  $\theta$

$\Delta \phi$  - rotational step-size for  $\phi$

We define the wire-frame,  $S_p$ , manually (see Figure 3.2), and restrict the position of  $\Psi_g$  to points along  $S_p$ . The grasp generation algorithm is

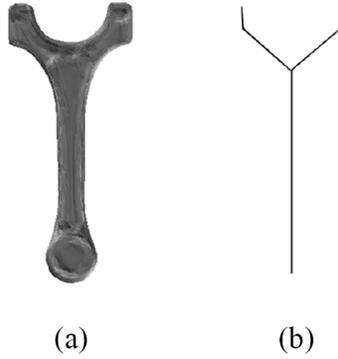


Figure 3.2: (a) Part model. (b) Corresponding wire-frame,  $S_p$ ;  $N_{sp} = 5$ .

described in Figure 3.4. In this algorithm, we translate the fully-opened gripper (and, correspondingly,  $\Psi_g$ ) in discrete steps along each  $s_i \in S_p$ , and at each translational step, we rotate  $\Psi_g$  through a sphere of discrete orientations. At each new pose of  $\Psi_g$ , the intersection between  $\Psi_g$  and the part is computed, resulting in a 2-D cross-section. Grasp points are defined at the extrema of the cross-section along the pinching direction, within a tolerance,  $\varepsilon$ , to account for soft gripper contacts (see Figure 3.5). Only grasps that do not result in a collision between the fully-opened gripper and the part are stored (otherwise, they are considered infeasible).

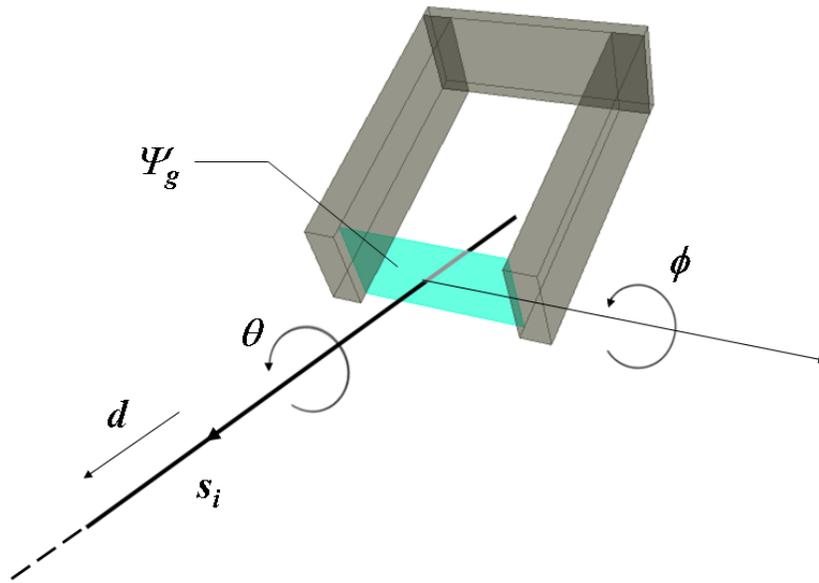


Figure 3.3: Illustration of  $\Psi_g$  (the 2-D region between the gripper fingertips), the gripper, and the sampling directions.

**Algorithm** GRASP\_GENERATOR

Let  $move(s_i, d, \theta, \phi)$  represent a function that translates and rotates gripper (and, correspondingly,  $\Psi_g$ ) to the pose defined by the input parameters

```

for  $i = 1$  to  $N_{sp}$ 
  for  $d = 0$  to  $L_i$ ; step  $\Delta d$ 
    for  $\theta = 0$  to  $(2\pi - \Delta\theta)$ ; step  $\Delta\theta$ 
      for  $\phi = 0$  to  $\pi$ ; step  $\Delta\phi$ 
         $move(s_i, d, \theta, \phi)$ 
        if fully-opened gripper does not collide with part // feasibility check
          compute intersection between  $\Psi_g$  and part
          compute grasp points from this intersection
          store grasp data (gripper pose + contact points)
        end if
      end for
    end for
  end for
end for

```

Figure 3.4: Grasp Generator Algorithm, which describes the intersection of  $\Psi_g$  with the part.

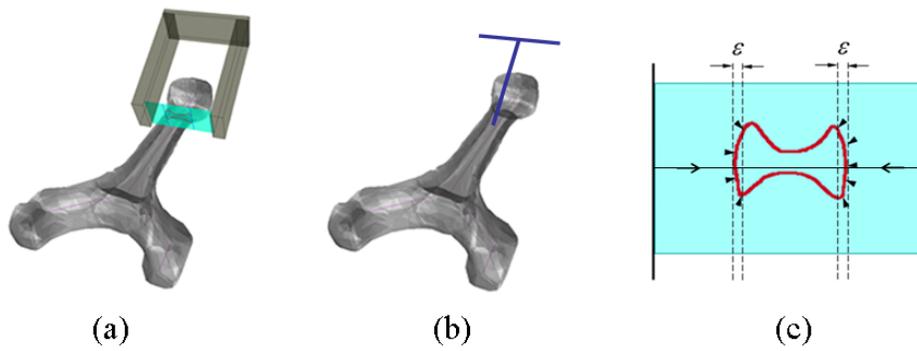


Figure 3.5: Illustration of generating a set of contact points from a 2-D cross-section. (a) Sample grasp. (b) Minimal representation of grasp using a  $T$  shape that depicts approach direction, pinching direction, and position of grasp (see Section 4.1 for details). (c) Contact points corresponding to sample grasp. Contacts are located at the extrema of the cross-section (indicated by the arrowheads) along the pinching direction of the gripper, within a tolerance,  $\epsilon$ .

Parameterizing the rotation with  $\theta$  and  $\phi$  ensures that the pinching direction of the gripper is perpendicular to the current wire-frame line segment. The justification for sampling this space is that grasps are, in general, more stable if the forces applied by the gripper fingers are perpendicular to the surfaces they contact; this minimizes the risk of slippage between the gripper fingers and the object.

The selection of the sampling step size is important to the grasp generation algorithm. Although a dense sampling is desired, there is a limit on the accuracy of the robot that would be used to grasp the part. It would be superfluous to use a step-size that is smaller than the positional error of the gripper. Thus, we use the robot’s positional accuracy as a lower bound on the translational step-size,  $\Delta d$ .

To uniformly sample the grasp space, it is desirable to use a similar step-size in all directions. This is complicated by the fact that one sample direction is translational, while the other two are rotational. To address this, we select rotational step-sizes ( $\Delta\theta$  and  $\Delta\phi$ ) such that the arc length spanned by each rotational step-size is approximately equal to  $\Delta d$ . In calculating arc length, we use a radius equal to the average radius of the part. Thus, the resulting rotational step-sizes are comparable to  $\Delta d$ , and dependent on  $\Delta d$  and the part geometry.

The grasp generation algorithm may be used to collect grasps for any object that can be roughly approximated by a wire-frame skeleton of line segments and that can be represented using a 3-D surface mesh model. The selected object should be small enough such that the gripper fingers can enclose some portion of the object, and light enough for the robot to lift without exceeding joint torque limits. Also, the algorithm is compatible with other gripper types, provided the gripper is capable of generating planar antipodal grasps (i.e., three-fingered). This provision allows us to model the space between the fully-opened gripper fingers with a 2-D planar region (namely,  $\Psi_g$ ) used for sampling the grasp space.

### 3.1.2 Grasp Evaluation

Herein, we consider both grasp stability and robustness in our evaluation of grasp quality. The simulator we use for evaluating grasp stability, *GraspIt!* [24], has been used in [25], [14], [10], and [26]. The *GraspIt!* quality measures are based on the magnitude of the largest disturbance wrench that can be resisted by a unit-strength grasp (see Section 2.4.1 for details). Henceforth, for a given grasp,  $g_i$ , we will refer to the *GraspIt!* quality measure as  $q_i$

and describe grasps with large values of  $q$  as being “highly stable”. A grasp is (technically) stable if  $q_i$  is greater than zero; therefore, we discard any grasps whose quality measure is less than or equal to zero.

We evaluate grasp robustness after evaluating grasp stability, since robustness depends on the stability evaluation. For a grasp,  $g_i$ , we denote the robustness as  $r_i$ . We describe robustness as a measure of the insensitivity of  $q_i$  (stability measure) to small variations in the pose (position and orientation) of the grasp. Robustness is important to consider for VGRBP since, due to the gripper pose error combined with the pose estimation error of the target part, the actual grasp is likely to be offset from the desired grasp. We propose that the robustness,  $r_i$ , of a grasp,  $g_i$ , is the the inverse of the standard deviation of  $q$  within a local region,  $\rho$ , centered on  $g_i$ . Thus, we present the following definition:

$$r_i = \frac{1}{\sqrt{\frac{\sum^{N_\rho} (q_j - \bar{q})^2}{N_\rho - 1}}}. \quad (3.1)$$

$N_\rho$  is the number of grasps within the local region,  $\rho$ , of the grasp in question, and  $\bar{q}$  is the mean stability measure within this region. Note that  $\rho$  is a discretized, 3-D region, with dimensions corresponding to the sampling directions,  $\theta$ ,  $\phi$ , and  $d$ , and the discretization is defined by the three sampling step-sizes. Thus,  $\rho$  can be visualized as a 3-D array, in which each array entry represents a single grasp with unique coordinates,  $(\theta, \phi, d)$ . The size of  $\rho$  is an input parameter, and is selected based on the position accuracy of the robotic system.

Herein, we impose the necessary condition that only feasible, stable grasps may be robust. Therefore, when we describe a grasp as being robust, we imply that it is also feasible and stable (see Figure 3.1). A feasible, stable grasp is considered to be robust (and is, therefore, accepted) if *all* neighbouring grasps satisfy the following three criteria:

1. They exist.
2. They are feasible (i.e., they will not result in collisions with the gripper).
3. They are stable ( $q_i > 0$ ).

Finally, we propose the following definition for the overall quality measure,  $Q_i$ , of a grasp  $g_i$ :

$$Q_i = q_i^\gamma \cdot r_i, \quad (3.2)$$

where  $q_i$  and  $r_i$  have each been normalized between 0 and 1 using the maximum value for each from their respective data sets, and  $\gamma$  is a tunable “stability” parameter that is greater than 1 that emphasizes grasp stability over robustness. It should be noted that changing the value of  $\gamma$  does not affect the ranking order of a set of evaluated grasps, provided that  $\gamma$  is greater than 1; it only affects the data spread. Henceforth, when we use the term “quality”, we are referring to  $Q$ .

Equation (3.2) ensures that the best grasps are those that both resist large disturbance wrenches and are insensitive to slight position changes. The factors are multiplied, rather than weighted and summed, since grasp quality depends on both factors simultaneously rather than either factor independently.

### 3.2 Online: Determining the Best Pick

In VGRBP, a 3-D vision system can be used to obtain a topographical map of the pile surface, providing information for part localization and obstacle avoidance (i.e., collision-checking) [6]. Our approach relies on such a system; we check for collisions between the pile and a list of pre-generated, evaluated grasps for each localized candidate part, at each cycle. Since each grasp-candidate combination is considered to be a unique pick, the resulting number of picks checked is the product of the number of grasps in the employed grasp set and the number of candidate parts. If a pick is collision-free, it is considered to be valid. To check for collisions between the gripper and the pile at various pick positions, we first generate a mesh model of the pile surface at each cycle (see Section 3.2.1). We then compute clear picks by projecting an enlarged version of the gripper into the pile at each pick position, and then performing a collision check (see Section 3.2.2). The gripper fingers are enlarged to incorporate a clearance buffer. The resulting valid picks are sorted according to each pick’s corresponding grasp quality measure,  $Q$ . The general procedure for generating many high-quality picks is described in Figure 3.6.

If only the part, the gripper, and the pile configuration were considered, the best pick would be the highest-quality clear pick. However, some picks may be impossible due to robot workspace and joint limit constraints. Therefore, in practice, an additional step is required to process the rated list to check for feasibility with the robot’s limits before finally selecting the highest-quality feasible pick. This final step may be executed using standard techniques, and, as such, is not discussed in detail herein.

**Procedure 1 - GENERATING MANY HIGH-QUALITY PICKS**

1. Localize set of candidate parts to pick from pile (located relative to camera frame).
2. Generate mesh surface model of pile (located relative to camera frame) according to Procedure 2 (see Figure 3.7).
3. Using (i) ranked set of highest-quality robust grasps (generated offline), (ii) pose estimates for localized parts, (iii) gripper model, and (iv) pile surface model, compute and rank collision-free picks according to Procedure 3 (see Figure 3.10).
4. Return ranked list of clear picks to robot control system to select the best feasible pick.

Figure 3.6: Overview of the procedure for generating many high-quality picks.

### 3.2.1 Generating a Model of the Pile Surface

In order to check for collisions between the pile and the gripper at various pick positions, a 3-D surface model of the pile is required. Thus, we generate a mesh of the pile surface using stereo data, according to the procedure described in Figure 3.7 involving standard mesh processing techniques.

**Procedure 2 - CREATING PILE SURFACE MODEL**

1. Localize a set of candidate parts to pick from the pile (located relative to the camera frame) using existing computer vision methods.
2. Take a snapshot of the pile using a stereo camera positioned directly above the pile, and generate a corresponding disparity map.
3. Cull statistical outliers in the disparity data.
4. Fill in regions of the disparity map that are invalid/unknown using Iterative Averaging Algorithm (described in Figure 3.8).
5. Convert disparity values at each pixel coordinate to  $(x, y, z)$  coordinates (relative to the camera frame) to generate a dense point cloud.
6. Triangulate the point cloud to create a surface mesh.
7. Smooth and down-sample the mesh.
8. Project instances of the part mesh model into the pile mesh at the estimated locations (generated in step 1).

Figure 3.7: Overview of the procedure for generating a mesh model of the pile surface. See Figure 3.9 for an illustration of this procedure.

Step 4 of Procedure 2 uses the Iterative Averaging Algorithm (see Figure 3.8) to fill in invalid/unknown regions of the disparity map. For the purpose of formally presenting this algorithm, we present the following parameter and variable definitions:

$D$  - initial disparity map

$D'$  - final disparity map with invalid/unknown regions filled in (has same dimensions as  $D$ )

$h_D$  - height of the disparity map, in pixels

$w_D$  - width of the disparity map, in pixels

$D(m, n)$  - disparity value at row  $m$  and column  $n$  in the map,  $D$

$B_{init}$  - user-defined disparity initialization parameter

$\Delta x$  - user-defined step-size parameter ( $0 < \Delta x < 1$ )

$DIFF_{max}$  - maximum difference between a disparity value  $D(m, n)$  and the average disparity of the four nearest neighbours of pixel  $(m, n)$

$DIFF_{init}$  - user-defined initialization parameter for  $DIFF_{max}$

$THRES_{stop}$  - user-defined threshold parameter that determines when the algorithm stops ( $THRES_{stop} > 0$ )

The Iterative Averaging Algorithm described in Figure 3.8 has the benefit of providing a conservative estimate of the clear space within the pile.

In the final step of Procedure 2, we use a-priori knowledge of the part to improve the pile surface model. The number of projected instances is directly proportional to the number of candidates we can localize within the pile; therefore, the quality of the pile model is dependent on how well we can recognize and localize parts within it. Figure 3.9 illustrates the process described in Procedure 2.

**Algorithm** ITERATIVE\_AVERAGING

Let  $isBorderPixel(D, m, n)$  represent a function that returns TRUE if the pixel indexed by  $(m, n)$  is along the border of  $D$ , and FALSE otherwise

Let  $isInvalidPixel(D, m, n)$  represent a function that returns TRUE if the pixel indexed by  $(m, n)$  has an invalid/unknown disparity value, and FALSE otherwise

Let  $neighbourAverage(D(m, n))$  represent a function that returns the average disparity of the nearest four neighbouring pixels to the pixel specified by  $(m, n)$  (must be an interior pixel)

```

for  $m = 1$  to  $h_D$ 
  for  $n = 1$  to  $w_D$ 
    if ( $isBorderPixel(D, m, n)$  OR  $isInvalidPixel(D, m, n)$ )
       $D(m, n) = B_{init}$ 
    end if
  end for
end for

```

```

 $D' = D$ 
 $DIFF_{max} = DIFF_{init}$ 

```

```

while( $DIFF_{max} > THRESSTOP$ )
  for  $m = 1$  to  $h_D$ 
    for  $n = 1$  to  $w_D$ 
      if ( $isInvalidPixel(D, m, n)$ )
         $diff = (neighbourAverage(D(m, n)) - D(m, n))$ 
        if( $DIFF_{max} < |diff|$ )
           $DIFF_{max} = diff$ 
        end if
         $D'(m, n) = D(m, n) + \Delta x \cdot diff$ 
      end if
    end for
  end for
   $D = D'$ 
end while

```

Figure 3.8: Iterative Averaging Algorithm used to fill in the invalid/unknown regions of the disparity map.

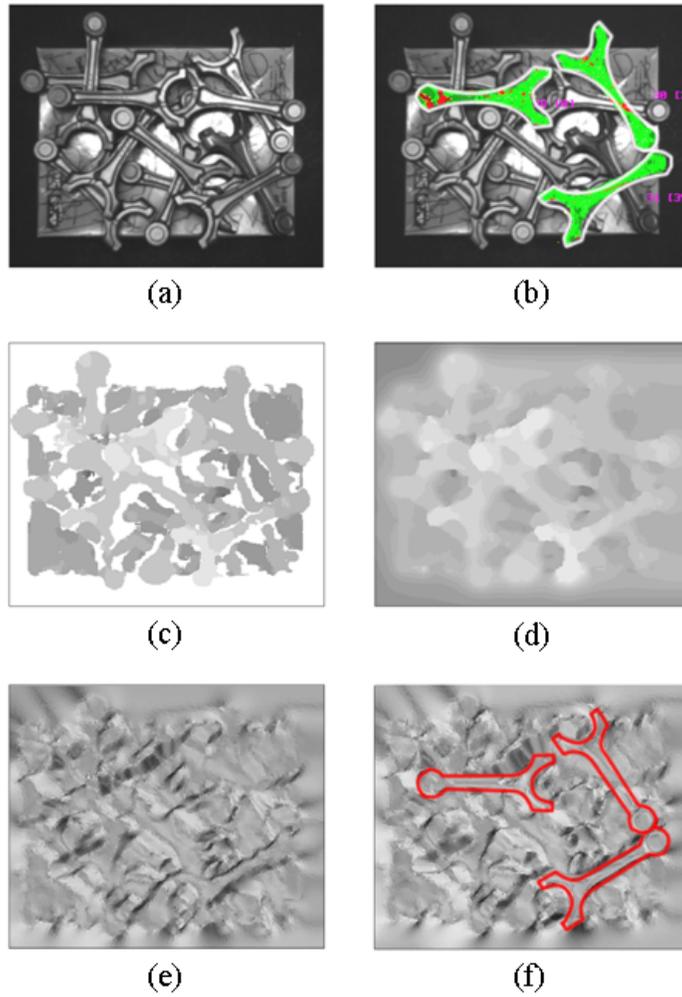


Figure 3.9: Illustration of Procedure 2 (see Figure 3.7) for generating a pile surface model from stereo data. (a) A top-view of a random pile of con-rods. (b) A screenshot from the vision software, eVisionFactory [6], used to localize candidates within the pile. Three candidates have been localized and highlighted. (c) Disparity map of the pile surface. White pixels represent regions where depth information is invalid/unknown. (d) Filled-in disparity map using Iterative Averaging Algorithm. (e) Resulting mesh after converting to  $(x, y, z)$  coordinates, triangulating, smoothing, and down-sampling. (f) Final model of the pile surface after projecting instances of con-rod model into the scene at the locations provided by eVisionFactory [6].

### 3.2.2 Computing Clear Picks

In our previous work (see [9]), the focus was on determining the best candidate *part* to pick up from the pile (as opposed to the best *pick*), and it was selected to be the part with the highest number of clear grasps. Once the best candidate was selected using this approach, we chose the best grasp as the highest-rated clear grasp *for that target part*. However, this approach is not ideal for the following reasons: (1) there is no explicit measure of the clearance around the gripper fingers (this is not computed), and (2) it may not always yield the highest-quality grasp possible for that pile.

These concerns led us to formulate the notion of a pick, which is collectively a grasp and a candidate part. Thus, we shifted our approach from “determining the best part to pick up” to “determining the best pick”. In the latter, we check each potential pick for validity by checking for collisions between the mesh model of the gripper at each pick configuration, and the pile mesh model. We then select the best pick as the one with the highest quality measure that is also collision-free, within a predetermined clearance buffer. Possible pick configurations are generated by applying a set of high-quality grasps (generated offline) to each localized candidate. The procedure for computing clear picks is outlined in Figure 3.10.

In a system where error can originate from the pose estimation of candidate parts, the incomplete and noisy stereo data used to form the pile mesh, and the robot position itself, we impose a buffer of clearance around the gripper fingers to reduce the likelihood of a collision. We decided to implement clearance as a binary filter: that is, if there is a minimum amount of collision-free volume around the gripper fingers, the pick is allowable, or “clear”, and eliminated otherwise. We reasoned that clearance beyond a predetermined minimum amount should not factor into the quality of a pick, since it is assumed that the VGRBP system is sufficiently accurate to operate within the clearance bounds (provided that the clearance buffer size is chosen based on the accuracy of the system). On the other hand, interference of any size can cause the pick to fail.

To generate a clearance buffer, we enlarge the gripper fingers by scaling the finger dimensions (in the  $x$ ,  $y$ , and  $z$  directions; see Figure 3.11). The dimensions of the gripper fingers are multiplied by scale factors to generate the dimensions of the enlarged gripper. When we check for collisions between the pile mesh and the gripper, we employ the enlarged gripper, resulting in collision-free picks with some predetermined amount of clearance (as dictated by the amount of scaling).

When checking for collisions, it is likely unnecessary to use the full list of

**Procedure 3 - COMPUTING CLEAR PICKS**

1. Obtain inputs: (i) ranked set of highest-quality robust grasps,  $\{G\}$  (generated offline), (ii) pose estimates for localized parts, (iii) gripper model, and (iv) pile surface model (relative to camera coordinate frame).
2. For each localized part:
  - (a) Using part's pose hypothesis, obtain homogeneous transformation that describes part's pose in camera coordinate frame,  ${}^cH_p$ .
  - (b) For each potential grasp,  $g_i$ , in grasp set  $\{G\}$ :
    - i. Compute transformation that describes gripper's pose in camera frame,  ${}^cH_g$ , using the following equation:  ${}^cH_g = {}^cH_p \cdot {}^pH_g$ , where  ${}^pH_g$  describes gripper's pose in the part frame.
    - ii. Apply  ${}^cH_g$  to gripper model to place gripper at current pick position.
    - iii. Check for collisions between gripper and pile surface model. If collision detected, eliminate pick.
3. Rate remaining picks based on their evaluated grasp quality measure,  $Q$ .

Figure 3.10: Overview of the procedure for computing clear picks.

robust grasps,  $\{G_{FULL}\}$ , generated offline. In addition, although all grasps in this list are technically “stable” (i.e., force-closed), some may be relatively poor in practice. Lastly, when evaluating picks online, computation time should be reduced where possible, in order to meet the cycle time requirement. For these reasons, we further reduce this list to a set of the highest-quality grasps,  $\{G\}$ , for collision-checking.

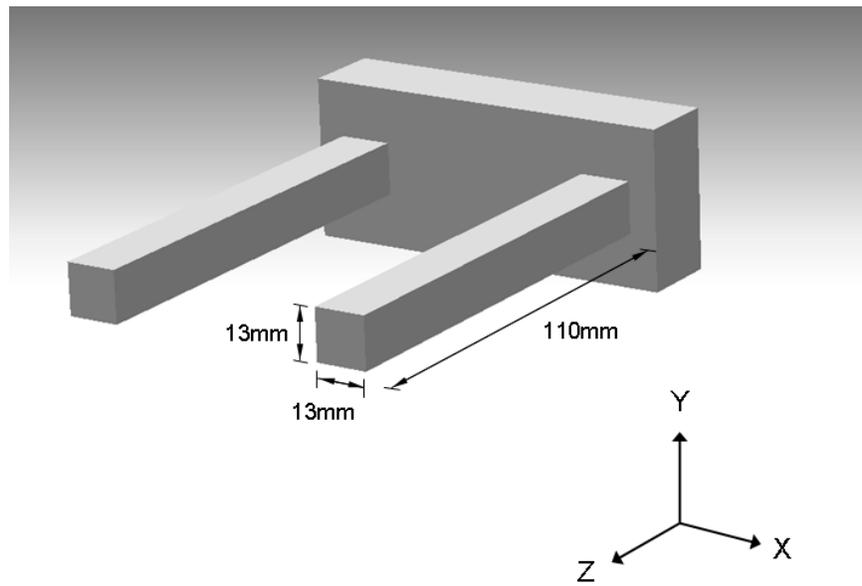


Figure 3.11: Illustration of dimensions and scaling directions of gripper, along which the gripper fingers are enlarged to generate clearance.

### 3.3 Summary

In this chapter, we described a novel method for determining the best pick on a given VGRBP cycle. This method involves (1) an offline portion, in which we densely sample the two-fingered grasp space of a choice part and evaluate the quality of all sampled grasps, and (2) an online portion, in which we evaluate a subset of the highest-quality grasps in the context of the pile to determine a collision-free set of rated picks. For (1), we presented an automatic grasp generation algorithm for sampling the grasp space, as well as a metric function for evaluating grasp quality. For (2), we described the process of generating a pile surface mesh model needed for collision-checking, and how we compute clear picks by considering gripper-finger clearance.

In the next chapter, we show the results from applying this method using a part that is commonly binned in automotive assembly lines.

## Chapter 4

# Experiments and Results

To evaluate our proposed pick selection method, we investigated bin-picking of a connecting rod, or con-rod (from a car engine). This part is typical in size and form of parts that would be suitable for bin-picking applications. Other parts to consider include screws, shafts, and caps, as they are simple in shape and typically delivered to the assembly line jumbled in bins. Future work aims to apply the proposed method to other types of parts, covering a variety of geometries, in order to strengthen and expand on the results presented herein.

Section 4.1 shows the results of generating a densely-sampled ranked set of robust grasps. This grasp set is used in Section 4.2, in which the proposed method for determining the best pick is experimentally validated using stereo data of real con-rod piles. In each section, a discussion accompanies the results shown.

### 4.1 Creating a Densely-Sampled Set of Evaluated Grasps

The parameters used for grasp generation for a con-rod are summarized in Table 4.1. Since typical values for robot accuracy and pose estimation accuracy are  $\pm 1\text{mm}$  and  $\pm 2\text{mm}$ , respectively, we estimated the accuracy of our system to be  $\pm 3\text{mm}$ , and thus, used a sampling step-size of  $3\text{mm}$ . To determine the 3-D region,  $\rho$ , for robustness calculations, we limited the neighbourhood around each grasp to include grasps within one step-size in all directions ( $\theta$ ,  $\phi$ , and  $d$ ). This can be visualized as a  $3 \times 3 \times 3$  array of grasp samples. We selected the tunable stability parameter  $\gamma = 2$ .

Table 4.1 summarizes the results of the grasp generation using the parameters shown in Table 4.1. Out of 26650 grasps sampled, 5011, or 19% were robust.

Figure 4.1 visualizes these grasps with respect to the con-rod model from different viewing directions of the model. We have chosen to visually represent each grasp using a  $T$  shape, which is to be interpreted as follows:

Table 4.1: Summary of input parameters used for grasp generation and evaluation.

Grasp generation sampling size (mm)	Soft gripper finger tolerance, $\varepsilon$ (mm)	Dimensions of region $\rho$ for robustness calculation	$\gamma$
3	3	$3 \times 3 \times 3$	2

Table 4.2: Grasp generation results. Percentages are in relation to number of grasp samples.

# of grasp samples	Feasible grasps		Stable and feasible grasps		Robust grasps	
	#	%	#	%	#	%
26650	20501	76.9	17367	65.2	5011	18.8

- The location of the  $T$  along the wire-frame represents the position of the grasp (as described by  $d$ ).
- The stem of the  $T$  represents the approach direction of the gripper.
- The top bar of the  $T$  represents the pinching direction of the gripper.
- The size of the  $T$  represents the quality,  $Q$ , of the grasp; it has been uniformly scaled according to  $Q$ .

Only robust grasps are shown. In Figure 4.1 (a), the part model is overlaid onto the wire-frame. In (b), just the wire-frame is shown.

The grasp qualities depicted in Figure 4.1 are consistent with what one would expect: high-quality grasps tend to be those for which (a) there exist many points of contact between the gripper fingers and the part, and (b) forces applied at the gripper finger contacts are generally normal to the part’s surface. As expected, the best grasps are clustered near the centre of mass of the part, where disturbance torque is minimized, and few good grasps are found in regions of high surface curvature. The grasps are not perfectly symmetrical because the con-rod model used was obtained from a

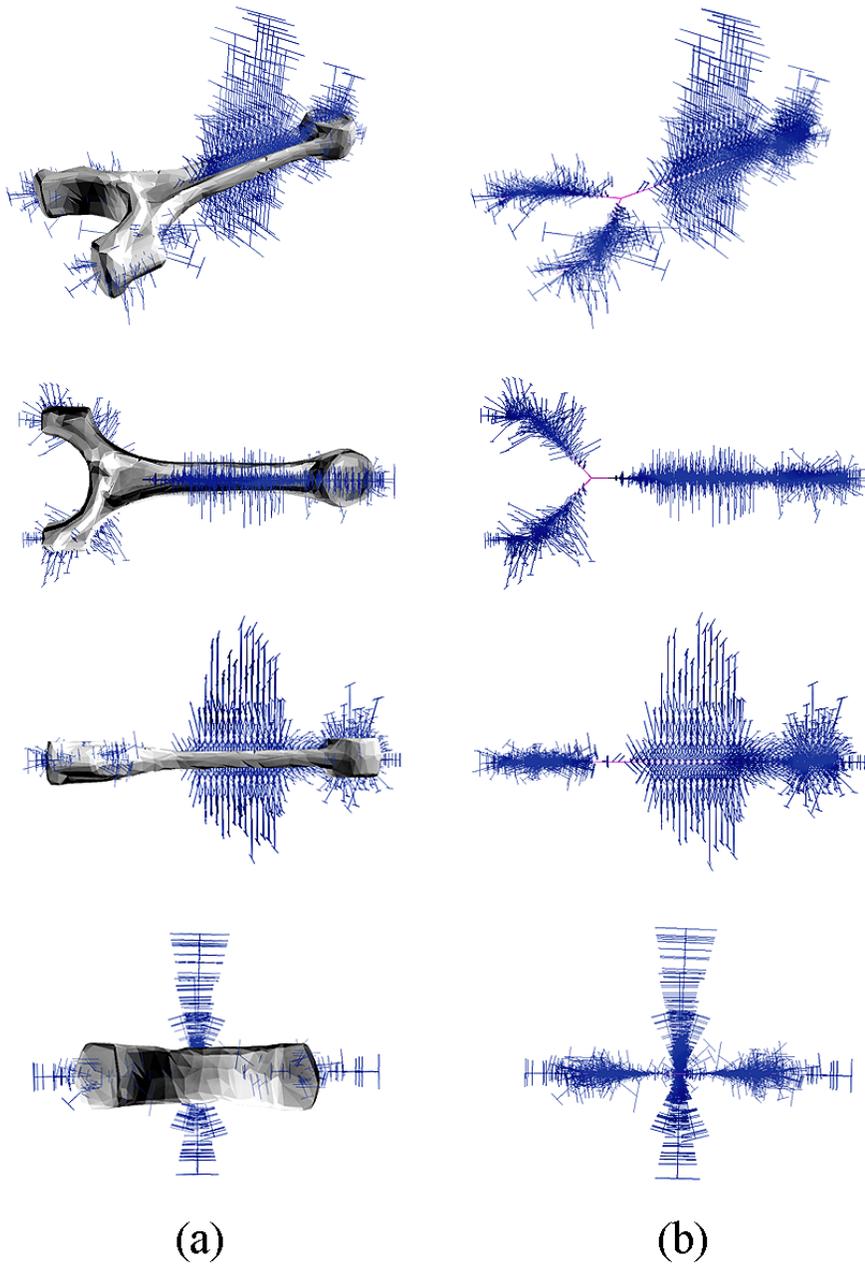


Figure 4.1: Visualization from different viewing directions of uniformly-spaced, densely-sampled list of generated grasps with respect to wire-frame,  $S_p$ . See text in Section 4.1 for details.

laser scan of a physical con-rod. Since grasp stability,  $q$ , is dependent on the number of contacts between the gripper fingers and the part, the final quality,  $Q$ , is sensitive to the deformability of the gripper fingertips (modeled in our system as  $\varepsilon$ ) and imperfections in the surface model of the part.

Grasp quality is also sensitive to the resolution of the part model. Recall from Section 3.1.1 that grasp points are defined at the extrema of the cross-section along the pinching direction, within  $\varepsilon$  (see Figure 3.5). If the part model is high-resolution (i.e., the mesh is formed from a large number of triangles), the cross-section will also be high-resolution; that is, the cross-section will comprise the same number of line segments or more than with a low-resolution model. In this implementation, grasp points are selected as the end points of each line segment comprising the cross-section, within  $\varepsilon$ . Therefore, the number of grasp points and corresponding stability measure calculated by *GraspIt!* are dependent on the part model resolution. This dependency may lead to different grasp rankings if computed at differing levels of resolution, and motivates the use of a high-resolution part model in order to achieve a high level of accuracy. The cost of increasing the part model resolution is computation time, but this is a minor concern since grasp generation is performed offline.

Another implication of selecting grasp points as the end points of each line segments is that the number of grasp points generated per line segment does not depend on the length of the segment. A potential consequence of this is selecting grasp points that poorly approximate the actual contacting surfaces of the grasp. This is one weakness of using a discrete set of point contacts to represent continuous surface area contacts. A possible improvement to this implementation is using the total length of the set of line segments that lie within  $\varepsilon$  to determine the number of grasp contact points.

In addition to using the evaluated densely-sampled set of generated grasps, illustrated in Figure 4.1 to provide many picks online, we can also use this data to establish high-quality grasp regions, aiding in the selection of nominal grasps offline.

## 4.2 Determining the Best Pick: Validating the Proposed Method

In our previous work (see [9]), we generated a densely-sampled set of evaluated grasps offline, and then used a subset of the highest-quality grasps,  $\{G\}$ , to select a pick online. However, in that paper, we focused on determining the best candidate part to pick up, and selected it to be the part

with the highest number of clear grasps. We then selected the target grasp to be the highest-quality, collision-free grasp for the chosen target. Thus, the main differences between earlier work and the proposed method herein concern the online computation portion, and are summarized by: (1) the notion of a pick (which enabled searching for the best pick as opposed to the best target part), and (2) computation of gripper finger clearance.

We validated our previous approach in simulation. Herein, we include a brief summary of the earlier method and the simulation results in Section 4.2.1. We include these results to reinforce the effectiveness of using a large grasp set to generate more picking options, and ultimately determine the best one. Section 4.2.2 explains our experiment and results for validating our current approach using stereo data from a real pile of parts. In both cases, we compare two grasp sets: (1) the set of top quality grasps from a generated list of evaluated grasps,  $\{G\}$ , and (2) a set of 12 nominal intuitive grasps,  $\{N\}$ , that would typically be used in an industrial application. The nominal grasps are illustrated in Figure 4.2. It should be noted that, due to the symmetry of the con-rod and the gripper, there exist two grasps (with diametrically opposed approach directions) that result in the same set of contact points on the surface of the object; thus, they share the same  $Q$  value. For both methods (earlier and current), we statistically validate the hypothesis that using a relatively large grasp set results in more picking options, on average.

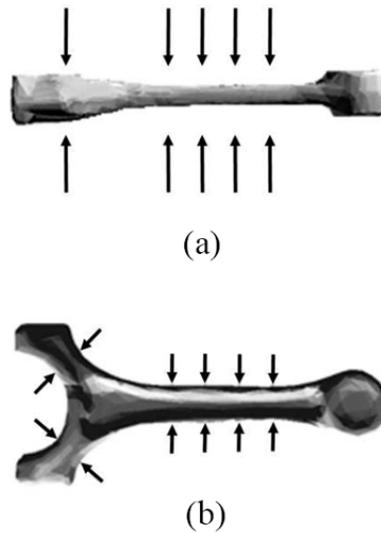


Figure 4.2: Visual description of  $\{N\}$ , comprising 12 nominal grasps used for experiments. (a) Side-view of part; arrows represent approach directions. (b) Top-view of part; arrows represent pinching directions. For each pinching direction shown in (b), note that there are two grasps, with diametrically opposed approach directions, that result in the same set of contact points on the surface of the object.

### 4.2.1 Determining the Best Pick in Simulation: Earlier Work

As previously stated, in earlier work (see [9]), we focused on selecting the best part to pick up out of a set of candidate parts, and it was chosen to be that with the highest number of clear grasps from a set of highest-quality grasps,  $\{G\}$ . Thus, we checked for collisions between the gripper and the pile for each grasp in  $\{G\}$ , for each candidate. We considered a candidate to be valid if there existed at least one collision-free grasp to retrieve it with, and rated valid candidates based on the number of collision-free grasps for each.

We evaluated candidates within multiple, simulated, randomized piles of con-rods, using two grasp sets,  $\{G\}$  and  $\{N\}$ , and compared the statistical results. Using simulated piles allowed us to have complete knowledge (i.e., pose information) of all obstacles. To create each pile, we randomly stacked parts, one at a time, using a physics simulator to model the rigid-body dynamics. To form  $\{G\}$ , we ranked all grasps from  $\{G_{FULL}\}$  based on  $Q$ , and selected the top 10% of grasps. It should be noted that all grasps from  $\{G_{FULL}\}$  could, potentially, have been included since all are robust. For each potential grasp, we checked for collisions with the ground plane and all other parts in the pile using the efficient hierarchical Oriented Bounding Box method described in [15].

We performed this evaluation with 100 different piles of 25 con-rods; for each pile, we selected the last 15 parts that had been added to the pile as our candidates in order to approximate the real-world situation wherein the candidates would likely be at or near the surface of the pile. Table 4.3 summarizes the input parameters for the experiment. Figure 4.3(a) shows an example of a simulated pile of parts, with the candidates highlighted in Figure 4.3(b). Valid candidates are highlighted and numbered according to their rating in Figure 4.3(c)-(d), for the grasp sets  $\{G\}$  and  $\{N\}$ , respectively.

The average number of valid candidates for the set of top grasps,  $\{G\}$ , and the set of nominal grasps,  $\{N\}$ , were 8 and 5, respectively. A paired t-test analysis of the null hypothesis that these two methods produce the same distribution of valid parts for picking had a probability of  $7.55 \times 10^{-25}$ , indicating that the distributions are significantly different. These results are summarized in Table 4.4. They confirm the hypothesis that increasing the number of possible grasps for the part results in an increased number of valid candidates and, accordingly, the number of picking options. However, this evaluation does not consider whether or not candidates are pinned down

Table 4.3: Summary of input parameters used for simulated experiment.

Input Parameters				
# of piles	# of parts per pile	# of candidates per pile	# Percentage of top robust grasps used	Resulting # of grasps used
100	25	15	10%	428

Table 4.4: Summary of results from simulated experiment.

Average # of parts with at least one valid grasp		Probability that distributions are the same (paired t-test)
Generated grasps $\{G\}$	Nominal grasps $\{N\}$	
8	5	$7.55 \times 10^{-25}$

by other parts, and if so, the extent to which they are buried. One would expect that a candidate for which there is an available grasp in the context of the pile, but is deeply embedded in the pile, would be a poor option, and should be eliminated. An example of this situation is illustrated in Figure 4.3(c) for the candidates rated 5th and 7th. It should be noted that we assumed that the probability distribution for the number of picks per pile could be approximated using a normal distribution.

In [9], we had reasoned that there would be a correlation between the number of grasps available for a particular candidate part and the amount of clearance around the candidate. As such, this approach did not provide an explicit computation of clearance - information that is essential for planning clear picks in practice. This motivated the addition of clearance computation in the proposed pick evaluation method herein. The results of our experiments with clearance included are described in the next section.

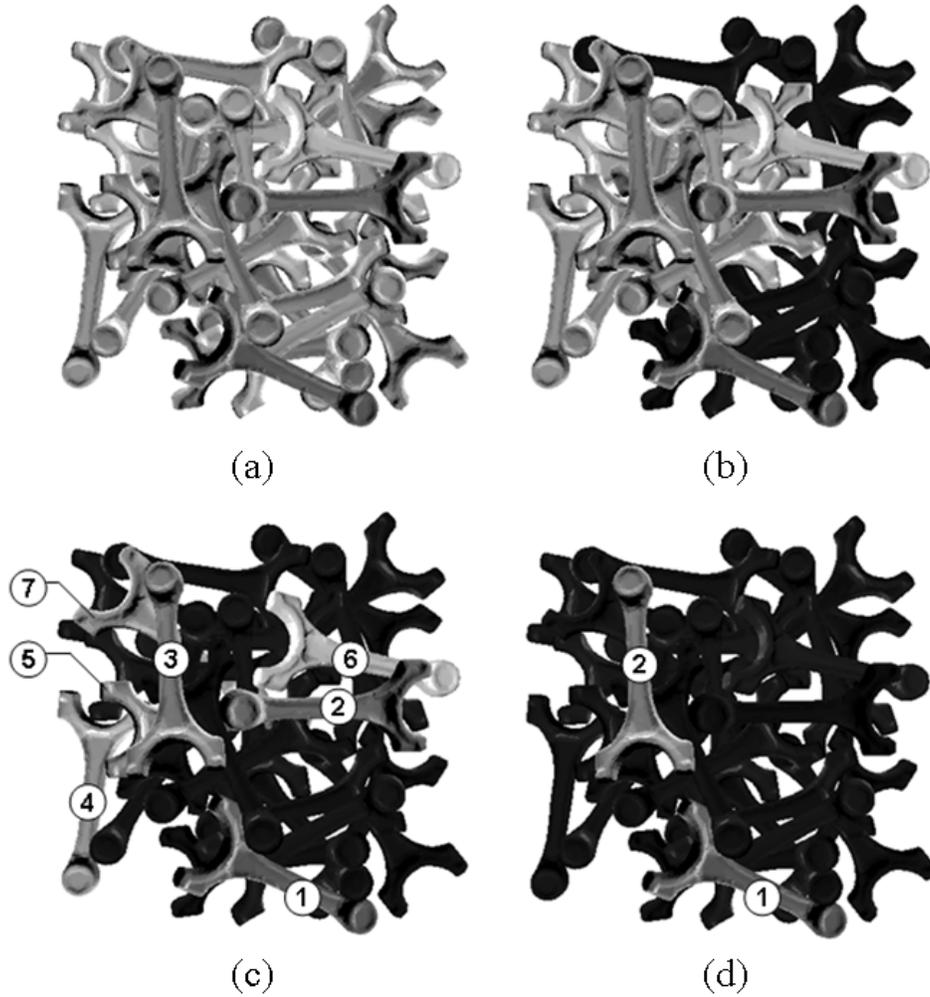


Figure 4.3: Comparison of valid candidates determined for an example pile using grasp sets  $\{G\}$  and  $\{N\}$ . (a) The simulated pile of parts. (b) Highlighted candidates. (c) Highlighted valid candidates found using  $\{G\}$ , numbered according to their rating. (d) Highlighted valid candidates found using  $\{N\}$ , numbered according to their rating.

### 4.2.2 Determining the Best Pick: Using Stereo Data of a Real Pile

In this experiment, we validate the online portion of the proposed method by executing it using stereo data from a real pile of parts. We also examine the effects of varying the amount of required clearance on the resulting number of generated picks. The tools we used for the experiment included a Point Grey Research Bumblebee2 stereo camera [29] mounted directly above a pile of con-rods (see Figure 4.4), and the Point Grey Triclops software [29] to read in stereo images and create corresponding disparity maps. For mesh manipulation, including smoothing, coarsening, and collision detection, we used the open-source GNU Triangulated Surface (GTS) library [22]. To localize candidate con-rods to pick within the pile, we used computer vision software called eVisionFactory, or eVF [6], which provides pose hypotheses of recognized con-rods using image data from the Bumblebee2. An eVF snapshot with a set of localized candidate parts highlighted is shown in Figure 3.9(b).

In our experiment, we computed and rated a set of the best available picks for a real pile of con-rods for each of two sets of grasps: (1) a set of highest quality grasps from our generated grasp list,  $\{G\}$ , and (2) a set of 12 nominal “intuitive” grasps,  $\{N\}$ , that would typically be used in an industrial application (shown in Figure 4.2). For the first set, grasps were ranked based on  $Q$ , and the top  $\eta = 120$  were selected, although all grasps could, potentially, have been included since all are robust. This quantity,  $\eta$ , is a tunable parameter, and optimizing this value depends on the quality of grasps generated, as well as limits on online computation time. For each potential pick, we checked for collisions between the gripper and the pile model, derived from stereo data. We repeated this for 30 random piles (30 trials), and 3 levels of clearance imposed on the gripper fingers. Table 4.5 summarizes the input parameters for the experiment, while table 4.6 summarizes the clearance parameters.

The data from this experiment are shown in Table 4.7. Statistical analysis of these data is summarized in Table 4.8 and Figure 4.4. We calculated the mean and standard deviation,  $\sigma$ , for each clearance level, and defined the error on the mean as  $\pm\sigma$ . In earlier work, we assumed a normal distribution in our analysis (see Section 4.2.1); however, applying a normal distribution here to model the probability density of these data results in a significant portion of the distribution extending into the range of negative numbers. Since the number of picks must be a non-negative quantity, we modeled the data with a gamma-type probability density function that is non-negative

Table 4.5: Summary of input parameters used for pick evaluation experiment.

Input Parameters				
# of piles	# of parts per pile	# of candidates per pile	# of top robust grasps used, $\eta$	# of clearance levels used
30	13	3	120	3

Table 4.6: Summary of scaling parameters used at each clearance level. The scaling directions for enlarging the gripper are shown in Figure 3.11. The dimensions of the gripper fingers are multiplied by the scale factor to generate the dimensions of the enlarged gripper. Note that the scaling parameters for Clearance Level 1 are all unity, resulting in zero clearance buffer.

Clearance Level	Scale Factor (ratio of gripper finger dimensions)		
	$x$ direction	$y$ direction	$z$ direction
1	1.00	1.00	1.00
2	1.20	1.20	1.035
3	1.40	1.40	1.07

by definition [23]. The key parameter that describes the shape of this distribution, and therefore defines  $\sigma$ , is  $\alpha$  (see Appendix A for a description of the gamma distribution). We chose  $\alpha = 2$  to provide the best fit to the histogram data (shown in Appendix B).

Table 4.7: Results of evaluating picks for the two grasp sets,  $\{N\}$  (nominal grasps) and  $\{G\}$  (grasps generated by our system), on multiple randomized piles of con-rods for three levels of gripper finger clearance (see Table 4.6 for clearance parameters). When observing these values, note that, the highest  $Q$  from the entire set of  $\{G\}$  is 0.079. Also note that, some of the grasps from  $\{N\}$  are unstable ( $Q = -1$ , represented by “x” in the table), as they were not selected based on quality, whereas the grasps from  $\{G\}$  must be stable by definition.

Pile #	Clearance Level 1				Clearance Level 2				Clearance Level 3			
	$\{N\}$		$\{G\}$		$\{N\}$		$\{G\}$		$\{N\}$		$\{G\}$	
	# of clear picks	$Q$ of best pick	# of clear picks	$Q$ of best pick	# of clear picks	$Q$ of best pick	# of clear picks	$Q$ of best pick	# of clear picks	$Q$ of best pick	# of clear picks	$Q$ of best pick
1	3	x	7	0.06	2	x	0	n/a	0	n/a	0	n/a
2	4	0.049	50	0.079	2	0.049	14	0.049	0	n/a	1	0.03
3	7	0.049	99	0.079	6	0.049	81	0.079	2	x	20	0.07
4	7	0.049	50	0.079	1	x	8	0.044	1	x	2	0.032
5	6	0.049	70	0.079	5	0.049	50	0.079	2	0.044	29	0.07
6	6	0.049	75	0.079	4	0.044	44	0.079	3	0.044	25	0.076
7	3	x	4	0.035	2	x	0	n/a	1	x	0	n/a
8	2	0.044	52	0.079	2	0.044	43	0.079	1	x	23	0.06
9	5	0.007	39	0.072	3	0.007	20	0.072	0	n/a	3	0.035
10	6	0.049	88	0.079	4	0.049	80	0.079	1	x	31	0.079
11	13	0.049	135	0.079	9	0.049	84	0.079	4	0.007	27	0.061
12	4	0.044	20	0.072	1	0.007	8	0.04	1	0.007	2	0.031
13	6	0.007	27	0.06	5	x	5	0.035	2	x	0	n/a
14	6	0.049	56	0.079	4	0.049	20	0.049	0	n/a	1	0.03
15	3	0.044	41	0.079	2	0.044	18	0.076	2	0.044	11	0.07
16	7	0.044	32	0.076	4	x	4	0.035	1	x	0	n/a
17	3	x	1	0.036	1	x	0	n/a	1	x	0	n/a
18	1	x	6	0.049	1	x	1	0.037	0	n/a	0	n/a
19	12	0.049	102	0.079	5	0.044	74	0.079	3	0.044	29	0.076
20	6	0.007	25	0.049	2	x	5	0.034	1	x	1	0.051
21	6	0.049	44	0.072	3	x	10	0.044	0	n/a	9	0.044
22	3	x	11	0.044	2	x	11	0.044	1	x	7	0.044
23	6	0.049	24	0.072	3	0.007	8	0.049	1	0.007	1	0.036
24	9	0.049	120	0.079	5	0.049	65	0.079	1	x	2	0.035
25	6	0.049	54	0.072	3	0.007	10	0.039	1	x	2	0.03
26	10	0.049	85	0.079	5	0.007	15	0.056	2	x	0	n/a
27	12	0.049	115	0.079	5	0.049	53	0.079	2	0.007	8	0.068
28	10	0.049	117	0.079	6	0.007	62	0.079	2	x	9	0.06
29	8	0.049	106	0.079	5	0.044	75	0.079	2	0.044	52	0.076
30	3	0.007	28	0.076	2	0.007	5	0.06	0	n/a	0	n/a

Table 4.8: Summary of statistical analysis of experimental data. Experiment used stereo data of 30 real piles of parts for three levels of gripper finger clearance.

Clearance level	Average # of clear picks		Standard dev., $\sigma$		% of trials where more picks found using $\{G\}$ than using $\{N\}$
	$\{N\}$	$\{G\}$	$\{N\}$	$\{G\}$	
1	6.1	56.1	4.3	39.7	97%
2	3.4	28.0	2.4	19.8	90%
3	1.3	9.8	0.9	7.0	83%

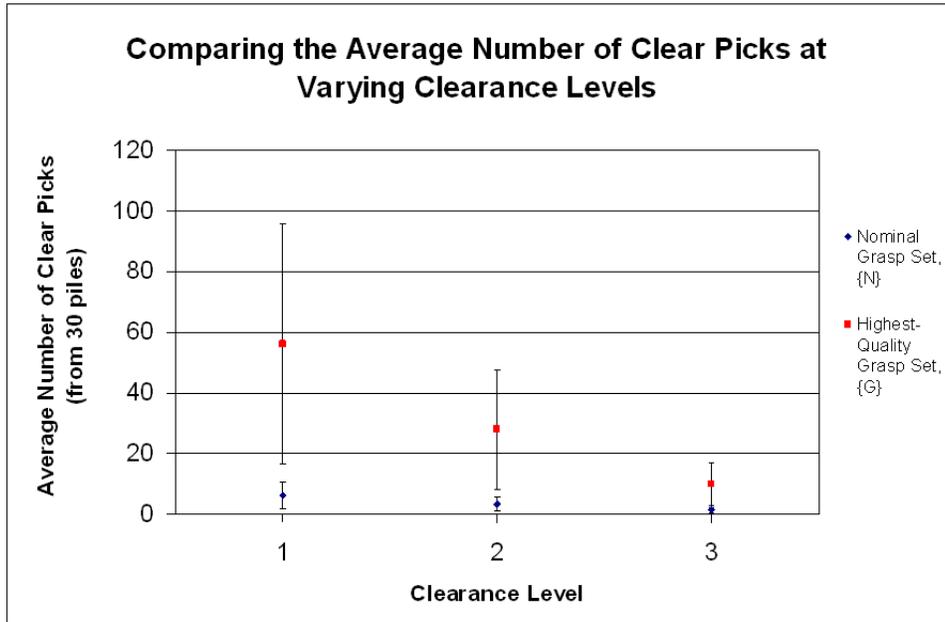


Figure 4.4: Comparison of the number of clear picks between  $\{N\}$  and  $\{G\}$  for three levels of gripper finger clearance. Error bars shown represent  $\pm\sigma$ , based on a gamma distribution with  $\alpha = 2$ . The parameters used to define the clearance levels are summarized in Table 4.6. The data used to construct this figure are summarized in Table 4.7.

From our results, we observe the following:

1. Using  $\{G\}$  generated significantly more picks, on average, than using  $\{N\}$ , for all three clearance levels.
2. In the majority of trials, more picks were found using  $\{G\}$  than using  $\{N\}$ , for all three clearance levels. Specifically, for clearance levels 1, 2, and 3, more picks were found using  $\{G\}$  97%, 90%, and 83% of the time, respectively.
3. In all trials where both grasp sets found at least one pick, the quality of the best pick found using  $\{G\}$  was higher than that using  $\{N\}$ .
4. In *all* trials where using  $\{N\}$  generated more picks, the best pick found using  $\{N\}$  was unstable, according to our evaluation (recall that the grasps comprising  $\{N\}$  were chosen as nominal, “intuitive” grasps, and not chosen based on their quality).

These results confirm the hypothesis that increasing the number of possible grasps for the part results in an increased number of clear picks, on average. Additionally, we have confidence that the resulting clear picks are high-quality since they have already been evaluated offline, and likely more stable and robust than those generated using a set of nominal grasps. Finally, for all piles and both grasp sets, the number of clear picks decreases as the minimum required clearance increases, as expected. This trend highlights the advantage of using a larger grasp set when increasing the minimum required clearance, as it increases the likelihood of finding a clear pick.

As with our earlier method (see Section 4.2.1), the proposed evaluation does not consider whether or not candidates are pinned down by other parts, and if so, the extent to which they are buried. Also, the current implementation has not yet been fully optimized for speed; the online computation time varies widely, but is on the order of minutes, with the bottleneck occurring during the mesh processing and collision-checking steps. It is currently a proof-of-concept at this stage, requiring computational speed-ups to be practical. Another issue is the variability inherent in randomized piles. This presents a challenge when attempting to statistically analyze the data since it is very difficult to accurately model the underlying distribution that drives the data. Future work aims to address these issues.

Figure 4.5 compares the best pick found using each grasps set, for pile 10 at the highest clearance level. A snapshot of the pile and the localized candidates using eVF [6] are shown in Figure 4.5 (a)-(b), respectively. Note

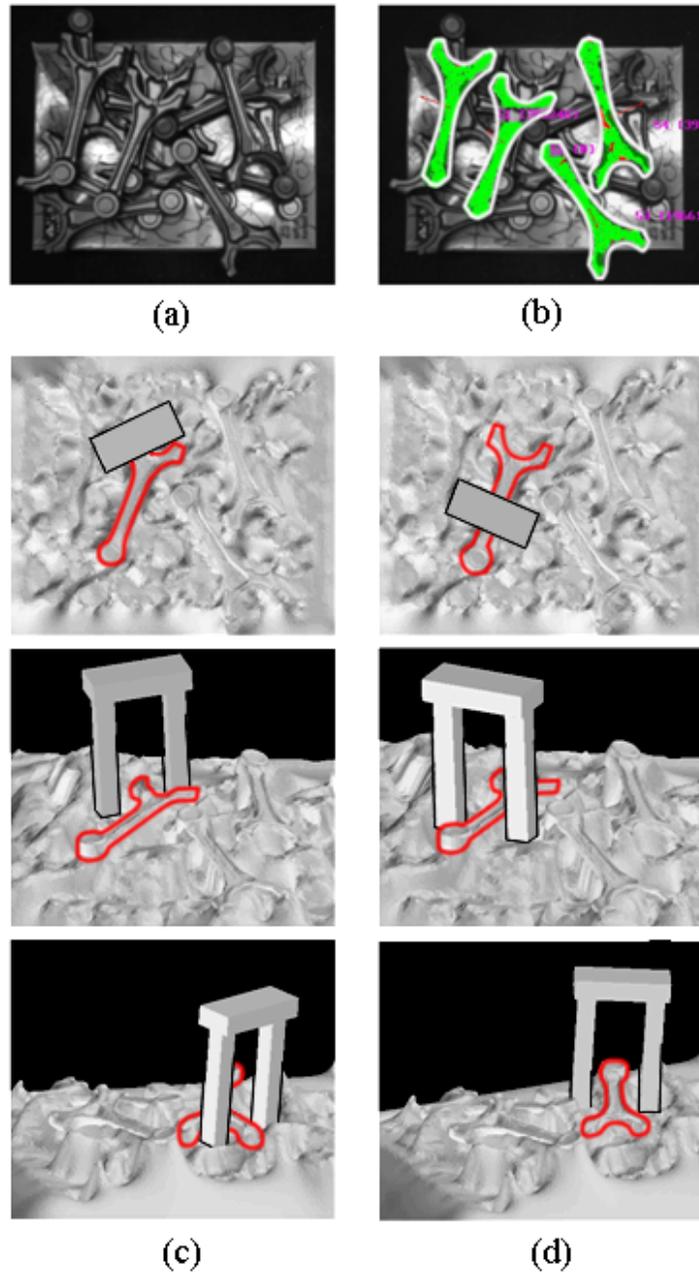


Figure 4.5: Comparison of the best picks generated from the two grasp sets,  $\{N\}$  and  $\{G\}$ , for pile 10 at clearance level 3. See text in Section 4.2.2 for details.

that, for the experiment, the number of candidate parts was limited to 3 (so only 3 of the 4 highlighted in Figure 4.5 (b) were used). Figure 4.5 (c)-(d) recreates the best pick for  $\{N\}$  and  $\{G\}$ , respectively, from three different points of view, with the target con-rod outlined. The gripper shown in Figure 4.5 (c)-(d) is the enlarged version of the gripper. In this situation, only one pick is clear from  $\{N\}$ , whereas 31 of the grasps from  $\{G\}$  are clear. Upon inspection, it is evident that the best pick provided by  $\{G\}$  results in the gripper positioned for a better grasp when compared to the best pick from  $\{N\}$ : in (c), the gripper is positioned to grasp the part on the (slightly curved) end of the con-rod, whereas, in (d), the gripper is positioned to pick much closer to the con-rod’s centre-of-gravity, and where the con-rod’s surface is flatter. This observation is validated by the computed quality measures of each grasp: “unstable” ( $Q = -1$ ) for set  $\{N\}$  and 0.079 for set  $\{G\}$  (which is the highest quality measure from  $\{G\}$ ).

### 4.3 Summary

In this chapter, we validated our proposed pick selection method using a con-rod (from a car engine) as our exemplar part. In Section 4.1, we present the results of generating a densely-sampled ranked set of robust grasps, offline. The highest-quality grasps generated were clustered around the centre of mass of the part, with the pinching directions roughly perpendicular to the contact surfaces, as expected. We found few robust grasps in regions of high surface curvature.

We used this generated grasp set in Section 4.2, as well as stereo data of real con-rod piles to validate the online portion of the proposed pick selection method. Our experiment consisted of computing and rating a set of the best available picks for a real pile of con-rods at three levels of gripper finger clearance for each of two sets of grasps: (1) a set of highest quality grasps from our generated grasp list,  $\{G\}$ , and (2) a set of 12 nominal “intuitive” grasps,  $\{N\}$ , that would typically be used in an industrial application. Our results confirmed the hypothesis that increasing the number of possible grasps for a part increases number of clear picks, on average. Also, the picks computed using the generated grasp set were higher quality than those generated using the nominal grasp set.

In the next chapter, we summarize the contributions of this thesis, and describe future directions of this work.

## Chapter 5

# Conclusions and Future Work

### 5.1 Conclusions

The focus of this work was determining the best pick in the context of VGRBP. The main contribution is a novel method for generating many high-quality (rated) pick options for a given cycle, enabling selection of the best pick. This main contribution requires development of the following supporting methods: (1) an automatic grasp-generation method to sample the space of all two-fingered grasps for the target part, (2) a metric function for evaluating grasps, and (3) a measure of the robustness of a grasp. Our work was tailored for a two-fingered gripper, as this is commonly used in industry, but is extendable to other gripper types. Throughout this thesis, we used a connecting rod as our exemplar part; however, the proposed method is theoretically generalizable for a wide range of part geometries.

The method developed for the main contribution to generate high-quality pick options requires, as inputs, a surface mesh model and corresponding wire-frame of the object to be picked, as well as a model of the gripper. It comprises an offline portion for generating high-quality grasps, and an online portion for evaluating these grasps in the context of a pile of objects. The method requires a stereo vision system to obtain depth information of the pile surface, as well as a method for recognizing and localizing candidates within the pile. The output of this method is a list of rated pick options for a given pile configuration, allowing the grasping system to choose the best pick.

The automatic grasp generation method, (1), developed herein, also requires a surface mesh model and corresponding wire-frame of the target object, and a two-fingered gripper model. The method is based on sampling and does not require a closed-form analytic description of the geometry of the chosen part; however, it is sensitive to the resolution of the part model used, and how accurately that model and the wire-frame model represent

the chosen part. The density of sampling is specified by the user. The output is a list of feasible grasps for the given object and gripper, which can then be evaluated for quality.

The metric function for evaluating a grasp, (2), requires two quantities: stability and robustness. Stability is computed using Ferrari and Canny’s widely accepted grasp quality metric [11]. Grasp robustness, as proposed in this thesis (3), is a measure of the insensitivity of a grasp to slight positional changes. We define it as the inverse of the standard deviation of grasp stability over a neighbourhood of grasp samples. Therefore, the robustness measure is dependent on the size of this neighbourhood, and on the grasp stability evaluation. The output of the metric function is a numerical value that represents the quality of a grasp.

We experimentally validated our pick selection method using stereo data of a real pile of parts. We compared the use of our proposed method to an approach typical in industry, and observed that our method resulted in significantly more picks, and higher quality picks. These results suggest that using our method would increase reliability within a VGRBP system by reducing the risk of a failed grasp attempt.

## 5.2 Recommendations for Future Work

One issue not considered in this thesis was that of part stacking; namely, whether or not candidate parts are pinned down by other parts, and if so, the extent to which they are buried. Exploring part stacking scenarios is a direction for future work, since a candidate part’s position within a random pile will affect how easily the part may be extracted from the pile. Herein, only one part, a connecting rod, was considered. In the future, the proposed method should be tested using other parts, covering a variety of part geometries, in order to support the generalizability of the method. Another area for exploration is the use of structured lighting to improve the quality of the stereo data, since the computation of clear picks is highly dependent on the stereo data quality. Also, the implementation used was not fully optimized; to be integrated into a commercially viable VGRBP system, computational speed-ups would be necessary, particularly for collision-detection. Finally, running more trials with randomized piles of real parts may strengthen the statistical results presented here.

# Bibliography

- [1] E. Al-Hujazi and A. Sood. Range Image Segmentation with Applications to Robot Bin-Picking Using Vacuum Gripper. *IEEE Transactions on Systems, Man, and Cybernetics*, 20(6):1,313–1,325, 1990.
- [2] M. A. Baumann, D. Dupuis, S. Léonard, E. A. Croft, and J. J. Little. Occlusion-Free Path Planning with a Probabilistic Roadmap. *IEEE/RJS International Conference on Intelligent Robots and Systems*, pages 2151–2156, 2008.
- [3] D. Berenson and S. Srinivasa. Grasp Synthesis in Cluttered Environments for Dexterous Hands. In *Robotics: Science and Systems Workshop - Robot Manipulation: Intelligence in Human Environments*, 2008.
- [4] A. Bicchi and V. Kumar. Robotic Grasping and Contact: A Review. *IEEE International Conference on Robotics and Automation*, pages 348–53, 2000.
- [5] F. Boughorbel, Y. Zhang, S. Kang, U. Chidambaram, B. Abidi, A. Koschan, and M. Abidi. Laser ranging and video imaging for bin picking. *Assembly Automation*, 23(1):53–59, 2003.
- [6] Braintech Inc. <http://www.braintech.com/products-evf.php>.
- [7] H. Bruyninckx, S. Demey, and V. Kumar. Generalized Stability of Compliant Grasps. In *International Conference on Robotics and Automation*, volume 4, pages 2396–2402, 1998.
- [8] A. Chan, E. A. Croft, and J. J. Little. Trajectory Specification via Sparse Waypoints for Eye-In-Hand Robots requiring Continuous Target Visibility. *IEEE International Conference on Robotics and Automation*, pages 3082–3087, 2008.
- [9] D. Dupuis, S. Léonard, M. A. Baumann, E. A. Croft, and J. J. Little. Two-Fingered Grasp Planning for Randomized Bin-Picking. In *Robotics: Science and Systems Workshop - Robot Manipulation: Intelligence in Human Environments*, 2008.

- [10] S. Ekvall and D. Kragic. Learning and Evaluation of the Approach Vector for Automatic Grasp Generation and Planning. In *IEEE International Conference on Robotics and Automation*, pages 4715–4720, 2007.
- [11] C. Ferrari and J. Canny. Planning Optimal Grasps. *IEEE International Conference on Robotics and Automation*, pages 2290–2295, 1992.
- [12] I. Fryndendal and R. Jones. Segmentation of sugar beets using image and graph processing. *International Conference on Pattern Recognition*, 2:1697–1699, 1998.
- [13] O. Ghita and P. F. Whelan. A bin picking system based on depth from defocus. *Machine Vision and Applications Journal*, 13(4):234–244, 2003.
- [14] C. Goldfeder, P. Allen, C. Lackner, and R. Pelossof. Grasp planning via decomposition trees. In *IEEE International Conference on Robotics and Automation*, pages 4679–4684, 2007.
- [15] S. Gottschalk, M.C. Lin, and D. Manocha. Obbtree: A Hierarchical Structure for Rapid Interference Detection. *Computer Graphics (SIG-GRAPH '96 Proceedings)*, 30:171–180, 1996.
- [16] K. Ikeuchi. Generating an interpretation tree from a cad model for 3-d object recognition in bin-picking tasks. *International Journal of Computer Vision*, 1(2):145–165, 1987.
- [17] J. Beis (Braintech Inc.). Personal communication, May 2008.
- [18] W. Iverson. Vision-guided robotics: In search of the holy grail. *Automation World*, page 28, February 2006.
- [19] J. Kirkegaard and T. B. Moeslund. Bin-Picking based on Harmonic Shape Contexts and Graph-Based Matching. *International Conference on Pattern Recognition*, 2:581–584, 2006.
- [20] S. V. Léonard, E. A. Croft, and J. J. Little. Dynamic Visibility Checking for Vision-Based Motion Planning. *IEEE International Conference on Robotics and Automation*, pages 2283–2288, 2008.
- [21] Z. Li, P. Hsu, and S. Sastry. Grasping and coordinated manipulation by a multifingered robot hand. *International Journal of Robotics Research*, 8(4):33–50, 1989.

- [22] GNU Triangulated Surface Library. <http://gts.sourceforge.net/>.
- [23] W. Mendenhall and T. Sincich. *Statistics for Engineering and the Sciences*, page 190. Pearson Prentice Hall, fifth edition, 2007.
- [24] A. Miller and P. Allen. Graspit! A versatile simulator for robotic grasping. *IEEE Robotics and Automation Magazine*, 11(4):1407–1422, December 2004.
- [25] A. Miller, S. Knoop, P. Allen, and H. Christensen. Automatic Grasp Planning Using Shape Primitives. *IEEE International Conference on Robotics and Automation*, 2:1824–1829, September 2003.
- [26] A. Morales, T. Asfour, P. Azad, S. Knoop, and R. Dillmann. Integrated Grasp Planning and Visual Object Localization For a Humanoid Robot with Five-Fingered Hands. In *IEEE/RJS International Conference on Intelligent Robots and Systems*, pages 5663–5668, 2006.
- [27] V. D. Nguyen. Constructing force-closure grasps. *International Journal of Robotics Research*, 7(3):3–16, 1988.
- [28] L. Perrault and P. Olivier. Bin-picking system for randomly positioned objects, Dec. 25 2007.
- [29] Point Grey Research. <http://www.ptgrey.com/products/bum>.
- [30] P. J. Sanz, A. Requena, J. M. Inesta, and A. P. Del Pobil. Grasping the not-so-obvious: vision-based object handling for industrial applications. *IEEE Robotics and Automation Magazine*, 12(3):44–52, 2005.
- [31] H. S. Yang and A. C. Kak. Determination of the identity, position and orientation of the topmost object in a pile. *Computer Vision, Graphics, Image Processing*, 36:229–255, 1986.
- [32] A. Zuo, J. Z. Zhang, K. Stanley, and Q. M. J. Wu. A Hybrid Stereo Feature Matching Algorithm for Stereo Vision-Based Bin Picking. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(8):1407–1422, 2004.

## Appendix A

# Gamma-Type Probability Distributions

The following description of a gamma-type probability distribution was taken from [23]. For a gamma-type random variable  $Y$ , the probability density function is given by:

$$f(y) = \begin{cases} \frac{y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} & \text{if } 0 \leq y < \infty; \alpha > 0; \beta > 0 \\ 0 & \text{elsewhere} \end{cases} \quad (\text{A.1})$$

where

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy. \quad (\text{A.2})$$

The mean is

$$\mu = \alpha\beta, \quad (\text{A.3})$$

and the variance is

$$\sigma^2 = \alpha\beta^2. \quad (\text{A.4})$$

This probability density function is non-negative, and is therefore appropriate for modeling non-negative random variables.

## Appendix B

# Histogram Data: Number of Picks

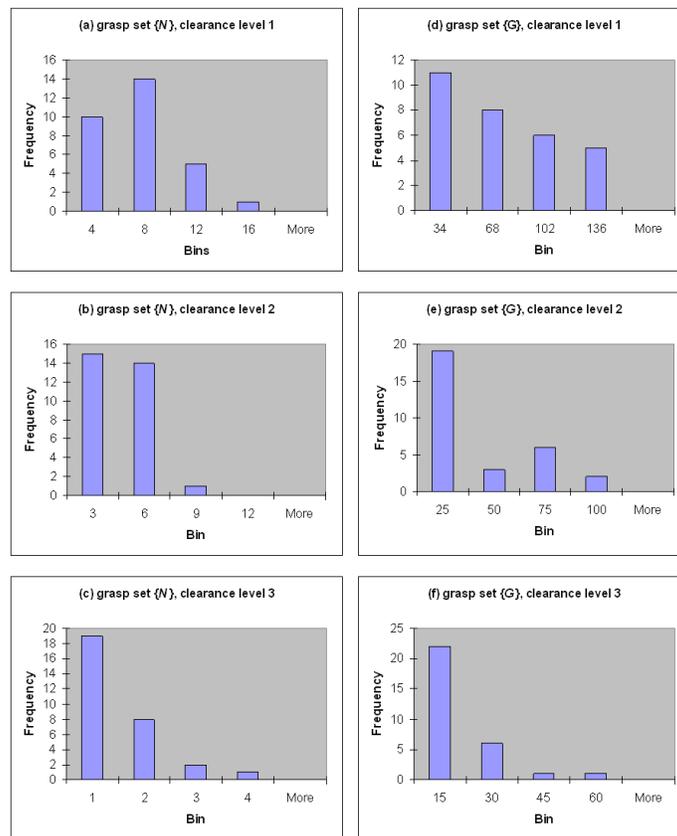


Figure B.1: Histogram data showing frequency of number of computed clear picks for 30 randomized piles, each consisting of 13 con-rods. Data is shown for grasp sets  $\{N\}$  and  $\{G\}$ , at three levels of clearance. For consistency, we limited the number of non-empty bins to 4.