Mining Haul Truck Pose Estimation and Load Profiling Using Stereo Vision

by

James Robert Borthwick

B.Sc., The University of British Columbia, 2003

A THESIS SUBMITTED IN PARTIAL FULFILLMENT 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

© James Robert Borthwick 2009
Abstract

Earthmoving at surface mines centers around very large excavators (mining shovels) that remove material from the earth and place it into haul trucks. During this loading process, the truck may be inadvertently loaded in a manner that injures the truck driver, or that results in an asymmetrically loaded or overloaded truck. This thesis presents two systems which aim to assist with haul truck loading: 1) a stereo-vision based system that determines a haul truck’s pose relative to the shovel housing as part of an operator loading assistance system, and 2) a system that can determine a haul truck’s load volume and distribution as the truck is being loaded. The haul truck pose estimation system is significant in that it is the first six-degrees of freedom truck pose estimation system that is sufficiently fast and accurate to be applicable in an industrial mine setting. Likewise, it is the first time that a system capable of determining a haul truck’s volume as it is being loaded has been described. To achieve this, a fast, resolution independent nearest neighbour search is presented and used within Iterative Closest Point (ICP) for point cloud registration. It also shown, for the first time, to the best of our knowledge, the possibility of using the Perspective-n-Point (PnP) pose estimation technique to estimate the pose a range-sensor derived point cloud model, and to use the same technique to verify the pose given by ICP. Camera errors, registration errors, pose estimation errors, volume estimation errors and computation times are all reported.
# Table of Contents

Abstract ................................................................. ii

Table of Contents ...................................................... iii

List of Tables ........................................................... vi

List of Figures ........................................................... vii

List of Algorithms ..................................................... ix

Acknowledgements ...................................................... x

Statement of Co-Authorship ............................................. xi

1 Introduction .......................................................... 1
   1.1 Introduction ...................................................... 1
   1.2 Safety and Efficiency in the Mining Industry ..................... 1
   1.3 The Need for Innovation .......................................... 3
   1.4 Thesis Context, Outline and Contributions ......................... 4
      1.4.1 Thesis Context and Goals ................................ 4
      1.4.2 Thesis Outline ............................................. 4
      1.4.3 Specific Contributions .................................... 4
   1.5 References ...................................................... 6

2 Rope Shovel Collision Avoidance System ........................... 8
   2.1 Introduction ...................................................... 8
      2.1.1 System Overview .......................................... 8
   2.2 Arm Geometry .................................................... 9
      2.2.1 Distal: Joint Sensor Based ................................ 11
      2.2.2 Proximal: Laser Scanner Based ............................ 11
   2.3 Swing Measurement .............................................. 12
      2.3.1 Encoder-Based Swing Measurement ........................ 13
      2.3.2 Camera-Based Swing Measurement ........................ 13
## Table of Contents

2.4 Truck Localization ............................................. 15  
2.5 Discussion and Conclusions ................................. 16  
2.6 References ....................................................... 18  

3 Haul Truck Pose Estimation Using Stereo Vision .......... 20  
3.1 Introduction ..................................................... 20  
3.2 System Design and Related Work ............................. 21  
3.3 Hardware Implementation ..................................... 24  
3.4 Software Implementation ...................................... 25  
3.5 Experimental Results .......................................... 29  
3.6 Conclusions .................................................... 33  
3.7 References ....................................................... 34  

4 Haul Truck Pose Estimation and Load Profiling .......... 37  
4.1 Introduction ..................................................... 37  
4.1.1 Shock and Vibration ....................................... 38  
4.1.2 Haul Truck Payload Volume, Mass and Distribution .... 38  
4.1.3 Research Goals and Contribution .......................... 39  
4.2 Related Work .................................................... 40  
4.2.1 Haul Truck Pose Estimation ................................. 40  
4.2.2 Haul Truck Payload Mass Estimation ...................... 41  
4.2.3 Haul Truck Payload Distribution and Volume Estimation ... 42  
4.3 System Requirements and Design ............................. 42  
4.3.1 Haul Truck Pose Estimation ................................. 43  
4.3.2 Load Profiling .............................................. 45  
4.4 Hardware Implementation ..................................... 46  
4.5 Software Design ............................................... 48  
4.5.1 Haul Truck Pose Estimation ................................. 48  
4.5.2 Haul Truck Payload Volume, Distribution and Mass ........ 60  
4.6 Experimental Evaluation ..................................... 61  
4.6.1 Experimental Setup ........................................ 61  
4.6.2 Qualitative Results ........................................ 61  
4.6.3 Accelerated Search Evaluation ............................. 63  
4.6.4 Error Analysis ............................................. 65  
4.6.5 Processing Speed ......................................... 67  
4.6.6 Load Profiling ............................................. 68  
4.7 Conclusions .................................................... 69  
4.8 References ....................................................... 71
Table of Contents

5 Conclusion ................................................................. 75
  5.1 Summary of Contributions, and Conclusions ..................... 75
  5.2 Directions for Future Work ........................................... 76
    5.2.1 Completing the Operator Loading Assistance System ........ 76
    5.2.2 Improving Haul Truck Pose Estimation and Load Profiling ... 77
    5.2.3 Further Capabilities ............................................ 78
  5.3 References ............................................................ 79

Appendices

A Equations ................................................................. 80
  A.1 Equation 4.2 ............................................................ 80
  A.2 Equation 4.3 ............................................................ 82
  A.3 Equation 4.5 ............................................................ 83
List of Tables

3.1 Error in ICP pose estimation. ........................................ 33

4.1 Comparison of an exhaustive search and resolution independent nearest neighbour search method. ........................................ 64

4.2 Uncertainty in ground truth determination using PnP ........................................ 66

4.3 Error in ICP pose estimation ........................................ 67

4.4 Uncertainty in load volume estimation ........................................ 69
# List of Figures

1.1 Early and modern mining shovels. .............................................. 2  
1.2 Mining fatalities in the United States. ................................. 3  

2.1 Shovel movements and the locations of the hardware components installed on the shovel. .................................................. 9  
2.2 The saddle block. ................................................................. 9  
2.3 Biaxial accelerometers mounted to boom joint. ...................... 12  
2.4 Accelerometers and range finder. ........................................... 12  
2.5 Rope shovel and laser scanner image. .................................... 13  
2.6 Swing angle camera mounting location. .................................. 14  
2.7 Swing angle camera image sequence. ..................................... 14  
2.8 Haul truck pose estimation hardware. .................................... 16  

3.1 A mining shovel loading a haul truck. ................................... 21  
3.2 Location of the stereo camera on the side of the shovel boom. ...... 25  
3.3 The stereo camera installed in its protective mount. .................. 26  
3.4 Before and after truck isolation. .......................................... 28  
3.5 A graphical representation of coarse-to-fine search. ................ 30  
3.6 Uncertainty in the stereo camera’s depth measurement (1280 × 960). 31  
3.7 Point cloud registration results. ........................................... 32  

4.1 A mining shovel loading a haul truck. ................................... 37  
4.2 The location of the stereo camera on the side of the shovel boom. . 47  
4.3 The stereo camera installed in its protective mount. ............... 47  
4.4 Before and after truck isolation. .......................................... 49  
4.5 Resolution independent search example. ............................... 51  
4.6 Comparison of exhaustive and resolution independent nearest neighbour search. .......................................................... 53  
4.7 Resolution independent search compared to a k-d tree approach. ... 56  
4.8 A truck body’s major features ............................................. 57  
4.9 Creating the haul truck model. ............................................. 58
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.10</td>
<td>The PnP control points.</td>
<td>60</td>
</tr>
<tr>
<td>4.11</td>
<td>Registration shown with a wireframe model.</td>
<td>62</td>
</tr>
<tr>
<td>4.12</td>
<td>The point cloud converging on the scene cloud.</td>
<td>62</td>
</tr>
<tr>
<td>4.13</td>
<td>Uncertainty in the stereo camera’s depth measurement (1024 × 768).</td>
<td>65</td>
</tr>
<tr>
<td>4.14</td>
<td>A haul truck before and after load profiling.</td>
<td>68</td>
</tr>
<tr>
<td>5.1</td>
<td>The operator loading assistance system’s subsystem hierarchy.</td>
<td>76</td>
</tr>
<tr>
<td>A.1</td>
<td>A single slice from a pinhole camera projection.</td>
<td>81</td>
</tr>
</tbody>
</table>
## List of Algorithms

1. Find Closest Point ................................................. 29
2. Find Nearest Neighbour ........................................... 53
Acknowledgements

I would like to thank my supervisors Drs. Peter Lawrence and Rob Hall who guided and encouraged me throughout this work.

Thank you also to my colleagues and friends Nicholas Himmelman, Ali Kashani, Mike Lin and Andrew Rutgers for the helpful discussions and insights, and for making the lab an enjoyable place to work.
Statement of Co-Authorship

This thesis presents research performed by James Borthwick.

- The introduction chapter was written by James Borthwick incorporating feedback and suggestions from his supervisors Peter Lawrence and Robert Hall.

- Chapter 2 was written by James Borthwick, Nicholas Himmelman, Ali Kashani and Mike Lin with feedback and suggestions from Peter Lawrence and Robert Hall. James Borthwick was primarily responsible for Section 2.4, which describes the hardware used for the research presented in this thesis. He also collaborated to prepare Figure 2.1, and collaborated to write the introduction and the abstract.

- Chapter 3 was written by James Borthwick with feedback and suggestions by Peter Lawrence and Robert Hall. The research and data analysis were performed by James Borthwick.

- Chapter 4 was written by James Borthwick with feedback and suggestions by Peter Lawrence and Robert Hall. The research and data analysis were performed by James Borthwick.

- The conclusion chapter was written by James Borthwick with feedback and suggestions by Peter Lawrence and Robert Hall.

The research program was identified and initiated by Peter Lawrence, Robert Hall and Tim Salcudean.
Chapter 1

Introduction

1.1 Introduction

Mining has been an essential element of human progress. Without the extraction of metal and coal, much of our present society would be profoundly different. While the benefits brought by mining are undeniable, the process of removing ore has often been undertaken at great human risk. Moving large quantities of earth, working underground or working with large machines can often be dangerous activities, as has been shown through many mining injuries and deaths.

From an alternate, business perspective, the need to provide product for a minimum cost has also been an ongoing concern. To maintain competitiveness and profitability, the mining operation must be as efficient as possible.

Although much progress has been made in both safety and efficiency, mines continue to be relatively dangerous worksites and the need to increase efficiency is ever present. The primary goal of this thesis is improve safety in surface (as opposed to underground) mining, and the secondary goal is to improve its efficiency.

1.2 Safety and Efficiency in the Mining Industry

As the goal of this thesis and the research project of which it forms a part is to improve safety and efficiency in mining, it is perhaps instructive to briefly consider the progress that has been achieved in these domains in order to appreciate where the industry stands today and where further improvements can be made.

It was the industrial revolution that is responsible for the creation of the large-scale mechanized mining that continues to this day. It greatly increased the demand for raw materials, as the ability to process them into goods grew dramatically; in turn, advances in mechanization and inventions such as the steam engine allowed mining to be undertaken at scales which were previously impossible. In fact, this trend of increased mechanization and increased scale has continued to this day, allowing for continually increased output and greater efficiencies within the mining industry [1].

An example of this improvement in efficiency can be seen in the technology of mining shovels, the large surface-mining excavators which remove material from the ground and
place it into transportation vehicles. Over the course of the 20th century, mining shovels have gone from being able to remove approximately 1 cubic meter per bucketload [2] to 76 cubic meters [3]. See Figure 1.1 for a photographic comparison of these machines.

![Early 20th century steam shovel operating in Kentucky.](image1) ![A modern shovel operating in British Columbia, 2008.](image2)

Figure 1.1: Early and modern mining shovels. For scale, take note of the size of the people beside and on the tracks in subfigure (b).

Although very efficient, large-scale mechanized mining has also been dangerous. Historically, large-scale mining has been an activity where accidental injury or death by falling, being crushed or being trapped in an explosion has, sadly, not been uncommon. It is not surprising then, that work towards improving the occupation’s safety was established long ago. For example, in 1815, scientist Sir Humphrey Davy was asked to improve safety by reducing the incidence of explosions in underground coal mines. After several months of study, he contributed what was perhaps the first scientifically derived improvement for mining safety: a specialized oil lamp that could be used safely in the presence of explosive methane gas, reducing the occurrence of deadly explosions and thereby saving many lives [4].

Government has also been instrumental in improving mining safety, which for many years had remained unregulated. In the United Kingdom, growing public consciousness of the unsafe conditions present in the country’s mines led to the passing of the Mines Act of 1842, the UK's first legislation aimed at improving conditions in the mining industry.

Through both technology and regulation, safety has continued to improve throughout the 20th century and into the 21st. In American mines, for the decade spanning 1911-1920, there was an average of 3,255 deaths every year. By the decade spanning 1999-2008, the average number of yearly deaths had fallen to 68 [5]. While a drop from 3,255 to 68 is a massive and impressive improvement, a life is still lost in an American mine, on average, more than once a week. See Figure 1.2 for a graph showing the number of yearly deaths in the United States.
1.3 The Need for Innovation

Although modern mines have never been safer or more efficient, the need for improvement on both fronts remains. Mining operations are still dangerous workplaces, where fatalities and major injuries are still too frequent. Efficiency is also an ongoing concern, as increasing the competitiveness and profitability of any business is always an objective.

It is therefore not surprising that research into techniques and products for improving safety and efficiency in mines continues to be pursued by private companies and academic researchers. There exist many recent patents, e.g. [6], [7], [8], and publications, e.g. [9], [10], [11], [12], with the goal of improving the safety and efficiency of the mining process.

As an illustrative example of high-tech research towards improving safety in surface mining, Ruff and Holden devised a system to reduce the incidence of haul trucks accidentally running over smaller vehicles or backing over cliffs at dump sites [13]. The system used GPS-based location sensors and peer-to-peer radio communication to create a proximity sensor capable of warning a haul truck driver when another vehicle or dangerous terrain was close by. Research such as this is highly valuable in that it can be used to reduce or eliminate injuries and deaths that regularly occur under similar circumstances.

On the efficiency side, the greatest leaps in surface mine efficiency have been achieved by engineering larger mining equipment: bigger shovels, bigger haul trucks, and bigger
crushers [1]. However, research by Bozorgebrahimi et al. has shown that efficiency gains through increased scale may be reaching its practical size limits [1]. This means that future improvements in efficiency will have to be realized through different means, such as making faster or more intelligent machines.

1.4 Thesis Context, Outline and Contributions

1.4.1 Thesis Context and Goals

At surface mines, mining haul trucks are massive off-highway trucks that carry ore from the pit to a destination where it is processed. The trucks are loaded in the pit by large excavators known as shovels. It is the aim of this thesis is to improve the safety and efficiency of truck loading at surface mines.

This work was conducted as part of a larger team effort, where the overall goal was to create a safety enhancing collision avoidance system to help prevent mining shovels from colliding with haul trucks. The primary goal of this thesis was to create one of the necessary subsystems of this collision avoidance system, specifically the ability to determined the position of a haul truck relative to the shovel that is loading it. A secondary goal was to enhance the efficiency of the mining operation by measuring the load in the haul truck as it is loaded, so that the shovel operator may deliver an optimal load.

1.4.2 Thesis Outline

The thesis is organized as follows:

1. Chapter 2 describes the shovel-truck collision problem, and presents the hardware and system architecture of the overall collision avoidance system. This chapter frames this thesis work in the context of the overall research project activities.

2. Chapter 3 is an introduction to the haul truck pose estimation system.

3. Chapter 4 describes the haul truck pose estimation system in detail, including a detailed error analysis. It also describes and assesses the haul truck load profiling system.

4. Chapter 5 contains concluding remarks about the system, its strengths and weaknesses, and potential areas for future work.

1.4.3 Specific Contributions

The specific contributions contained in this thesis are:
1.4. Thesis Context, Outline and Contributions

1. A method for determining the pose of a haul truck with high accuracy from a mining shovel.

2. A method for determining the volume and profile of a haul truck’s contents.

3. A method for determining the centre of mass of a haul truck’s contents.

4. A method for fast nearest-neighbour search within the context of ranger-sensor 2.5D data.

The thesis also contains experimental evaluations of each of the methodological contributions above.
1.5 References


1.5. References

Chapter 2

Rope Shovel Collision Avoidance System

2.1 Introduction

Rope shovels are used extensively in open pit mining to extract material from the earth and load it into haul trucks. The rate at which they are able to extract and load the material is often the limiting factor in a mine’s throughput, and as such, the shovels need to be run continuously in order to meet production targets. Unfortunately, the truck loading process is not without risk, as the dipper can collide with the haul truck during loading and self collisions are possible between the dipper and the caterpillar tracks. Although collisions do not typically occur on a daily or even weekly basis, when they do occur they can result in serious injury to the truck driver and expensive downtime and machine repairs for the mine. A system that is capable of warning the shovel operator when the dipper is on a collision course with a haul truck or with the shovel tracks could significantly reduce the likelihood of collisions, and therefore be of significant value to a mine.

2.1.1 System Overview

During the loading process, we assume that the truck is parked beside the shovel and that the shovel’s tracks are stationary. In order to avoid collisions during loading, we must determine the exact state of the shovel and the exact position of the truck. The state of the shovel is defined by the current position, velocity, and acceleration of the joints; the position of the truck is defined as its position relative to the shovel carbody. The task of determining the state of the shovel and position of the truck has been divided into three subsystems: shovel arm geometry, shovel swing angle, and truck location.

This is not the first attempt to instrument an excavator or mining shovel [1–3] or to locate a haul truck [1, 4, 5]. It is, however, to the best of our knowledge, the first...

---

published work describing a sensing system for a real-time collision avoidance system for a full-scale electric rope shovel.

In order to reduce the costs associated with installation and maintenance, the collision avoidance system is designed such that all the subsystems are installed on a shovel. This allows the data collected by each subsystem to be easily gathered by a central computer where collisions will be predicted. The information collected could be used to warn the operator of an impending collision, and if no corrective action is taken, briefly override the operator’s controls to prevent a collision. The following sections describe the hardware components and the data collected by each of the three subsystems. An overview of the hardware layout comprising all subsystems is shown in Figure 2.1.

![Figure 2.1: Shovel movements and the locations of the hardware components installed on the shovel: A1) accelerometers placed on the rotary joints, A2) rangefinder placed on the saddle block prismatic joint, C) central computing system located in house underneath operator cab, L) planar laser scanner placed below boom foot, S) stereo camera placed underneath house looking inwards, and T) truck localization stereo camera mounted on boom.](image)

![Figure 2.2: The saddle blocks are not rigidly connected to the dipper handle and can rotate back and forth depending on how much space is left between the dipper handle and the saddle block slide bars.](image)

### 2.2 Arm Geometry

Obtaining arm geometry variables is an important step towards building the collision avoidance system. The arm geometry of the shovel is determined by the position of the boom, the extension of the dipper handle, and the length of the hoist cables. The boom angle is set by adjusting the length of the boom suspension ropes whose length is kept constant during operation but can change slightly when the boom is lowered for maintenance.
2.2. Arm Geometry

During typical operation the arm geometry is controlled by the crowd and hoist functions. To determine the location of the dipper in space and the rate at which the joints are moving, the arm geometry can be directly measured in two ways. The hoist rope length and the dipper handle extension can be used to determine the location of the dipper. One difficulty that arises when measuring the hoist rope length is estimating the stretch in the hoist rope as it depends on the load and arm geometry. Alternatively, the angle of the dipper handle with respect to the boom can be used with the dipper handle extension to locate the dipper. By measuring the angle of the dipper handle with respect to the boom the uncertainty associated with stretch in the hoist rope can be avoided.

The angle of the dipper handle with respect to the boom can be measured between the saddle block and the boom. The saddle block acts as a brace which keeps the dipper handle on the shipper shaft. As the dipper is moved, the saddle block pivots with the dipper handle, making it ideal for measuring the angle of the dipper. One problem related to measuring the dipper handle angle on the saddle block is that the saddle block can twist back and forth on the dipper handle as the hoisting direction changes or as the direction of the torque applied to the shipper shaft changes. The way the saddle block moves on the dipper handle is shown in Figure 2.2.

Traditionally, joint angle or extension sensors are used in forward kinematics to determine the workspace location of an end-effector [6]. In this work, the arm geometry was measured both directly and indirectly: Directly by placing joint angle and extension sensors at each joint (distal sensors) and indirectly by tracking the position of the dipper with a laser scanner (proximal sensor) and using inverse kinematics to determine the dipper handle angle, dipper handle extension, and hoist rope length. Advantages of using distal sensors are:

- Complex algorithms are not required for obtaining the arm geometry.
- The processing simplicity makes the required computing power minimal.
- Visibility of the dipper is not required as it is with a proximal sensor where occlusion can degrade the arm geometry measurements.
- Distal sensors may be less sensitive to environmental factors such as dust and lighting

Advantages of using a proximal sensor are:

- Proximal sensors are less vulnerable to mechanical damage compared with distal joint sensors [7].
• Linkage singularities cannot cause numerical instabilities in calculations when using a proximal sensor, as they can for distal joint sensors [6].

• In some excavators, the dipper is not rigidly connected to the body, making forward kinematics impossible [8].

2.2.1 Distal: Joint Sensor Based

The first method for measuring the arm geometry used sensors placed at each of the joints. Pairs of biaxial accelerometers were used to measure the angles of the rotary joints (boom and dipper handle) and a laser rangefinder was used to measure the extension of the prismatic joint. The accelerometer-based angle measurement system used is part of a payload measurement system for hydraulic shovels called LoadMetrics, designed and manufactured by Motion Metrics International [9]. To measure the joint angle one biaxial accelerometer is placed on the proximal link of the joint and the other is placed on the distal link of the joint. Figure 2.3 shows the accelerometers installed on the boom joint. A second pair of accelerometers was installed on the saddle block joint to measure the dipper handle angle. The accelerometers are connected to a Motion Metrics International LoadMetrics computer that runs an algorithm which determines the difference in angle between the two sensors with an accuracy of ±1°. An offset is used to adjust the measurement according to the placement of the sensors.

The laser rangefinder used to measure the extension of the dipper handle was a SICK DT 500. It has an accuracy of ±3mm between 0.2 and 18 m on a surface with 6% reflectivity (black). The rangefinder was mounted on the saddle block aiming along the dipper handle, away from the dipper. A target was mounted on the end of the dipper handle opposite the dipper. The rangefinder was connected to the same computer used for the angle sensors via an RS422 serial connection. An offset was used to adjust the distance measured by the rangefinder to make it correspond to the dipper handle extension. Figure 2.4 shows the installation of the rangefinder and target on a shovel.

2.2.2 Proximal: Laser Scanner Based

Rather than using distal joint sensors, a proximally-mounted imaging sensor can be used to track the dipper. This section focuses on the development and implementation of a method for estimating the dipper location using a laser scanner.

To be operational on a mining shovel, a range sensor must be: reliable with a Mean Time Between Failures (MTBF) of greater than 1 year [10], accurate within 1% of the measurement range [10], and capable of range measurements up to 50m [11].

Computer vision, radar, and laser scanning have all been used for object tracking as well as dipper localization. While radars can provide the reliability and range, they are
slow and expensive compared with laser scanners and cameras [10, 12]. Winstanley et al. successfully utilized a laser scanner for their research after finding cameras to be less reliable than laser scanners when faced with environmental factors such as rain, dust, and shadows.

Laser scanners provide information in the form of point-clouds which are a set or array of points that describe object outlines and surfaces in the workspace. Each data point represents a vector from the scanner origin to the intersection of the laser beam and an object in the workspace. Typically, as in the laser scanner used in this work (SICK LMS 221), the laser beam is rotated to produce 2D measurements in a scan plane. Hence, the output can be described in a 2D scanned plane of the environment [13].

The laser scanner was mounted vertically underneath the boom pointing toward the dipper. In this position, the laser scanner maintained a consistent view of the dipper and occlusions were minimal. A typical scan plane of the laser scanner at the given position is provided in Figure 2.5 b) where the shovel boom, dipper, ground, and truck are evident in the point-cloud. The laser scanner provides 40 readings per second via an RS422 serial connection with a resolution of 1cm, accuracy of 5cm, and maximum viewing range of 80m.

### 2.3 Swing Measurement

Swing angle, the angle between the house and the lower carbody, is one of the measurements required for the computation of the dippers 3D position in workspace. Without swing angle, the collision avoidance system cannot determine the dippers position and cannot predict common collisions such as the collision of the dipper with the excavator
2.3. Swing Measurement

Figure 2.5: a) A P&H 2800 Electric Rope Shovel. b) A sample laser scanner image superimposed on equipment diagrams. The laser scanner is mounted vertically underneath the shovel boom. The laser scanners point cloud image, shown by a series of dots, represents a contour of the environment. Note that the above image is not taken during a normal digging cycle. Here, the laser sees the back of the truck, whereas during a normal loading sequence, the truck is located sideways and the laser only sees its side.

tracks or external objects such as a truck. Unfortunately, many shovels do not have a swing angle measurement system in place and one must be designed and retrofitted for this project.

2.3.1 Encoder-Based Swing Measurement

One method for obtaining the swing angle is to use an encoder to track the change in rotation angle that the swing motor shaft makes. For this purpose, a 100 line BEI HS35F encoder was attached to the shaft of one of the swing motors. In this configuration each quadrature increment on the encoder corresponds to a house rotation of $0.002^\circ$. The encoder was mounted to a stub shaft which was attached to the motor shaft, rising through the brake assembly. The brake assembly was removed to allow the stub shaft to be mounted to the motor shaft. The stub shaft had to be trued to the motor shaft to minimize wobble when the motor was spinning to prevent damage to the encoder. This laborious process would have to be repeated any time motor maintenance was required. A flexible coupling could be used to connect the encoder to the motor shaft but this would require a more complex mounting assembly which would in itself impede maintenance.

2.3.2 Camera-Based Swing Measurement

Given the drawbacks of using an encoder, a novel camera-based swing angle sensor which can be easily retrofitted and does not get in the way of regular maintenance was investigated. The proposed swing angle sensor consists of a Point Grey Bumblebee 2 stereo camera mounted on the bottom skirt of the excavator housing, looking down,
2.3. Swing Measurement

toward the center of the carbody (Figure 2.6). As the house rotates, the camera will rotate with the house and revolve around the stationary carbody, seeing differing views of the carbody (Figure 2.7). An ethernet cable carries digital video images from the camera to a computer in the house. The computer analyzes the images in real-time and determines the angle from which the camera is viewing the carbody. The desired swing angle accuracy of the system is $\pm 0.25$ degrees which corresponds to approximately $\pm 10$ cm error in dipper localization.

Figure 2.6: The left image shows the camera attached to bottom skirt of the house, its position indicated by white bounding box. The middle image shows how the camera is aimed downwards, toward the carbody centre. The right image shows a closeup of the camera.

Figure 2.7: Sample images taken by the camera as the house rotates clockwise.

For easy retrofitting, we are designing a system such that the camera need not be exactly positioned or angled when mounted to the excavator. As long as the lower carbody covers most of the view of the camera, the system will function properly. Further, the swing sensor should not need a prior model of the excavator as there are many variations of shovels. Thus the swing angle sensor automatically learns the 3D features on the carbody and calibrates the camera position and orientation with respect to the swing rotation axis.
2.4 Truck Localization

Once the shovel is fully instrumented, the task still remains of precisely locating the haul truck to avoid. As with the system developed by Stentz [1], we wish to determine the truck’s exact position, or “pose”, which can be fully described by six parameters – three for translation and three for rotation. However unlike [1], we require the system to work in real time without imposing new requirements or restrictions on how the shovel operator digs or loads a truck. The system must work in the background to quickly and accurately determine the six position parameters.

As stated previously, a goal was to place all equipment for the collision avoidance system on the shovel. This requirement restricts us from installing beacons or specialized GPS equipment on the trucks. As such, the use of a shovel-mounted 3D imaging sensor was seen as the best solution. Several 3D imaging sensors, namely stereo cameras, laser scanners and millimeter-wave radar units, were considered. Laser scanners and radar units were attractive because they deliver highly accurate 3D measurements, but unfortunately, also have slow image acquisition rates and low resolution images [1, 11]. Stereo cameras, meanwhile, deliver high-resolution 3D images with a high acquisition speed. However, stereo cameras suffer from the fact that their depth accuracy falls off exponentially with the distance between the camera and the measured object. This stems from the fact that stereo cameras work using the same triangulation principle as the human visual system. In order to mitigate this weak point, we selected a stereo camera with high-resolution sensors (1280 × 960) and a large separation between cameras (24 cm). This will allow for a triangulation depth accuracy on the order of ten centimeters at our working distance of about sixteen meters.

We chose to mount the camera high on the side of the shovel boom, far from potential sources of dust and damage. As the camera is not intended for use in outdoor or rough environments, we constructed a waterproof, heated camera enclosure and secured the camera on shock mounts. Figure 2.8 shows the cameras location on the boom, its enclosure, and its view of a haul truck.

The data produced by a stereo camera is called a point cloud, which is the same as a regular image except that for each pixel the precise \((x, y, z)\) location relative to the camera (and hence the shovel) is known. For the system to function, it must know which areas of the cloud represent the truck, as they must be avoided, and which represent the payload of the truck, as they may be gently brushed against by the dipper. Additionally, the \((x, y, z)\) measurements of the point cloud will contain errors which must not confuse or overwhelm the system.

What we wish to achieve is to be able to use this data to locate the truck from distances on the order of fifteen meters with an accuracy of about ten centimeters.
Figure 2.8: The left image shows a white arrow pointing at the location on the boom where the stereo camera is mounted. The middle image shows the stereo camera, mounted on the boom, in its protective enclosure. The right image shows the view of a haul truck delivered by the stereo camera.

Furthermore, this must be accomplished quickly enough to operate within the real-time time constraints of a collision avoidance system. We believe that the described hardware platform and resultant data will provide the basis for such a system.

2.5 Discussion and Conclusions

Three measurement subsystems have been described for a system to prevent collisions between the shovels dipper and a haul truck, or between the shovels dipper and its own tracks. Together, the three subsystems can provide to a collision avoidance system the position of the dipper, and the position of the haul truck with respect to the shovel’s carbody. The sensing subsystems are for:

- Arm Geometry, which measures the position of the dipper relative to the revolving frame. Two different approaches to obtain this information have been developed: a joint sensor based method which can be compared to the results from a laser scanner based method.

- Swing Angle, which measures the angle between the revolving frame and the fixed frame. This system relates the dipper position as found from the Arm Geometry subsystem, to the track body frame. We have also developed two approaches here: an encoder-based angle measurement subsystem which can be compared to the camera-based swing angle measurement subsystem.
2.5. Discussion and Conclusions

- Haul Truck Localization, which measures the pose of a haul truck with respect to the revolving frame. A camera based approach to localize the truck has been developed.

All the sensors have been developed so that they could be easily retrofitted. They are all attached externally to the shovel without requiring the shovel to be disassembled or extensively modified. A practised team could install all the hardware components in one 12 hour shift.

These sensors have been installed and tested on a shovel at the Highland Valley Copper mine. The current installation described here forms the test bed for determining the most appropriate set of sensors, and for the development of the collision avoidance system itself. Results will be presented at the conference.
2.6 References


Chapter 3

Haul Truck Pose Estimation Using Stereo Vision

3.1 Introduction

At surface mines, the earth is removed by large excavators (mining shovels), and placed into haul trucks. This process works whereby the shovel operator fills the bucket, swings the bucket over the truck’s bed, deposits the bucket load into the truck, swings back to the digging face, and repeats the work cycle. Each truck is typically filled with three to four bucket loads.

Ideally, the shovel operator deposits the contents of the shovel’s bucket into the centre of the haul truck. However, a momentary lapse of attention or judgement by the shovel operator can have severe consequences for the truck driver. If the shovel operator accidently moves the bucket so that it strikes the truck, or so that it dumps its contents on to the rear of the truck bed, it can cause the truck driver to experience a violent, unexpected shock and harsh vibrations. For the truck operator, this often means whiplash, injury to the back, or both. Unfortunately accidents such as this are not uncommon. An analysis of data available from the United States’ Mine Safety and Health Administration (MSHA) revealed that there were 21 such injury-causing events in 2008 at one mine alone.

In order to reduce the occurrence of these injurious events, a shovel operator loading assistance system is envisioned, and it is within this context that the following truck localization system was developed. Note that a loading assistance system requires some of the same capabilities as an autonomous shovel but only rarely removes control decisions from the operator. This partially autonomous system could function by intervening only in emergency situations, such as by preventing the operator from being able to release the contents of the bucket at a inappropriate truck bed location, or by very briefly altering the bucket’s trajectory if a collision were imminent. To develop a loading assistance system, two distinct sets of information must be determined: 1) the precise state of the shovel; that is, the position of each arm link and the shovel’s swing angle, and 2) the

---

1 A version of this chapter has been submitted. Borthwick, J.R., Lawrence, P.D., Hall, R.A., “Mining Haul Truck Pose Estimation Using Stereo Vision”.

20
precise position of the external object being interacted with, relative to the shovel (in this case the haul truck).

![Image of shovel dumping material into a haul truck](image)

Figure 3.1: A shovel dumping a load of material into a haul truck. The haul truck is approximately 5.9 m (19 ft) high.

This paper’s initial contribution is, to the best of our knowledge, the first presentation of the development and implementation of a system that determines in real-time the six degrees of freedom of the relative positions of a full-size shovel body and nearby haul truck. Secondly, the system makes use of a novel modification of previous ICP search methodology to permit real-time, time-guaranteed search for this application. The system was also designed with robustness and cost in mind. As such, it was developed to reside entirely on the shovel, to be easy to install via a retrofit, and to use relatively inexpensive commercially available components.

### 3.2 System Design and Related Work

The goal is to accurately determine the position of the haul truck relative to the mining shovel. Two different strategies are available: a) to independently and globally locate the shovel and the truck, and combine the information to determine their relative positions; or b) to use sensors and possibly beacons on one or both platforms to detect one in the reference frame of the other and process that information to determine the relative positions. The first strategy, which would effectively require a GPS-based system, was rejected for both cost and accuracy reasons. Although sub-centimeter accurate GPS systems have been demonstrated by employing carrier-phase differential error correction
[1], such systems in surface mines, where signal availability is limited by the high walls of the pit, are complex, require extra infrastructure to be installed around the mine, and do not achieve the overall instantaneous accuracy strived for in this application [2], [3].

The second strategy, determining the relative equipment positions locally, requires sensors and (optionally) beacons on the shovel, the trucks, or both. Given the choice, modifying solely the shovel has the following advantages:

- Lower system cost (because the trucks do not have to be modified),
- Single point of installation,
- Single point of maintenance.

For this reason the decision was made to modify the shovel only. In order to detect a truck from the shovel, several sensor options were considered:

1. monocular cameras
2. structured light (monocular camera and a projected grid of laser light)
3. millimeter-wave radar systems
4. laser scanners
5. stereo cameras

Below is a suitability assessment of each sensor, including when possible, how each has performed in previous work in a similar application.

1. Monocular cameras do not inherently provide depth information. As the goal is to determine the relative six degrees of freedom pose of the truck, such a system would have to rely on detecting natural features or artificial markers that occur at known locations on the truck. A good survey of this type of feature-dependent pose estimation was written by Lepetit and Fua [4]. As stated earlier, altering the trucks is not desirable, which means natural features would have to be used for the pose estimation as demonstrated by Gordon and Lowe [5]. A natural feature-based method was rejected primarily due to concerns that the muddy, dusty mine environment would conspire with shadows to create many features unique to individual haul trucks, and worse, to have these features change over time.

2. A monocular camera used in combination with a grid of laser light (structured light) can also be used to determine depth. Most structured light applications thus far have been used in indoor situations, where the laser light does not have to overpower
the sun. However, there has been some research towards using this type of system outdoors [6], [7]. What is demonstrated by these papers, as far as structured light for this application is concerned, is that no high-speed eye-safe grid laser scanner with sufficient range under sunlight has yet been developed.

3. Millimeter-wave radar has been successfully employed outdoors in the mining industry for environment monitoring and mapping [8], [9], [10]. The GHz radiation used in these devices has the desirable property of being able to better image through dust and fog than the frequencies used by lasers or cameras [11]. However, mm-wave radar demonstrated to date is not fast enough for a high resolution real-time application. The custom built high speed unit developed by Widzyk-Capehart et al. attained a data rate of 173 data points/second [8], which does not deliver the necessary resolution quickly enough for this real-time application.

4. Time-of-flight based laser scanners have previously been the sensor of choice for excavator-mounted truck localization systems [12], [13], [14]. In 1999, Stentz et al. presented a system that used a 3D laser scanner to locate trucks next to a construction-sized excavator [12]. However, the laser scanner was required to remain stationary during its three second exposure time, which is not suitable for this application because the operator could choose to swing the shovel (and thus the laser scanner) at any time, corrupting the scan data. The three second pause is also unacceptable from a production point of view. As with mm-wave radar, this slow exposure time is a general problem with 3D laser scanners because they need to physically scan point-by-point to acquire the data.

A laser scanner was also employed by Duff, but the system was significantly different in that a horizontally mounted 2D \((r, \theta)\) laser scanner was used to localize the truck [13]. The system was developed on 1/7th scale models of a shovel and haul truck, where the objective was to be able to guide a haul truck into a loading position. This system located the truck using a single horizontal scan line, and assumed that the shovel and truck resided on the same plane. Although sufficient for guiding a truck into a loading position, the same-plane (three degrees of freedom) assumption does not deliver the required accuracy for a loading assistance system.

As with [13], Green et al. also propose to use a 2D laser scanner, but to mount it vertically instead of horizontally on the side of the shovel [14]. This allows the system to combine scans at various angles as the shovel swings, and consider them together as a single scene. Unfortunately, using simulated laser scan data, the system required approximately 15 seconds worth of data to localize the truck, and was demonstrated using the same same-plane assumption as [13].

5. Stereo cameras have also been deployed in the mining industry [15], [16], although less frequently than either laser or radar scanners. Stereo cameras offer several
advantages over laser and radar scanners:

- **Speed**: The time to acquire the raw data is constrained only by the shutter speed of the camera system.

- **Parallelism**: All data points are acquired simultaneously and not via a point-by-point procedure, ensuring all data points represent the same moment in time.

- **Data Density**: Using megapixel cameras a stereo image can contain more than a million data points. Laser scanners and millimeter wave radar systems are so far incapable of delivering the same resolution.

- **Cost**: Digital cameras are ubiquitous and inexpensive.

- **Reliability**: Cameras with no moving parts can be employed, which decreases susceptibility to damage from shock and vibration.

However, stereo cameras do require a visible scene texture to determine depth, and require an initial calibration. These two shortcomings are not generally a problem as the mining environment is rich in texture and the initial calibration, once complete, should not have to be repeated. Of potentially greater concern is that cameras suffer difficulties when imaging through atmospheric disturbances such as thick dust, fog or heavy rain or snow. This means that there may be times when the system is unable to operate, but a human operator also requires sufficiently visible scene elements to control the machine safely and effectively. Finally, perhaps the greatest weakness of stereo cameras is that their depth accuracy falls off exponentially with the distance of the object being measured. However, by using a camera with a sufficiently wide baseline and high-resolution sensors, it is possible to obtain the depth accuracy needed for this application's working distance.

### 3.3 Hardware Implementation

The stereo camera selected for the system is the commercially available Bumblebee XB3 produced by Point Grey Research Inc. It was chosen for its 24 cm wide baseline, its high resolution 1280 × 960 CCD sensors and its wide 66° horizontal field of view. The camera was mounted high on the side of the shovel’s boom, as shown in Figure 3.2. This position was chosen to give the camera a high perspective of the haul truck and to keep it far from sources of dust and potential damage. Also, since the camera is not constructed for outdoor industrial use, it needed to be protected from the environment. A waterproof, shock dampened, heated enclosure was designed and constructed for the...
camera, and in Figure 3.3 it is pictured mounted on a shovel boom with the camera installed.

![Location of Stereo Camera]

Figure 3.2: Location of the stereo camera on the side of the shovel boom.

### 3.4 Software Implementation

The software system is responsible for processing the data produced by the stereo camera to determine the six degrees of freedom (three translation, three rotation) relative position of the haul truck.

The data produced by the stereo camera may be processed into real-world $x$, $y$ and $z$ co-ordinates. This data is sometimes referred to as 2.5D, because it represents 3D data projected onto a two dimensional plane (the image sensor) from a single perspective, and so the scene data is incomplete. The collection of $(x, y, z)$ data points, collectively called a point cloud, is used to determine the truck position. This is accomplished by positioning a second “model” point cloud representing a model of the haul truck directly onto the scene point cloud generated by the stereo camera. Since the dimensions and original orientation of the model cloud are precisely known, once it is aligned or registered with the haul truck in the scene cloud, it can be inferred that the model haul truck cloud position is the same as the scene’s haul truck position.

The standard method for registering two point clouds with high accuracy is the iterative closest point (ICP) method, first introduced by Besl and McKay [17]. A good review of the ICP procedure and techniques for increasing the algorithm’s speed was
3.4. Software Implementation

Figure 3.3: The stereo camera installed in its protective mount. The camera is connected to a computer located in the shovel housing by wires run through waterproof conduit.
written by Rusinkiewicz and Levoy [18]. A more recent analysis, comparing ICP to other procedures for fine point cloud registration, showed that few other methods have been developed and that ICP is still the basis for the fastest methods [19]. ICP was thus selected as the base algorithm to use to align the scene and model point clouds.

As its name suggests, ICP registers two clouds in an iterative fashion. It does so by matching points from one cloud to their closest points in the other, and minimizing an error metric based on these pairs at each iteration. The essential steps taken at each iteration are: 1. select points from the model cloud, 2. match each point to its nearest neighbor in the scene cloud, 3. optionally weight and/or reject point pairs, 4. calculate an error, and 5. minimize the error. Variants optimizing one or more of the steps for speed and robustness have been developed, and [18] and [19] may be referred to for an overview. The ICP variant that has been implemented for this application uses a point-to-plane error metric [20], [18] and a new accelerated point-matching method, described below.

The point-matching step is often a target for improving the overall speed. This is because unlike the other steps, a “naïve” approach runs in $O(N_M N_S)$ time, where $N_M$ is the number of model points, and $N_S$ is the number of data points.

The first measure taken to reduce the runtime of the search is to simply reduce the number of model points and scene points involved. The number of model points, 150, is selected as a compromise between speed and accuracy. The number of scene points is reduced by removing all the non-truck points from the scene cloud (see Figure 3.4), which increases both speed and accuracy. The pruning of the scene cloud is accomplished by only including points that are between 3 and 9 meters off the ground, and closer than 30 meters to the shovel.

The second and more notable strategy is to search the scene cloud in a more efficient manner. This is a classic optimization problem known as the nearest neighbor search (NNS). A now standard acceleration method is to structure the scene data into a k-d tree [21] as first demonstrated in the ICP context by Zhang [22], however this method requires an $O(N_S \log N_S)$ tree-building operation every frame, making it unsuitable for this real-time application. Other search acceleration methods such as [23] and [24] have been developed but they too rely on a pre-processing step which effectively rules out their use in this application due to time constraints. A technique called closest-point caching presented by Simon et al. does not require pre-processing, but was shown to speed up the searches by only approximately 10% [25]. Another approach is to perform reverse calibration [26] which is very fast ($O(n)$) and requires no pre-processing. However, reverse calibration does not truly attempt to find nearest neighbors and as a result does not produce as accurate a registration as do nearest neighbor algorithms.

In [27], Jost and Högli present a method that takes advantage of the hypothesis
that points close to each other in the first cloud will be matched to points that are
close to each other in the second cloud. This method has a best-case time complexity
of $O(N_M)$ and (if no pre-processing is done) a worst case complexity of $O(N_M N_S)$. They also suggest a “multi-resolution” approach where the number of matched points
increases as the registration error decreases so that each iteration can be made with as
little computational effort as possible [28]. For this application, it was found that this
technique improved the performance when using a large number of matched points, but
that better performance was obtainable by using a smaller number of matched points
and the method described below. (The methods in [28] may be more beneficial when
the cloud surfaces being matched are more intricate than those presented here as they
will need to use more matched points to achieve a good registration.)

The relatively simple search method used by this system has proved efficient, runs in
guaranteed time and requires no pre-processing. As with [28], the method can be thought
of as multi-resolution; however, as opposed to [28], it does not involve increasing the
number of matched points as the clouds converge, but rather increasing the spatial
sampling frequency of the scene cloud during each closest-point search in the matching
step.

Specifically, it functions by taking advantage of the fact that the scene points are
not randomly ordered. They are arranged in a 2D grid (as delivered by the cameras’
CCDs) that is the projection of an approximately smooth physical world. Consequently,
a physical \((x_1, y_1, z_1)\) location stored in the scene cloud point projected to pixel position \((i, j)\) may be approximated by the physical location \((x_2, y_2, z_2)\) in the scene cloud point’s neighbor projected to pixel \((i\pm1, j\pm1)\). Therefore, instead of sampling every point in the scene cloud, a coarse-to-fine search procedure may be adopted. This procedure works, for example, whereby every 20th scene point (pixel) along each axis of the projection is sampled, taking note of the closest. Then a search window is formed around the closest scene point found, and sample that area more finely. This coarse-to-fine method can have as many “layers” as desired. (Three layers have been used for our implementation.) Figure 3.5 shows an example of a 2-layer search procedure, and Algorithm 1 shows it expressed as pseudo code.

Note that in common with other approximation methods (e.g. [27]), this method does not absolutely guarantee that the nearest neighbor will be returned. The results, however, for this application are very good and the execution time, which is guaranteed, is reduced in time complexity to \(O(N_M \log N_S)\) with no need for pre-processing. Note that the effective use of the search method requires that each projected surface patch on the object sampled be larger than the separation between samples at the coarsest sampling level. This requirement is in some respects analogous to the Nyquist-Shannon sampling theorem: the sampling must be sufficiently frequent, or the true structure of the data cannot be determined.

Algorithm 1 Find Closest Point

```pseudo-code
closest_point ← VERY_DISTANT_POINT
for every 8th projected point per axis do
    if point is closer than closest_point then
        closest_point ← point
    end if
end for
for every point in a 9 × 9 window centered at closest_point do
    if point is closer than closest_point then
        closest_point ← point
    end if
end for
return closest_point
```

3.5 Experimental Results

Two sources of error in the system were analyzed. The first source of error is in the data points produced by the stereo camera. The depth accuracy, \(\Delta z\), is determined by \(\Delta z = fB/(d - m) - fB/d\) where \(f\) is the focal length of the camera, \(B\) is the stereo baseline, \(d\) is the disparity and \(m\) is the correlation error. At distances of 12 m and
Figure 3.5: A graphical representation of a 2-layer coarse-to-fine multi-resolution search. The coarse search measures the distance to all points located at the intersection on the large grid 1 (25 points), which is followed by a fine search 2, centered at the closest point on the coarse grid (81 points). In this example, the coarse-to-fine search is approximately 12 times faster than an exhaustive search.
16 m, the uncertainties in \( z \) are ±0.061 m and ±0.109 m respectively. See Figure 3.6 for a graph of depth uncertainties. The \((x, y)\) translational errors are determined by \( \Delta x = \Delta y = pz/f \), where \( p \) is the stereo calibration error, \( z \) is the depth co-ordinate of the data point and \( f \) is the focal length of the camera. At distances of 12 m and 16 m, the \((x, y)\) uncertainties are ±0.005 m and ±0.007 m respectively. Note that although individual data points have an error greater than some high precision GPS systems, the stereo camera errors are approximately Gaussian distributed [29] and as ICP is a least-squares method, the large number of stereo data points help to mitigate the errors' overall effects.

![Graph of depth uncertainties](image)

Figure 3.6: The uncertainty in the stereo camera’s depth measurement at 1280 × 960.

A second source of error exists in the fit provided by the ICP image registration process. The implemented point-to-plane ICP (described in detail in [20]) operates by minimizing the sum of the square of the distances between the planes defined at model cloud points that have been matched to scene cloud points. The RMS error is the error in fit between the model and image clouds, which is present due to errors in the stereo camera data and in imaging the truck from different perspectives. Testing showed that the RMS error at distances on the order of 15 m would typically lie between 0.05 and 0.1 m, or about 0.5%. See Figure 3.7 for an example of unregistered and registered clouds.

Because the system is required to operate in real-time, the speed at which the stereo data can be processed and the truck position determined is also an important measure of performance. Operating on a 2.4 GHz Intel Core 2 Duo CPU with 667MHz DDR2 SDRAM, at a resolution of 1280 × 960, the system performs stereo processing at 3.8 Hz, and performs both stereo processing and pose estimation at 2.5 Hz. See Table 3.1 for speeds for lower camera resolutions. Note that the system is single threaded, and
Figure 3.7: The first row shows examples of the truck model (white) adjacent to a truck image cloud (grey). The left image is of the model image cloud, which is used to align the clouds, and the right image is the model’s wireframe representation. The two clouds are not registered. The second row is an example of the model registered with the truck cloud after 29 ICP iterations. The rotation was determined to be 171.13 degrees about axis $[0.193, -0.335, 0.921]^T$ and the translation was $[-2.95, 0.18, 16.74]^T$ meters. The RMS ICP matching error was 0.063 meters.
as both stereo processing and ICP point-matching can be programmed in a parallel manner, the speed of the system should scale with the number of cores used.

<table>
<thead>
<tr>
<th>Resolution (Pixels)</th>
<th>Stereo Processing (Hz)</th>
<th>Stereo Processing + Pose Estimation (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>320 × 240</td>
<td>12.0</td>
<td>9.3</td>
</tr>
<tr>
<td>640 × 480</td>
<td>10.0</td>
<td>6.9</td>
</tr>
<tr>
<td>1024 × 768</td>
<td>5.3</td>
<td>3.8</td>
</tr>
<tr>
<td>1280 × 960</td>
<td>3.8</td>
<td>2.5</td>
</tr>
</tbody>
</table>

Table 3.1: Error in ICP pose estimation.

3.6 Conclusions

The present state of development of a real-time stereo-vision based haul truck localization system has been described. The system is to form a core component of a haul truck loading assistance apparatus that will help reduce injuries caused by shovel/truck collisions and dangerous loading errors. The truck localization system was designed to reside entirely on a mining shovel and be comprised of commercially available components to reduce the initial and ongoing costs. It was demonstrated using data gathered from a full-size industrially operated shovel and haul truck.

The novel six degrees of freedom haul truck localization system used a fast, nearest neighbor search algorithm without pre-processing as part of an ICP approach. As no sufficiently fast nearest neighbor algorithm was found, a novel, multi-resolution ICP search strategy was employed to obtain real-time rates. This allowed the complete system to achieve a speed of 2.5 Hz, which meets industrial production requirements. The experimental results showed an RMS point-to-plane matching error of 0.05 to 0.10 m at typical truck loading ranges of 12 to 16 m from the camera. The proposed methods could prove most advantageous with similar real-time ICP range-data problems in which data pre-processing is not possible.
3.7 References


3.7. References


Chapter 4

Haul Truck Pose Estimation and Load Volume Estimation and Profiling using Stereo Vision

4.1 Introduction

At surface mines, the earth is removed by large mining shovels, and placed into haul trucks. The shovel operator fills the bucket, swings the bucket over the truck bed, deposits the bucket load into the truck, swings back to the digging face, and repeats the work cycle. Each truck is typically filled with three to four bucket loads.

Figure 4.1: A shovel dumping a load of earth into a haul truck.

Two major things that can go wrong during this process are: 1) the shovel operator can injure the truck driver by accidently subjecting the truck body to shocks and vibrations, and 2) the shovel operator can deposit a load with poor volume, mass or

\footnote{A version of this chapter will be submitted. Borthwick, J.R., Lawrence, P.D., Hall, R.A., “Mining Haul Truck Pose Estimation and Load Profiling Using Stereo Vision”}. 

37
distribution characteristics. Each of these problems is detailed below.

4.1.1 Shock and Vibration

Mining shovels are massive machines: a modern shovel (e.g. P&H 4100XPB) has a mass on the order of 1,100 metric tons, and its bucket may contain a payload of over 100 metric tons. When a haul truck ready for loading is parked beside a shovel, the shovel operator must skillfully load the truck in order to avoid injuring the truck driver. Ideally, this is done by depositing the bucket’s payload into the centre of the truck bed from a minimal height. However, a momentary miscalculation or lapse of attention on behalf of the shovel operator can cause the shovel to accidently transfer dangerous amounts of kinetic energy to the truck and its driver.

For instance, a shovel operator may accidently strike the truck with the bucket, subjecting the truck driver to a violent, unexpected shock and harsh vibrations. Similarly, the operator may, in error, dump the 100 tons of earth onto the rear of the truck bed (its “duck tail”), again creating a dangerous shock and large vibrations. These jolts can immediately injure the truck driver’s back, whiplash his/her neck, or both. Unfortunately accidents such as these are not uncommon. An analysis of the data available from the United States’ Mine Safety and Health Administration (MSHA) revealed that there were 21 such injury-causing events in 2008 at one mine alone.

In addition to immediate injury, truck drivers are also at risk for longer-term problems. The shovel operator may deposit the bucketload in a manner that does not produce a shock severe enough to cause immediate injury, but which does produce vibrations which are dangerous when one is exposed to them repeatedly. In fact, a study by S. Kumar showed that haul truck drivers are routinely exposed to vibrations above ISO standards (ISO 2631-2:2003) throughout the hauling process, including during loading [1]. Although the study did not directly consider the quality of the loading, it stands to reason that the poorer the loading technique, the more severe the vibrations and thus the greater the long-term health risks.

4.1.2 Haul Truck Payload Volume, Mass and Distribution

In addition to loading a truck in a safe manner, the shovel operator must also strive to fill the truck with an optimal load. An optimal load can be defined in different ways, but it is always designed so that the haul truck will carry the greatest amount of material without overloading it. At some mines, this optimal load is defined as the total mass that each truck should carry, while at others it is defined in terms of volume. In both cases, the load should be symmetrically distributed from left to right.

While underloading a haul truck results in a drop in mine efficiency, overloading is
in fact the greater concern. Overloading a haul truck can cause premature failure of its components [2], greatly increasing maintenance costs. Overloading may also have warranty implications: certain haul truck manufacturers stipulate that if a truck is overloaded by more than 20% must be immediately unloaded or the truck’s warranty may be voided.

Further, uneven left-right load distributions are a problem as they can effectively overload one side of the truck even if the total mass is within limits. An unevenly loaded haul truck may also damage its body while travelling, because it can cause twisting of the body’s frame (frame racking) [3]. In fact, frame racking can further increase the vibration experienced by the truck’s driver. For these reasons, an evenly distributed load is important.

4.1.3 Research Goals and Contribution

This research has two major goals. The primary goal is to mitigate the safety-related problem: the accidental positioning of the shovel bucket such that it strikes the truck or deposits its payload in a manner that injures the truck driver. The secondary goal is to improve the load quality which will reduce damage to the trucks and the cost to maintain them.

To accomplish the primary safety-related goal, two distinct sets of information must be determined: 1) the precise state of the shovel; that is, the position of each arm link and the shovel’s swing angle, and 2) the precise position of the haul truck relative to the shovel body. This paper develops a method to determine the relative position of the haul truck. The primary contribution is to thoroughly present the entire six degrees of freedom real-time truck pose estimation system, to analyze the methods used within and to assess in detail its overall accuracy.

An additional contribution is to introduce for the first time an excavator-based system that can help the shovel operator achieve an optimal haul truck load by quantifying multiple aspects of a haul-truck’s load. Specifically, it can determine during loading, the volume and the distribution characteristics of the load, as well as present a visually instructive relief map of the truck’s payload.

Both contributions make use of a novel adaptation of previous ICP search methodology to permit fast, time-guaranteed search for this application. The entire system was also designed with robustness and cost in mind. As such, it was developed to reside entirely on the shovel, to be easy to install via a retrofit, and to use relatively inexpensive commercially available components.
4.2 Related Work

4.2.1 Haul Truck Pose Estimation

This work is not the first with a requirement to determine the pose of a haul truck relative to a shovel. In a similar application, Stentz et al. created an autonomous construction-size excavator control system capable of locating and loading a truck [4]. The system found the truck’s location by fitting planes to 3D laser scanner data, and from that inferring the six degrees of freedom pose of the truck. The system was capable of loading a truck as quickly as a human operator despite what were reported as slight inaccuracies in the determination of the truck’s position. The system also required the excavator to stay stationary for three seconds while the scan data was obtained, which is a restriction that is not practical for the assistive system that is proposed here, nor is it acceptable from a production point of view.

In 2006, Duff presented a real-time system that was based on a 1/7th scale model of a shovel and haul truck, where the objective was to guide a haul truck into a loading position [5]. For this purpose, a 2D laser scanner was used to determine the relative position of the truck based on the minimum enclosing rectangle of its data points. The system functioned in real-time, but assumed that the shovel and truck resided on the same plane and hence only determined two translation and one rotation parameters. Although these three degrees of freedom are sufficient for guiding a truck into a loading position, it does not deliver the required accuracy for a loading assistance system.

Another approach for relative haul truck pose estimation was designed and reported by Green et al. [6]. The proposed system involves vertically mounting a 2D laser scanner to the side of a shovel to acquire multiple scans as the shovel swings, and to combine these scans together into a single scene representation. The system then uses a Bayesian probability method to segment truck data points based on which plane of the truck’s structure they reside in. Planes are then fit to each group of data points, and then these best-fit planes are used to derive the position of the truck. Like [5], the system was demonstrated using only three degrees of freedom, although it does offer the possibility of being directly extendible to six. Unfortunately the demonstrated system, based on simulated data, required approximately 15 seconds worth of data to achieve a reasonable pose estimate, and even then had a translation error of about 0.6 m. The authors speculated that the final error in position may be due to an implementation error, and that further investigation was required.
4.2. Related Work

4.2.2 Haul Truck Payload Mass Estimation

The traditional method of weighing the contents of a truck is to weigh the truck on a set of scales before and after loading, and then infer that the difference in mass is the mass of the truck’s contents. Although this is a mature technology [7], its major drawback is that it generally slows down the hauling process. This is because the truck may need to slow down or completely stop on the scales, wait in a queue to use the scales, or take a detour if the scales are not directly between the shovel and the dumping location.

More recent innovations have eliminated the need for trucks to visit a truck scale. The most common technology is to include pressure sensitive transducers in the support system of a haul truck. Recently patented commercial systems of this type claim to have an accuracy that is usually better than 3%, which on a 218 metric ton capacity truck equates to 7 metric tons [8]. While a truck-based load measurement system eliminates the need to visit a scale, it imposes a new requirement that a separate weigh system be purchased and maintained for each individual truck. This requirement means that the cost of the system increases with each truck in the fleet. As well, it adds truck maintenance overhead which potentially includes periodic sensor re-calibration.

With the disadvantages of direct truck-load measurement systems apparent, indirect shovel-based systems have been developed. These systems work by weighing the contents of each bucket load, either by monitoring the entire shovel system (e.g. its motor currents) or by using a point-sheave mounted load cell, and then adding the bucket load weights together on a per-truck basis to determine the truck’s total payload. Most research towards electric rope shovel-based payload systems has been performed by commercial entities, and as such detailed, refereed publications on the systems’ designs and performances have not been made available. These systems, which are made by both original equipment manufacturers and by third parties, claim to have accuracies ranging from 2% to 5% on a per-truck basis. It is however in the author’s experience in speaking with mine personal at six mines in two countries that mass measurement systems (both truck and shovel based) are not fully satisfactory due to reliability problems.

Although academic research into shovel payload monitoring systems has been less extensive, it is an area of interest. In his Ph.D. thesis, Wauge develops a Kalman filter-based mining shovel payload monitor with a mean error of 1.96%, and a standard deviation of 4.76% [9].

Although existing systems that measure the payload of each shovel bucket are potentially more useful and less expensive than truck-based systems, it is hard to compare relative accuracies due to an absence of refereed publications. An independent field study of competing systems would likely be necessary to truly know the relative strengths and weaknesses. One area, however, where these shovel-based system are not as capable as
truck-based or scale-based systems, is their ability to determine the load distribution inside the truck. Although the total load may be within limits, it may be loaded in a manner where one side is in fact overloaded. Existing shovel-based systems are so far incapable of detecting this problem.

4.2.3 Haul Truck Payload Distribution and Volume Estimation

Payload distribution is usually found as a byproduct of measuring the mass using external or truck-mounted scales. However, it can also be determined directly, by optically measuring the surface of the haul truck’s payload. In fact, measuring the load distribution optically as the shovel is loaded provides information and other data opportunities that a mass-only measurement system cannot. This information includes the total volume, the dipper fill factor, and the opportunity to analyze the material fragmentation of the load. Additionally, with knowledge of the payload’s density, one can also estimate the mass. All of this supplementary data would be of significant value to a mine.

A study into optically measuring the payload of haul trucks was conducted by Duff [10]. In this study, two planar laser scanners were mounted on a pole which was extended over the haul truck road from a large scissor lift. This enabled the system to view the payload of the truck as it slowly (9 km/h) drove under the system. Unfortunately, details on the system’s overall accuracy were not given. This stationary, external measurement has the same drawbacks as external mass measuring scales (discussed in Section 4.2.1). The technology has been successfully deployed to the mining industry by licensing it to the mining services company Transcale Pty Ltd.

Other commercial laser-based truck volume measurement systems exist, such as those produced by Tally Clerk Ltd. and Woodtech Measurement Solutions. Although not specifically designed for the mining industry, these systems are also based on laser scanners mounted above a truck in order to measure the volume of its contents. They are often used to measure the quantity of wood chips, gravel or coal leaving an operation, where the material is sold by volume rather than weight in order to discount any extra weight added by rain or other moisture.

4.3 System Requirements and Design

The system’s overriding goal is to locate the haul truck as part of an operator loading assistance system, and as such it was designed around this requirement. Because an operator loading assistance system must operate in real-time, the truck pose must be available on a real-time basis; however, since a haul truck does not move during the loading process, it needs to be located only once, with the relative shove/truck positions
thereafter being computed solely based on the movements of the instrumented shovel. Therefore, the truck must be accurately located once, in the time during which the shovel’s initial bucket is filled, which is typically 5-10 seconds.

To determine an acceptable system accuracy, it was reasoned that the bucket should be able to safely move within the central 90% of the truck box’s internal area. As the truck’s width is approximately seven meters, this necessitates an overall system accuracy of ±0.35 m. However, errors in determining the state of the shovel will also be present. If both systems exhibit the same level of accuracy, then the truck pose estimation must be accurate to within ±0.25 m.

4.3.1 Haul Truck Pose Estimation

To develop a truck pose estimation system, two different strategies are available: a) to independently and globally locate the shovel and the truck, and combine the information to determine their relative positions; or b) to use sensors and possibly beacons on one or both platforms to detect one in the reference frame of the other and process that information to determine the relative positions. The first strategy, which would effectively require a GPS-based system, was rejected for both cost and accuracy reasons. Although sub-centimeter accurate GPS systems have been demonstrated by employing carrier-phase differential error correction [11] such systems in surface mines, where signal availability is limited by the high walls of the pit, are complex, require extra infrastructure to be installed around the mine, and do not achieve the overall instantaneous accuracy strived for in this application [12], [13].

The second strategy, determining the relative equipment positions locally, requires sensors and (optionally) beacons on the shovel, the trucks, or both. Given the choice, altering solely the shovel has the following advantages:

- Lower system cost (because the trucks do not have to be modified),
- Single point of installation,
- Single point of maintenance.

For this reason the decision was made to modify the shovel only. In order to detect a truck from the shovel, several sensor options were considered:

1. monocular cameras
2. structured light (monocular camera and a projected grid of laser light)
3. millimeter-wave radar systems
4.3. System Requirements and Design

4. laser scanners

5. stereo cameras

Below is a suitability assessment of each sensor, including when possible, how each has performed in previous work in a similar application.

1. Monocular cameras do not inherently provide depth information. As the goal is to determine the relative six degrees of freedom pose of the truck, such a system would have to rely on detecting natural features or artificial markers that occur at known locations on the truck. A good survey of this type of feature-dependent pose estimation was written by Lepetit and Fua [14]. As stated earlier, altering the trucks is not desirable, which means natural features would have to be used for the pose estimation as demonstrated by Gordon and Lowe [15]. A natural feature-based method was rejected primarily due to concerns that the muddy, dusty mine environment would conspire with shadows to create many features unique to individual haul trucks, and worse, to have these features change over time.

2. A monocular camera used in combination with a grid of laser light (structured light) can also be used to determine depth. Most structured light applications thus far have been used in indoor situations, where the laser light does not have to overpower the sun. However, there has been some research towards using this type of system outdoors [16], [17]. What is demonstrated by these papers, as far as structured light for this application is concerned, is that no high-speed eye-safe grid laser scanner with sufficient range under sunlight has yet been developed.

3. Millimeter-wave radar has been successfully employed outdoors in the mining industry for environment monitoring and mapping [18], [19], [20]. The GHz radiation used in these devices has the desirable property of being able to better image through dust and fog than the frequencies used by lasers or cameras [21]. However, mm-wave radar demonstrated to date is not fast enough for a high resolution real-time application. The custom built high speed unit developed by Widzyk-Capehart et al. attained a data rate of 173 data points/second [18], which does not deliver the necessary resolution quickly enough for this real-time application.

4. As described in Section 4.2.1, time-of-flight based laser scanners have thus far been the sensor of choice for excavator-mounted truck pose estimation systems [4], [5], [6]. However, 3D laser scanners \( (r, \theta, \phi) \) require relatively long data acquisition periods [4], and faster 2D laser scanners \( (r, \theta) \) [5], are best suited for determining only three degrees of freedom. For these reasons, laser scanners were not viewed as an optimal sensor for this real-time six degrees of freedom application.

5. Stereo cameras have also been deployed in the mining industry [22], [23], although less frequently than either laser or radar scanners. Stereo camera data is inherently 3D
and does not degrade in the presence of shadows or shifting dirt patterns, as might monocular camera systems. For an overview of how stereo cameras compute 3D positions, see [24], pp. 240-246. Compared to other inherently 3D sources (radar and laser scanners), stereo cameras offer several advantages:

- **Speed**: The time to acquire the raw data is constrained only by the shutter speed of the camera system.

- **Parallelism**: All data points are acquired simultaneously and not via a point-by-point procedure, ensuring all data points represent the same moment in time.

- **Data Density**: By using megapixel cameras, a stereo image can contain more than a million data points. Laser scanners and millimeter wave radar systems are to date incapable of delivering the same resolution.

- **Cost**: Digital cameras are ubiquitous and inexpensive.

- **Reliability**: Cameras with no moving parts can be employed, which decreases susceptibility to damage from shock and vibration.

However, stereo cameras require a visible scene texture to determine depth, and require an initial calibration. These two shortcomings are not generally a problem as the mining environment is rich in texture and the initial calibration, once complete, should not have to be repeated. Of potentially greater concern is that cameras suffer difficulties when imaging through atmospheric disturbances such as thick dust, fog or heavy rain or snow. This means that there may be times when the system is unable to operate, but a human operator also requires sufficiently visible scene elements to control the machine safely and effectively. Finally, perhaps the greatest weakness of stereo cameras is that their depth accuracy falls off exponentially with the distance of the object being measured. However, by using a camera with a sufficiently wide baseline and high-resolution sensors, it is possible to obtain the depth accuracy needed for this application’s working distance.

### 4.3.2 Load Profiling

An added benefit of using a stereo camera for truck pose estimation is that it is also highly suitable for load profiling. The same stereo-camera strengths that apply for truck pose estimation apply for load profiling: the data is inherently 3D, it is dense and it is acquired essentially instantaneously. An additional advantage when applied to load profiling is that as the load profile information is to help the operator make better decisions, the fact that normal camera images are available for data visualization makes
the sensor especially well suited. Furthermore, while the previously stated weaknesses (reliance on texture to determine depth, exponential falloff in depth accuracy, etc.) still apply, none of them preclude using a stereo camera for load profiling.

An additional advantage to using the same stereo camera for load profiling is not the sensor itself but rather its location. By being placed on the shovel, it offers load profiling on a per-bucket basis, with live feedback, which allows the shovel operator to make corrections as the truck is filled. This kind of operator feedback and opportunity for corrective measures is not present in any existing truck load profiling system, such as those discussed in Section 4.2.3.

However, the positioning of the camera presents challenges. This is because the same camera is used for both haul truck pose estimation and load profiling, and the optimal position for one may not be optimal for the other. The position of the camera, discussed in Section 4.4, provides an isometric view of the haul truck load. While the camera can see most of the payload, some of it is not visible due to occlusion by the truck’s side wall and due to the load occluding itself. This necessitates a degree of data interpolation, however the overall utility of the system will not be affected.

Lastly, a limitation of using any optical measuring device for load profiling is that the sensor delivers the structure of the surface of the load, which makes it perfect for determining the volume and (assuming constant density) the centre of mass, but it does not directly measure the mass. Mass must be inferred by combining the volume of the load with its density, and although the density of the material will be known approximately, it will change depending on its composition, its fragmentation and its water content. A stereo camera is therefore well suited for determining load volume and distribution, and its measure of mass (by way of assumed material density) will be approximate.

### 4.4 Hardware Implementation

The stereo camera selected for the system is the commercially available Bumblebee XB3 produced by Point Grey Research Inc. It was chosen for its 24 cm wide baseline, its high resolution $1280 \times 960$ CCD sensors and its wide $66^\circ$ horizontal field of view. The camera was mounted high on the side of the shovel’s boom, as shown in Figure 4.2. This position was chosen to give the camera a high perspective of the haul truck and to keep it as far as possible from sources of dust or potential damage.

Since the camera is not constructed for outdoor industrial use, it needed to be protected from the environment. A waterproof, shock dampened, heated enclosure was designed and constructed for the camera. Figure 4.3 shows the camera mounted on a shovel boom with the camera installed.
4.4. Hardware Implementation

Figure 4.2: The location of the stereo camera on the side of the shovel boom.

Figure 4.3: The stereo camera installed in its protective mount. The camera is connected to a computer located in the shovel housing by wires run through waterproof conduit.
4.5 Software Design

The software system is responsible for using the stereo camera data to both determine the pose of the truck and the profile of its load. The truck pose estimation and load profiling subsystems are detailed in Sections 4.5.1 and 4.5.2 respectively.

4.5.1 Haul Truck Pose Estimation

To determine the relative position of the haul truck, the haul truck pose estimation subsystem must determine all six degree of freedom parameters that define it’s unique location and orientation.

The data produced by the stereo camera is processed into real-world $x$, $y$ and $z$ co-ordinates. This data is sometimes referred to as 2.5D, because it represents 3D data projected onto a two dimensional plane (the image sensor) from a single perspective, and so the scene data is incomplete. The collection of $(x, y, z)$ data points, collectively called a point cloud, is used to determine the truck position. This is accomplished by positioning a second model point cloud representing a model of the haul truck directly onto the scene point cloud generated by the stereo camera. Since the dimensions and original orientation of the model cloud are precisely known, once it is aligned or registered with the haul truck in the scene cloud, it can be inferred that the model haul truck cloud position is the same as the scene’s haul truck position.

The standard method for registering two point clouds with high accuracy is the iterative closest point (ICP) method, first introduced by Besl and McKay [25]. A good review of the ICP procedure and techniques for increasing the algorithm’s speed was written by Rusinkiewicz and Levoy [26]. A more recent analysis, comparing ICP to other procedures for fine point cloud registration, showed that few other methods have been developed and that ICP is still the basis for the fastest methods [27]. ICP was thus selected as the base algorithm to use to align the scene and model point clouds.

ICP Overview

As its name suggests, ICP registers two clouds in an iterative fashion. It does so by matching points from one cloud to their closest points in the other, and minimizing an error metric based on these pairs at each iteration. The essential steps taken at each iteration are: 1. select points from the model cloud, 2. match each point to its nearest neighbor in the scene cloud, 3. optionally weight and/or reject point pairs, 4. calculate an error, and 5. minimize the error. Variants optimizing one or more of the steps for speed and robustness have been developed, and one can refer to [26] and [27] for an overview. The ICP variant that has been implemented for this application uses
a point-to-plane error metric [28], [26] and a new accelerated point-matching method, described next. It does not perform any form of tracking on the truck, treating each frame completely independently.

The point-matching step is often a target for improving the overall speed. This is because unlike the other steps, a “naïve” approach runs in $O(N_M N_S)$ time, where $N_M$ is the number of model points, and $N_S$ is the number of scene points. The first measure taken to reduce the runtime of the search is to simply reduce the number of model points and scene points involved. The number of model points, 1661, was selected to deliver high accuracy while maintaining the necessary processing speed. The number of scene points is reduced by removing all the non-truck points from the scene cloud (see Figure 4.4), which increases both speed and accuracy. The pruning of the scene cloud is accomplished by only including points that are between 3 and 9 meters off the ground, and closer than 30 meters to the shovel.

![Figure 4.4: The scene cloud before and after truck isolation.](image)

The second and more notable strategy is to search the scene cloud in a more efficient manner. This is a classic optimization problem known as the nearest neighbor search (NNS). A now standard acceleration method is to structure the scene data into a k-d tree [29] as first demonstrated in the ICP context by Zhang [30]. A k-d tree is attractive because it offers a fast expected search time of $O(\log N_S)$ [31]. Although studies have
focused on improving this search time even further [32],[33], they all require an initial tree be built, an operation that (for 3D data) requires a minimum $3N_S \log N_S$ tree-building steps [31]. In a realtime application where new scene clouds are generated in rapid succession, the time to build the initial tree for each cloud is in fact the bottleneck which may prevent the system from running in realtime. Other search acceleration methods such as [34] and [35] have been developed but they too rely on a pre-processing step which effectively rules out their use in this application due to time constraints. A technique called closest-point caching presented by Simon et al. does not require pre-processing, but was shown to speed up the searches by only approximately 10% [36]. Another approach is to perform reverse calibration [37] which is very fast ($O(n)$) and requires no pre-processing. However, reverse calibration does not truly attempt to find nearest neighbors and as a result does not produce as accurate a registration as do nearest neighbor algorithms.

In [38], Jost and Hugi present a method that takes advantage of the hypothesis that points close to each other in the first cloud will be matched to points that are close to each other in the second cloud. This method has a best-case time complexity of $O(N_M)$ and (if no pre-processing is done) a worst case complexity of $O(N_MN_S)$. They also suggest a multi-resolution approach where the number of matched points increases as the registration error decreases so that each iteration can be made with as little computational effort as possible [39]. Implementing this technique improved the performance while maintaining registration accuracy, and so this technique is used in conjunction with the method described below.

**Resolution-Independent Nearest Neighbour Search**

The search method used by this system has proved efficient, runs in guaranteed time and requires no pre-processing. The method is motivated by the observation that nearest neighbour search strategies typically search (either directly, or as with a tree, indirectly) every available data point. This, however, is not an intuitive or efficient method when the data is somewhat organized, as shown in Figure 4.5. The problem is that the sampling frequency of a typical search method is ruled by the resolution of the sensor that created the data, not by the real-world characteristics of the data itself. However, if the search strategy could be changed so that the nature of the data is considered, then the search could potentially be performed much more quickly.

Consider the structure of a stereo camera’s 2.5D data. Each camera pixel $(i, j)$ maps directly (via the lens’ projection function) to a single $(x, y, z)$ point in space. The vector connecting the $(x, y, z)$ point in space to the $(i, j)$ sensor pixel is the path travelled by the light ray between a surface in the world and the camera’s 2D sensor. By considering
Figure 4.5: In the above subfigures, each line dash represents a data point. Consider, for both subfigures, the task of manually finding the closest data point on the dotted line to the $\times$. For either the low resolution data (a) or the high resolution data (b), a human would require about the same amount of time to complete the task. This is because a person will immediately dismiss most of the data, and concentrate solely on determining which data point of the large feature near the $\times$ is the closest. The resolution independent search strategy adopts this approach.
the pinhole projection,

\[
\begin{pmatrix}
    u \\
    v
\end{pmatrix} = -\frac{f}{z} \begin{pmatrix}
    x \\
    y
\end{pmatrix}
\]  

(4.1)

it can be seen that \((x, y, z)\) locations that are close to each other in space will project to 2D points \((i, j)\) that are close to each other on the sensor. (The position \((u, v)\) in meters corresponds to the pixel position \((i, j)\) in the image.) Likewise, so long as the imaged world is sufficiently smooth, 2D points \((i, j)\) that are close to each other on the sensor are likely to have been projected from \((x, y, z)\) locations that are close to each other in space. Consequently, by assuming that the 3D point \((x_1, y_1, z_1)\) projected to pixel \((i, j)\) is similar in location to the 3D \((x_2, y_2, z_2)\) point projected to pixel \((i \pm 1, j \pm 1)\), the \((i, j)\) pixel location may be treated as an index from which to construct a coarse-to-fine search strategy.

To understand the coarse-to-fine search method presented here, consider a single search where one model point is to be matched to the closest scene point. An exhaustive search would measure the distance (defined in the Euclidian sense,

\[
\sqrt{(x_s - x_m)^2 + (y_s - y_m)^2 + (z_s - z_m)^2}
\]

from the model point to every scene point, storing the scene point that produced the minimum distance. However the coarse-to-fine method works by first sampling, for example, every 20th scene pixel \((i, j)\) along each axis of the 2D projection, storing the closest 3D point \((x, y, z)\). A 2D projection space search window is then formed around the closest point, and that area is sampled more finely. This process is then repeated, producing as many search “layers” as needed, until a single pixel is reached. Figure 4.6 presents a visual comparison between an exhaustive search and a three-layer coarse-to-fine search, and Algorithm 2 shows a two-layer search expressed as pseudo code. The strategy allows the search to proceed at low resolution speeds, while maintaining the accuracy afforded by high resolution data.

From an intuitive perspective, the method works by using the first, coarsest search layer to locate the nearest “large feature” in the scene, with the subsequent search layers locating the actual closest point within the large feature. It is the first search layer’s sampling frequency that in essence defines how smooth the world is assumed to be. As shown next, if the frequency is chosen based on the physical attributes of the scene, rather than the resolution of the camera sensor, the speed of the search may be largely decoupled from the resolution of the camera sensor.

More formally, the coarsest sampling rate (pixels) is bound by the fact that it must be chosen such that the projection onto each sensor axis of the smallest detectable feature area of the sampled object be larger than the separation between samples. This requirement is in some respects analogous to the Nyquist-Shannon sampling theorem: the sampling must be sufficiently frequent, or the true structure of the data cannot be
Figure 4.6: The left image is a Euclidean distance map showing the distance between a single model point (shown as a $\times$ positioned directly above the image plane in the left image) and a haul truck (lighter is closer, darker is further). An exhaustive search would measure the distance between the model point and every truck point, effectively creating a distance map like this for every model point. In the coarse-to-fine method, shown on the right, only a subset of points are measured. The different sampling rates of the three search layers of the coarse-to-fine search are clearly visible.

Algorithm 2 Find Nearest Neighbour

```plaintext
nearest_neighbour ← VERY_DISTANT_POINT
{Start Layer 1}
for every 8th projected point per axis do
    if point is closer than nearest_neighbour then
        nearest_neighbour ← point
    end if
end for
{End Layer 1}
{Start Layer 2}
for every point in a $9 \times 9$ window centered at nearest_neighbour do
    if point is closer than nearest_neighbour then
        nearest_neighbour ← point
    end if
end for
{End Layer 2}
return nearest_neighbour
```
determined. Using the pinhole camera model (Equation 4.1), this bound on the sampling rate may be expressed as (derivation in Appendix A.1):

\[
\alpha = \frac{rp\cos(45^\circ) \cos(\theta_x) \cos(\theta_y)}{2d \tan\left(\frac{\text{FOV}}{2}\right)}
\]  

(4.2)

where \(\alpha\) (pixels) is the coarsest per-axis sampling rate in projection space, \(r\) (pixels) is the linear resolution of the camera, \(p\) (m) is length in the scene of the smallest detectable feature, \(\theta_x\) and \(\theta_y\) (degrees) are the maximum angles of rotation of the patch around the \(x\) and \(y\) axes relative to the camera, \(d\) (m) is the maximum distance between the patch and the camera, and \(\text{FOV}\) (degrees) is the angular field of view of the camera. The \(\cos(45^\circ)\) term ensures that the patch will be found no matter how it is rotated about the \(z\) axis. The desired level of sampling subsequent layers can also be determined with Equation 4.2 (by using smaller values for \(p\)), or by simply selecting a sampling level where all layers will visit the same number of total points.

This strategy essentially de-couples the speed of the search from the resolution of the input data, making it instead dependant on the physical properties of the object being sampled. The speed increase from this approach over the standard approach of building and using a k-d tree can be considerable, especially within search parameters typical of ICP applications. However, their relationship is complex and one strategy is not uniquely faster than the other. Below is an analysis of the relative speeds of both approaches.

To analyze the presented method, it is assumed that all layers after the first layer sample the same number of points. The total number of points sampled (or touched) is (see Appendix A.2):

\[
iN_M \left[ \frac{N_S}{\sigma^2} + L \alpha^{2/L} \right]
\]  

(4.3)

On the other hand, the theoretical minimum number of operations required to build the three-dimensional k-d tree is \([31]\):

\[
3N_S \log N_S
\]

and the average number of operations to search (once per model point per iteration) is \([31]\):

\[
iN_M \log N_S
\]
giving the total number of operations to build and search as:

\[ \log N_S(3N_S + iN_M) \]  \hspace{1cm} (4.4)

In all the formulas, \( i \) is the number of iterations, \( N_M \) is the number of model points connected to scene points, \( N_S \) is the number of scene points, \( \alpha \) is the result of Equation 4.2, and \( L \) is the number of search layers used after the first search layer.

To compare the methods, the base parameters for determining the coarsest search layer in the presented algorithm (Equation 4.2, or \( \alpha \)) were set to those which are appropriate for this particular application, \( r = 1024 \) pixels, \( p = 3.2 \) m, \( \theta_x = 67^\circ \), \( \theta_y = 23^\circ \), \( d = 16 \) m, \( \text{FOV} = 66^\circ \) and \( L = 3 \). The length of the smallest feature, \( p \), was set to the shortest dimension of the truck’s canopy plane, which is about 3.2 m, and the resolution of the camera was set to 1024 \( \times \) 768 (rather than the maximum 1280 \( \times \) 960) to meet speed requirements. The number of points in the search space, i.e., the number of points comprising the truck in the scene cloud (\( N_S \)), was set to 25\% of the total scene cloud points, the number of connected model points (\( N_M \)) was set to 1661 and the number of iterations (\( i \)) required for convergence was 40. A single parameter was then varied and the speed of the presented method relative to building and searching a k-d tree was recorded. The results are shown in Figure 4.7.

These results reveal that the method presented here is fastest with objects that have large, relatively smooth features, with high-resolution scene data, with fewer matched points and with a fast convergence. Note that in common with other approximation methods (e.g. [38], [32]), this method does not absolutely guarantee that the nearest neighbor will be returned. The results, however, for this application are very good and the execution time, which is guaranteed, is reduced in time complexity to \( O(n \log n) \) with no need for pre-processing.

**ICP Model Creation**

The design of the haul truck model point cloud is important for accurate cloud registration, and therefore haul truck pose estimation. A simplified haul truck body, shown in Figure 4.8, is comprised of five planes: two side walls, a bottom, a front wall, and a protective canopy.

During the loading process, the bottom plane, right wall and front wall (planes 3, 2 and 4) all become obscured by the material placed into the haul truck. Therefore, they are not consistently visible features of the truck and cannot be relied upon for determining its pose. This leaves two planes that do remain consistently visible, the left side wall and the canopy, which are used to form the ICP model.

The base information used to create the model is also important. Using a computer
4.5. Software Design

(a) Size of the minimum guaranteed detectable feature. (Equation 4.2, parameter $p$)

(b) Number of search layers used after initial layer. (Equation 4.3, parameter $L$)

(c) Total points matched between model and scene. (Equation 4.3, parameter $N_M$)

(d) Linear resolution of camera. (Equation 4.2, parameter $r$; will also change Equation 4.3, parameter $N_S$)

(e) Number of iterations to convergence. (Equation 4.3, parameter $i$)

Figure 4.7: The presented search method’s performance relative to building and searching a k-d tree. Each sub-figure represents altering one parameter from Equation 4.2 or 4.3. 100% signifies equal performance.
generated model cloud derived from the truck body’s design drawings would initially seem to be the best choice as it would be the most perfect copy of reality. While this may be true, the stereo camera which is used to generate the scene cloud does not produce a perfect copy of reality. Aside from noise (which is Gaussian [40]), there are consistent and predictable distortions in the stereo camera cloud, such as missing data due to occlusions and a smoothing effect over small sharp features. Since the goal is to match the model of the haul truck with the haul truck in the scene, a more accurate registration is possible by using a model truck that mimics the scene truck as closely as possible, including the stereo camera produced distortions. Therefore, instead of using a perfect computer generated model of the haul truck, a model was created from stereo camera data. An example of the truck model can be seen in Figure 4.9.

A complication of using a stereo camera sourced haul truck model is determining the model’s initial six degrees of freedom pose relative to the stereo camera. While this information is inherently present in a drawing-sourced computer model, it is missing from one created from real-world data because the data contains no reference point. This camera-centric initial position is absolutely necessary because the model truck will be used to find the position of a scene truck by measuring the change in position that the model must undergo to achieve registration with the scene. As the model’s initial position is the reference point to which all subsequent positions are measured, the initial position must be in terms of the camera (and thereby the shovel).

Clearly one cannot use the described ICP method to determine the truck’s position at the time of model data acquisition, as the model and the scene are in the same position and there exists no external reference point. Another method to determine this initial position must be employed. One such method would be to physically measure the relative positions of the camera and the truck during model data acquisition using standard
Figure 4.9: The haul truck used for model creation. The first image shows the haul truck normally, the 2nd shows the portion that will be used as the model cloud, and the 3rd shows the model cloud in isolation.

physical surveying techniques. This method would be accurate but also dependent on extra specialized equipment as well as being time consuming.

Another approach, and the one used here, is to determine the initial position directly from the same data used to create the initial stereo model. This can be done using an alternative pose estimation technique that does not depend on 3D data but rather solely on its projection, i.e. a simple 2D image as delivered by one of the two stereo camera imagers. The method, known as the Perspective-n-Point, or PnP method, consists of identifying $n$ points on an object in a 2D image, and by using knowledge of their relative 3D positions, determines where the object is in space in order to produce the observed projection [41]. This pose determination concept is much the same as closing one eye, yet still having a sense of depth due to knowledge of an object’s actual dimensions and its apparent size within one’s field of view.

In order to guarantee a unique pose solution using PnP, the number of identified points $n$ must be greater than or equal to six ($n \geq 6$), or there must be at least four co-planar points of which no three are co-linear [42]. The identification of the points in the 2D image need not be automated as this procedure is only performed once when the model cloud is created. An iterative Gauss-Newton approach to solving the PnP problem was used here as it has been shown to be among the most accurate methods [43].

The accuracy of a PnP solution depends on the use of a calibrated camera, that is a camera of a known focal length that produces images as if projected through a pinhole (i.e. a rectified image, one with no lens distortion). The process of precisely determining
a camera’s focal length and removing its lens distortion, known as camera calibration, can be achieved using a method such as the one proposed by Tsai [44]; however, the stereo camera used here was calibrated by the manufacturer. The rectification of the image was verified by performing PnP using a large number of points and confirming that the residual errors of the PnP pose for all points were less than one pixel. Similarly, the focal length was verified by measuring the rectified image’s field of view at a known distance, and confirming that it too was accurate to within one pixel. This verification assures that error in the system caused by any imperfections in the calibration will be minimal when compared to other sources of error.

The accuracy of the PnP pose estimation system is limited by the accuracy of the localization of the known points in the image, and the accuracy of their real-world 3D relationships. Their 3D relationships were provided with millimeter accuracy by the haul truck body’s manufacturer. Precisely locating the known haul truck points, shown in Figure 4.10, is somewhat less precise as they are not sharp high-contrast corners. This means that unlike the checkerboard corners used in [44], they cannot be found to sub-pixel accuracy, and consequently lead to multiple but very similar solutions. Therefore, multiple solutions for the haul truck were found, with the pose containing the median rotation being selected as the best possible pose. (A median was chosen over the mean because as all parameters for a single pose are determined in conjunction with one another, it was reasoned that the integrity of a pose could be compromised by averaging multiple poses’ individual components.)

As this is an alternative, independent method of determining a haul truck’s pose, it will also be used to determine the ground truth to verify the pose results generated by the presented ICP-based method. The accuracy to which the pose can be determined with PnP is presented in Section 4.6.4.

Using PnP to determine pose is a general technique, and so it is applicable in any situation where ICP is used and a model based on sensor data is desired, or, similarly, to determine the ground-truth of an ICP derived pose estimation.

**Initial Position of the ICP Model**

While traditional ICP methodology has been proven to converge, there is no guarantee that it will converge to the global minimum. The minimum that the algorithm converges to is largely dependent on the initial pose of the model, where the closer the initial pose is to the global minimum, the greater the likelihood that it will converge to it. Many coarse, initial alignment techniques have been developed [27], but a sophisticated one is not necessary for this application as the haul truck presents itself in roughly the same initial position each time. This is taken advantage of by using a method that is basic
but effective and efficient.

As the haul truck always presents itself with approximately the same rotation (±10°), the initial rotation is set as a constant. The model cloud’s initial translation is calculated by determining the centre of mass of the truck in the scene cloud, and offsetting the model cloud by a constant distance from it. Experiments have shown this to be a reliable method for achieving convergence to the global minimum.

4.5.2 Haul Truck Payload Volume, Distribution and Mass

As with truck pose estimation, the task of load profiling was accomplished with the 2.5D data provided by the stereo camera. Before load profiling can be done it is essential to have the truck accurately located, for it is the truck that forms the reference frame in which the payload must be measured. Once this reference frame has been established, profiling the load is, simply put, measuring how the observed surface of the truck’s payload area differs from where the known bed of the truck lies. This information will reveal the volume, the distribution, and, with density information, the mass of the contents of the truck.

Conceptually, the method to determine the load volume (and distribution) is to simply calculate the average data point height of the observed surface above the known location of the truck bottom. However the process is complicated somewhat by the effects of projection and perspective foreshortening, effects which result in surfaces that
are closer to the camera or orthogonal to its direction of view having a disproportionate number of data points compared to surfaces that are further away or presented at an angle. For this reason, the payload area is divided into a $3 \times 3$ grid, with the volume of each element determined separately, and then the volumes of these elements summed to arrive at the total volume. As the surface of the material in the truck is mostly smooth, the effects of projection-derived data and perspective foreshortening are effectively eliminated.

Another complication is that due to camera’s perspective of the truck’s payload, it is not able to accurately image the entire contents. This is because parts of the payload are occluded, and that the stereo processing becomes unreliable where the material creates a sharp angle against the far side wall. For this reason, only the middle 90% of the contents is measured, with the unavailable data being interpolated from the observed surface.

Finally, the results of the payload profiling are not entirely numerical as the analysis will also be presented in a form which is easily used by the shovel operator to improve the quality of the loading. For this, a color-height encoded relief surface with contour lines is created for display purposes. Additionally, the ideal centre of mass and the current centre of mass of the truck is shown, so that they may be compared to improve the loading process.

4.6 Experimental Evaluation

4.6.1 Experimental Setup

The system was installed on an production P&H 2800 XPB electric mining shovel. Stereo image pairs were recorded at 10 Hz as the shovel loaded two haul trucks. Qualitative and quantitative results are presented and discussed in the following sub-sections.

4.6.2 Qualitative Results

The most immediate measure of the system’s performance is its visual feedback: does the model haul truck appear to be correctly registered with the scene haul truck? Figure 4.11 shows a wireframe of the haul truck positioned on a scene truck that is empty, partially and mostly full. In Figure 4.12, three stages of ICP convergence are shown: the initial position, the position after one iteration and after convergence. The initial results look accurate, and are investigated numerically in Section 4.6.4.
Figure 4.11: A wireframe representation overlaid on the scene cloud’s haul truck. The haul truck is shown carrying three different payloads.

Figure 4.12: The model cloud converging on the scene cloud. The lines connecting the model cloud and the scene cloud represent a few of the points matched during the nearest neighbour search.
4.6.3 Accelerated Search Evaluation

The accelerated ICP nearest neighbour search method was evaluated by comparing it to an exhaustive nearest neighbour search at multiple sampling resolutions. The results are presented in Table 4.1, where the \((a, b, c)\) triplet in column 1 represents the pixel sampling frequency on both projected axes for each successive search layer. Differences in computation time and final position have been recorded, showing that coarser sampling leads to faster performance but also an increased discrepancy in the final position. Note though, that these differences in the final position are minimal, especially when considered on a percentage basis.
<table>
<thead>
<tr>
<th>Search Operations Per Iteration</th>
<th>Total Time (s)</th>
<th>Iterations</th>
<th>Time per Iteration (s)</th>
<th>Relative Time Taken Per Iteration</th>
<th>Change in Translation (m)</th>
<th>Change in Rotation (°)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exhaustive</td>
<td>282,345,085</td>
<td>546.415</td>
<td>43</td>
<td>12.71</td>
<td>-.-</td>
<td>-.-</td>
</tr>
<tr>
<td>Fast Search (5,2,1)</td>
<td>11,379,154</td>
<td>18.42</td>
<td>40</td>
<td>0.46</td>
<td>0.0362</td>
<td>0.010</td>
</tr>
<tr>
<td>Fast Search (20,3,1)</td>
<td>896,659</td>
<td>1.667</td>
<td>39</td>
<td>0.04</td>
<td>0.0034</td>
<td>0.043</td>
</tr>
<tr>
<td>Fast Search (40,3,1)</td>
<td>530,099</td>
<td>1.14</td>
<td>42</td>
<td>0.03</td>
<td>0.0021</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Table 4.1: Comparison of an exhaustive search and resolution independent nearest neighbour search method.
4.6.4 Error Analysis

Stereo Camera Error

Stereo camera error is difficult to characterize because the error is dependent on the accuracy of the correlation used in the stereo processing, which itself is dependent on the texture and geometry of the scene being imaged. As texture and geometry are different for virtually every data point, so is the uncertainty. However, it is still informative to have a notion of the approximate uncertainty for a point at a given working distance. The manufacturer of the stereo camera used in this application states that the correlation is usually accurate to within 0.1 pixels (one standard deviation).

The error in the $z$ direction is determined by (see Appendix A.3):

$$\Delta z = \frac{z^2}{fb} \varepsilon$$

(4.5)

where $f$ is the focal length of the camera, $b$ is the stereo baseline, and $\varepsilon$ is the correlation error. At distances of 12 m and 16 m, with a camera resolution of $1024 \times 768$, the uncertainties in $z$ are $\pm 0.076$ m and $\pm 0.135$ m respectively. See Figure 4.13 for a graph of depth uncertainties. The $(x, y)$ translational uncertainty is provided by the manufacturer, is directly proportional to the depth ($z$) value of the data point, and is far less significant than the uncertainty in $z$. At distances of 12 m and 16 m, the $(x, y)$ uncertainties are $\pm 0.005$ m and $\pm 0.007$ m respectively.

Figure 4.13: The uncertainty in the stereo camera’s depth measurement at $1024 \times 768$.  

65
ICP Residual Lack of Fit Error

A second source of error exists in the fit provided by the ICP image registration process. The implemented point-to-plane ICP (described in detail in [28]) operates by minimizing the sum of the square of the distances between the planes defined at model cloud points that have been matched to scene cloud points. The RMS error is the error in fit between the model and image clouds, which is present due to errors in the stereo camera data and in imaging the truck from different perspectives. Testing showed that the RMS error at 15 m distances typically lie between 0.05 and 0.1 m, or about 0.5% of the total distance.

Ground Truth Consistency Error

The internal matching error reported by ICP is not the true error—it is a measure of how well the two point clouds have been registered, but not how well the model cloud has been matched to ground truth. As explained in Section 4.5.1 (ICP Model Creation), Perspective-n-Point (PnP) can be used to estimate the ground truth by manually identifying five known truck points in a 2D picture pose (Figure 4.10). However, as stated in Section 4.5.1, the known five points cannot be located at sub-pixel accuracy, which leads to a range of possible ground truths. To estimate the ground truth for a particular picture, PnP pose estimation was performed seven times. For the same reasons as those stated in Section 4.5.1, the pose with the median rotation angle for the seven estimations was selected as ground truth. It should be noted that the differences for a single pose using seven rounds of PnP were small. For every different picture, the standard deviation of its seven possible ground truth poses was calculated, with the mean standard deviation shown in Table 4.2. This gives an indication of the accuracy of a ground truth estimate.

<table>
<thead>
<tr>
<th>PnP</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Standard Deviation of Translation (m)</td>
<td>0.012</td>
</tr>
<tr>
<td>Mean Standard Deviation of Rotation (degrees)</td>
<td>0.037</td>
</tr>
</tbody>
</table>

Table 4.2: Uncertainty in ground truth determination using PnP

Registration Errors

With ground truth established, ICP-based pose estimation was performed, with the resultant pose compared to ground truth. Both exhaustive and fast nearest neighbour search were used for comparison purposes. The results are shown in Table 4.3, and again
4.6. Experimental Evaluation

the fast search’s \((a, b, c)\) triplet represents the sampling frequency in pixels at the three successive search layers.

<table>
<thead>
<tr>
<th></th>
<th>Exhaustive</th>
<th>Fast Search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(12, 3, 1)</td>
<td>(12, 3, 1)</td>
</tr>
<tr>
<td>Mean Translation Error (m)</td>
<td>0.062</td>
<td>0.072</td>
</tr>
<tr>
<td>Std. Dev. of Translation Error (m)</td>
<td>0.013</td>
<td>0.021</td>
</tr>
<tr>
<td>Maximum Translation Error (m)</td>
<td>0.079</td>
<td>0.090</td>
</tr>
<tr>
<td>Mean Rotation Error (degrees)</td>
<td>0.27</td>
<td>0.27</td>
</tr>
<tr>
<td>Std. Dev. of Rotation Error (degrees)</td>
<td>0.24</td>
<td>0.18</td>
</tr>
<tr>
<td>Maximum Rotation Error (degrees)</td>
<td>0.58</td>
<td>0.51</td>
</tr>
</tbody>
</table>

Table 4.3: Error in ICP pose estimation

It is also instructive to measure what the error is at the outermost (extreme) points of the haul truck, where a shovel-truck collision is most likely to occur. Using the fast nearest neighbour search with ICP, the mean error resulting from translation and rotation errors at the very rear of the truck’s bed was 0.11 m with a standard deviation of 0.04 m. The worst case observed error was 0.16 m. This maximum error can be used to define the exclusion zone for the shovel bucket control system.

4.6.5 Processing Speed

Because the system is required to operate in real-time, the speed at which the stereo data can be processed and the truck position determined is an important measure of performance. Implemented using a combination of C++ and (for the GUI) C# on an Intel T7700 2.4 GHz Core 2 Duo CPU with 667MHz DDR2 SDRAM, at a resolution of 1024×768, the system performed stereo processing at 5.3 Hz, and performed both stereo processing and pose estimation at 0.5 Hz. This pose estimation speed was sufficiently fast for an operational loading assistance system.

The system was single threaded, and as both stereo processing and ICP point-matching can be programmed in a parallel manner, the speed of the system should scale with the number of cores used. If programmed to run on a graphics processing unit (which are massively parallel), large performance improvements could be realized, where the performance bottleneck would likely stem from image acquisition speed rather than computational limitations.
4.6.6 Load Profiling

Load profiling can be performed following haul truck pose estimation, as shown in Figure 4.14. The contour lines represent even elevation gains, the white line represents the ideal centre of mass, and the dotted line represents the current centre of mass. A view similar to this could be presented to the shovel operator to aid in delivering the best possible payload profile.

![Image](image_url)

Figure 4.14: A view of a haul truck before and after load profiling. This load was measured to have a volume of 108 cubic meters, and a centre of mass (dotted line) that is 0.47 meters off centre (white line).

The absolute accuracy of profiling the load is difficult to assess, as no alternate load profiling system or ground truth data was available at the time this research was performed. However, it is possible to determine the expected accuracy of the system by considering the errors found in truck pose estimation.

Volume estimation was performed by measuring the payload’s surface, and calculating how far above the known location of the truck’s bottom that it lies. Error in the measurement of the surface itself results from errors in the stereo camera data points, and as there are many points and the error is Gaussianly distributed [40], error in the measurement of the surface should not be significant. However, the position of the bottom of the truck is determined by the truck pose estimation, so an error in pose estimation will lead directly to an error in volume estimation. As shown in Table 4.3, the mean truck translation error is 0.072 m, and the mean error in rotation is 0.27
degrees. To assess how this affects the volume results, an experiment was conducted where the haul truck pose, as found by ICP, was displaced in a random direction by the mean errors in translation and rotation, and the volume of the truck was re-calculated. This process was repeated 100 times per base volume, and the results are shown below in Table 4.4. For each base volume, the mean absolute error was less than 0.1 m$^3$.

<table>
<thead>
<tr>
<th>Base Volume (m$^3$)</th>
<th>Standard Deviation of Volume Errors After Displacement (m$^3$)</th>
<th>Standard Deviation of Volume Errors as a % of Total Volume</th>
<th>Absolute Maximum Error (m$^3$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32.8</td>
<td>1.3</td>
<td>4.0</td>
<td>2.6</td>
</tr>
<tr>
<td>60.3</td>
<td>1.4</td>
<td>2.3</td>
<td>2.5</td>
</tr>
<tr>
<td>108.0</td>
<td>1.5</td>
<td>1.4</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Table 4.4: The observed change in calculated payload volume by moving the truck’s position by the mean error, 100 times per base volume.

The percent errors present in the volume estimation of a fully loaded truck are slightly better than those found in modern weighscale or truck scale mass measurement systems (as given in Section 4.2.2), however a rigorous error analysis using an independent, validated volume estimation system would be necessary to conclusively determine the error. As implemented, the load volume and profile information requires 0.35 seconds to compute. It is possible, however, to implement it so that it can be computed in approximately 0.15 seconds.

### 4.7 Conclusions

A real-time stereo-vision based haul truck pose estimation system has been described. The system is to form a core component of a haul truck loading assistance apparatus that will help reduce injuries caused by shovel/truck collisions and dangerous loading errors. The haul truck pose estimation system was extended to also provide the volume and profile of the load carried by the haul truck. This is the first time this information has been provided by a shovel-based system, an advantage because it provides for the opportunity to actually correct loading errors if the occur. The hardware was designed to reside entirely on a mining shovel and be comprised of commercially available components to reduce the initial and ongoing costs. The system was demonstrated using data gathered from a full-size industrially operated shovel and haul truck.

The novel six degrees of freedom haul truck pose estimation system used a fast, resolution independent nearest neighbor search algorithm that operates without pre-
processing as part of an ICP approach. The search algorithm was compared to building and searching a k-d tree, and was found to require fewer computations for many situations typical in an ICP nearest neighbour search. This allowed the complete system to achieve a speed of 0.5 Hz, which meets industrial production requirements. The method used PnP to find the initial pose of a range-sensor derived model point cloud, and used the same technique to verify results given by ICP. To the best of our knowledge, this is the first time PnP has been used for these purposes. The experimental results showed a mean haul truck pose estimation error of 0.090 m (translation) and 0.27° (rotation). The maximum observed error was 0.16 m, which may be used to form the bucket-exclusion zone around the haul truck.

Load volume and profile estimation was demonstrated using the same stereo image data used in haul truck pose estimation. It was shown that the expected uncertainty in the volume calculation to be 1.4% of a fully loaded haul truck. This uncertainty compares favourably to existing mass estimation systems, however an independent measure of a trucks volume would need to be used to absolutely determine the accuracy of the measurement.
4.8 References


4.8. References


4.8. References


Chapter 5

Conclusion

5.1 Summary of Contributions, and Conclusions

This research contributes to two major goals in the mining industry: improving safety and efficiency. A method for improving safety by preventing dangerous shovel-truck interactions has been described. Similarly, a method to help a shovel operator optimally load a haul truck has been presented.

From an application standpoint, Chapter 2 introduced a complete hardware architecture that can be used to determine the relative positions of a haul truck and mining shovel. In particular, a stereo camera based system was determined to be the most appropriate sensor choice for determining the relative position of a mining shovel and haul truck. In Chapters 3 and 4, a fully operational truck pose estimation system built upon this hardware was presented. Chapter 4 described the importance of a haul truck’s payload volume and profile, and presented the first shovel-based system to determine these quantities as the truck is being loaded.

Chapter 4 assessed the speed and accuracy of the haul truck pose estimation, and found it to be sufficiently fast and accurate for industrial use. Specifically, the pose estimation operated at 0.5 Hz and had a worst case error of 0.16 m. This is a significant improvement over previous systems ([1], [2], [3]) as it offers the best reported accuracy and does not interfere with normal shovel operation. Chapter 4 also assessed the shovel-based haul truck load volume estimation and profiling system. Its expected uncertainty in volume estimation was shown to be 1.4% for a fully loaded haul truck, which if verified would be an improvement over truck and shovel-based mass estimation systems which report errors in the 2-3% range [4], [5].

It is worth noting that the discrepancy in processing speeds reported in Chapters 3 and 4 was due to the fact that Chapter 3 relied on the internally reported ICP measure of registration accuracy, and not the ground truth. It was found during the analysis reported in Section 4.6.4 that a greater number of points was necessary for truly accurate pose estimation. It should, however, be appreciated that in systems where fewer points are needed for accurate pose estimation, that the presented nearest neighbour search method would offer consistently and significantly greater performance than a k-d tree method (note the trend when using fewer matched points Figure 4.7(c) on page 56).
From a theoretical standpoint, Chapter 3 presented a novel, resolution-independent nearest neighbour search method for use with 2.5D range sensor data. The search method is used within ICP to allow high speed point cloud registration without the need to pre-process the data as required by previous work [6], [7]. Chapter 4 also detailed a method to use PnP to find the initial pose of a range-sensor derived point cloud, and used the same technique to verify pose results obtained with ICP. To the best of our knowledge, PnP has not previously been used for these purposes.

In conclusion, the goal of accurately determining the pose of a haul truck from a mining shovel was achieved. The stereo camera was shown to provide sufficiently accurate data on which to base such a system, and a new nearest-neighbour search method was used with ICP to provide the necessary speed.

### 5.2 Directions for Future Work

#### 5.2.1 Completing the Operator Loading Assistance System

A system to determine the pose of a haul truck was outlined in this work; however, to build an operator loading assistance system this work will have to be integrated into a larger system that includes knowledge of the shovel’s swing angle and arm geometry. Additionally, a collision prediction algorithm that considers all the information provided by the subsystems will have to be built. Figure 5.1 shows a block diagram with each subsystem, and what each subsystem relies upon to perform its function.

![Subsystem Diagram](image)

**Figure 5.1:** The operator loading assistance system’s subsystem hierarchy. The arrows point to blocks that are relied upon by the originating subsystem.

Integrating all of these systems will not be trivial as they will have to share the resources of a single computer, or use a communication system to relay information between separate computers. Once integrated, the system will need to be tested as a whole to ensure that all the components are working together as planned. Finally, a shovel operator interface will have to be designed and tested.
5.2. Directions for Future Work

5.2.2 Improving Haul Truck Pose Estimation and Load Profiling

There are a number of avenues that could be explored to improve the haul truck pose estimation and the load profiling. The two primary aspects that would likely be targets for improvement would be speed and accuracy.

**Improving Speed**

Increasing the speed would allow the system to process images at a higher frame rate which would be more pleasing from a visual perspective, or alternatively, use fewer system resources to allow for easier integration or the addition of new capabilities.

As noted in Sections 3.5 and 4.6.5, the presented system currently runs in a single thread. However, the most time consuming portions of the algorithm, processing the stereo data and finding the nearest neighbour in the point matching step of the ICP algorithm, are readily programmed in a parallel manner (e.g. [8]). By doing so, the system could run these time consuming algorithms faster by a factor equal to the number of processing cores available. If they were programmed to run on graphics hardware (e.g. [9]), which is highly parallel, great increases in speed could be expected.

Another avenue worth exploring is that of “low quality” k-d trees, that is k-d trees that do not provide the fastest search time but that are much faster to construct [10], [11]. As the construction time is the most time consuming element of using a k-d tree, the ability to construct one quickly would make up for any time penalty incurred during the search phase. It would also be beneficial to implement a k-d tree search method to conclusively determine its speed, rather than solely a statistical analysis of the number of operations performed (Section 4.5.1). This is because other factors such as memory access, caching and implementation-specific constants are not contained for either method in such an analysis.

**Improving Accuracy**

Increasing the accuracy would be beneficial, as an increase in accuracy would likely reduce any false alarms which may occur with an integrated system.

In Section 4.5.1, it is stated that stereo cameras do not produce a perfect copy of reality. Indeed, the copy that is produced is dependent on the exact geometry of the scene they are imaging. It was for this reason that stereo camera data was used to create the model for use in the ICP matching algorithm, as the highest degree of accuracy is achieved when the point clouds being registered are as identical as possible. With this in mind, the closer the scene haul truck is to the pose of the truck during model acquisition, the greater the expected accuracy of the pose determined by ICP. Therefore, if multiple models were available, each at a different translation and rotation, then as the model...
converged, the model being used could be switched to match as closely as possible the current pose in the scene. This method would likely increase accuracy, while using essentially zero additional computational power.

Another method that may offer an increase in accuracy is by employing some form of tracking, where a frame’s pose is checked for consistency with previous frame poses. This would be especially powerful if combined with information from a swing angle sensor, so that the position change of the camera is known. A Kalman filter may be particularly helpful in implementing a tracking system.

5.2.3 Further Capabilities

There are likely a number of new capabilities that could be added to the presented hardware and software system. Below are listed three which may be most useful.

1. **Bucket fill factor.** The volume of material contained per shovel bucket load is an important measure because it is one of the prime indicators as to how efficiently the shovel is being used. By measuring the truck volume on a once-per-bucket basis, the bucket fill-factor could be relayed back to the shovel operator. This would allow the operator to objectively measure how efficiently the available bucket volume is being used. Additionally, when the operator is digging the last load to be dumped into a truck, the system could recommend a fill factor to avoid over- (or under-) loading the truck.

2. **Recommended dump location.** If a load of material is dumped in an incorrect spot, the truck will not be centre-loaded. In order to correct this, the following bucket load will have to be dumped off-centre, but on the opposite side of the truck. The presented system detects when and by how much the haul truck is from being centre loaded, and by using models of how material will slump when deposited onto a slope, the system could recommend where the following bucket be dumped in order to re-establish a centrally loaded truck.

3. **Fragmentation analysis.** Fragmentation is a measure of the size distribution of the individual rock fragments that are produced by a blast. Fragmentation is an important measure for a mining operation because the fragmentation can have a great effect on the efficiency of the downstream ore processing. The stereo camera has a clear view of the contents of the truck, and could be used to measure the fragmentation.
5.3 References


Appendix A

Equations

A.1 Equation 4.2

To determine the coarsest sampling rate, the smallest feature length $p$ must be found in units of pixels when it is projected from a distance $d$ onto the image sensor. The camera is modeled as a pinhole camera with a linear resolution, $r$ (pixels) and field of view, FOV (degrees) (see Figure A.1). At a distance $d$, the camera’s linear field of view (m) is given by

$$2d \tan\left(\frac{\text{FOV}}{2}\right)$$

If the camera has a linear resolution $r$, then this field of view is projected onto $r$ pixels, where each pixel represents the same linear distance:

$$\frac{r}{2d \tan\left(\frac{\text{FOV}}{2}\right)}$$

Thus, if the previous equation represents the number of pixels per linear meter at distance $d$, the numbers of pixels spanned by $p$ will be

$$\frac{rp}{2d \tan\left(\frac{\text{FOV}}{2}\right)}$$

However, the previous equation assumes that $p$ is presented without any angular rotation. If $p$ is rotated away from the camera (rotated about the $x$ or $y$ axis), then its projected size length will shrink by a factor of the cosine of the rotation angle. Additionally, a rotation about $z$ changes the proportion of the feature that is projected in the combined $x$ or $y$ axis of the projection. As the coarsest regular sampling is to be determined, the smallest simultaneous projection in the $x$ and $y$ axes must be considered, which occurs when the angle of rotation about the $z$ axis is $45^\circ$. Because all of these rotations shrink the size of the projection, the final equation for $\alpha$ must be discounted by a factor of each:

$$\alpha = \frac{rp \cos(45^\circ) \cos(\theta_x) \cos(\theta_y)}{2d \tan\left(\frac{\text{FOV}}{2}\right)}$$
Figure A.1: A single slice from a pinhole camera projection. $p$ is the smallest image feature (m), $d$ is the $z$ distance from the patch to the pinhole (m), FOV is the camera’s field of view (degrees), $r$ is the sensor’s linear resolution (pixels), and $\alpha$ is the minimum sampling rate (pixels).
A.2 Equation 4.3

To find the closest scene point to a single model point, the number of search operations takes the form \( p + q \), where \( p \) is the number of operations for the first search layer and \( q \) is the number of operations for subsequent layers. For an exhaustive search, \( p + q \) would equal \( N_S \), the total number of points in the scene cloud. However for the coarse-to-fine search, the total number of searches is reduce by sub-sampling. The number of search operations in the first layer is a function of \( N_S \) and the sampling rate, \( \alpha \), determined in Equation 4.2. Since \( \alpha \) is the number of points skipped along both axes, the number of points searched in the first layer is reduced by a factor of \( \alpha \) twice, giving

\[
p = \frac{N_S}{\alpha^2}
\]

Any number of sampling strategies can be adopted for subsequent layers, but for the purposes of this analysis it is defined as follows:

- A search window is square, and the length of the sides are equal to the previous layer’s sampling rate;
- A search window is centered at the closest point found in the previous search layer;
- All search layers in \( q \) sample the same number of points;
- The final search layer does not sub-sample.

The number of search operations in \( q \) depends only on \( \alpha \), and the number of search layers it employs, \( L \). If only a single search layer is used, then it is also the last search layer and so is sampled exhaustively, performing \( \alpha^2 \) search operations. If more than one search layer is used, then as each layer must sample the same number of points, each layer must sample \( \sqrt[2/L]{\alpha} \) points. (This can be seen by reasoning that the length of the final search layer’s window, \( w \), when raised to the power of the number of layers, must equal \( \alpha \): \( w^L = \alpha \).) Since each of the \( L \) search layers searches the same number of points, and there are \( L \) layers, the total number of points search in \( q \) is

\[
q = L\alpha^{2/L}
\]

and therefore

\[
p + q = \frac{N_S}{\alpha^2} + L\alpha^{2/L}
\]

Because a nearest neighbour must be found for all considered model points at each iteration, the total number of search operations is:
\[ iN_M \left[ \frac{N_S}{\alpha^2} + L\alpha^2/L \right] \]

### A.3 Equation 4.5

For a rectified stereo camera with the cameras on the same plane, the depth of a point is determined by the distance separating the cameras \( b \) (m), their focal length \( f \) (m) and the disparity of the feature between the two cameras \( D \) (pixels):

\[ z = \frac{fb}{D} \]

Therefore the sensitivity of \( z \) to an error in \( D \) (caused by an error in correlation) is

\[ \frac{\delta z}{\delta D} = -\frac{fb}{D^2} \]

Since \( D = \frac{fb}{z} \),

\[ \delta z = -\frac{z^2}{fb} \delta D \]

or

\[ \delta z = -\frac{z^2}{fb} \varepsilon \]

where \( \varepsilon \) is the uncertainty in the correlation match. (The negative sign which does not appear in Equation 4.5 is due purely to the arbitrary coordinate system that is chosen.)