Positioning for Mobile NDE Inspection Robots

Rahul Summan

Department of Electronic and Electrical Engineering

University of Strathclyde

A thesis submitted for the degree of

Doctor of Philosophy

November 2012

Copyright

Date:

Signed:

I would like to dedicate this thesis to my beloved grandfather
Darshan Lal Bhatti.

Acknowledgements

Firstly, I would like to thank my supervisor Dr Gareth Pierce of the Centre for Ultrasonic Engineering for the opportunity to conduct this research and his support, inspiration and patience throughout. This research was supported by the Engineering and Physical Sciences Research Council and the Research Centre for Nondestructive Evaluation. I would also like to thank Gordon Dobie, Liam Mackenzie and Charles Macleod for their input throughout this work. Finally, I would like to thank my parents, brother and sister for their support and encouragement through my studies.

Abstract

Ageing infrastructure worldwide requires periodic inspection, often in-situ, in order to ensure continued safe and economic operations as well as adherence to stringent quality and performance requirements. In service automated Nondestructive Evaluation, where feasible, is highly attractive, and potentially allows inspection of operational plant. The use of such technology is very attractive in terms of safety, cost and the potential for minimal disruption to the inspection site especially if plant operations can remain online. Automated Nondestructive Evaluation in the form of remotely operated robotic vehicles is an active area of research. Knowledge of position relative to a frame of reference is a key aspect for a robotic Nondestructive Evaluation system in order to associate sensor measurements with locations on the structure being investigated.

This thesis investigates relative and absolute positioning techniques for a single robot. The accuracy and repeatability of a photogrammetry system is characterised over a large volume using a high accuracy metrology instrument. It was found that the photogrammtery system was most accurate in the centre of the volume and least accurate at the edges. This photogrammetry system is then used to evaluate the performance of algorithms developed in subsequent research. An image based positioning system is implemented which extracts motion information from a camera carried onboard a robot. The system is evaluated on surfaces typically found in industrial environments. Ultrasonic ranging techniques are investigated for robot positioning. In particular a low cost, modular, ultrasonic positioning system is characterised and calibrated. Bayesian filtering in the form of an Extended Kalman and Particle Filter are implemented to fuse noisy optical encoders estimates available at 100 Hz and the ultrasonic positioning measurements available at 3 Hz. The Extended Kalman Filter, at lower computational cost, was found to produce the lowest error.

Contents

List of Figures

List of Abbreviations

2D	Two Dimensional
3D	Three Dimensional
ADM	Absolute Distance Meter
BIRA	Bio-Inspired Range Finding Algorithm
CCD	Charge Coupled Device
DOG	Difference of Gaussians
EKF	Extended Kalman Filter
FIRST	Facility for Innovation and Research in Structural Testing
FM	Frequency Modulated
FOV	Field of View
GPS	Global Positioning System
GPU	Graphical Processing Unit
IFM	Interferometer
IID	Independent and Identically Distributed
IMU	Inertial Measurement Unit
KF	Kalman Filter

LAT	Leica Absolute Tracker
LED	Light Emitting Diodes
LFM	Linear Frequency Modulated
LIDAR	Light Detection And Ranging
MFL	Magnetic Flux Leakage
MSE	Mean Square Error
NDE	Non Destructive Evaluation
NN	Nearest Neighbour
PDF	Probability Distribution Function
PF	Particle Filter
PIG	Pipeline Inspection Gauge
PMF	Probability Mass Function
RANSAC	Random Sample Consensus
RRR	Red Ring Reflector
RSA	Remote Sensing Agent
RV	Random Variable
SIFT	Scale Invariant Feature Transform
SLAM	Simultaneous localisation and mapping
SMR	Spherically Mounted Reflector
SURF	Speeded Up Robust Features
SVD	Singular Value Decomposition

TDoA	Time Difference of Arrival
ToF	Time of Flight
UAV	Unmanned aerial vehicle
UKF	Unscented Kalman Filter
USB	Universal Serial Bus
VMCS	Vicon motion capture system
VO	Visual Odometry
VSK	Vicon Skeleton
WCS	World Coordinate System
XML	Extensible Markup Language

Chapter 1

Introduction

1.1 Nondestructive Evaluation

Ageing infrastructure worldwide requires periodic inspection, often in-situ, in order to ensure continued safe and economic operations as well as adherence to stringent quality and performance requirements. Many of the structural assets comprising this infrastructure have been subject to operating life extensions.

The Health and Safety Executive estimates that over 50% of offshore platforms resident in the UK Continental Shelf have exceeded their original design life [1]. A number of factors can be attributed to this extension namely the continual need for oil and gas coupled with improved low pressure oil extraction techniques which have allowed installations to remain operational. Clearly the harsh environmental conditions present offshore have a significant impact on the structural integrity of the rig and therefore it needs to be regularly inspected. Indeed the recent Deepwater Horizon disaster in the Gulf of Mexico in which eleven workers were killed and an estimated four million barrels spilled into the ocean serves to highlight the severity of accidents that can occur [2]. A similar case of prolonging operational lifetime is true of the nuclear industry. Fundamentally it is cheaper to implement lifetime extension than to construct a new plant; it is estimated that a new build costs in region of £600 billion while decommissioning and associated

disposal and waste treatment costs £250 billion [3]. A specific example of the extending lifetime occurred in 2008 when the Nuclear Decommissioning Authority (NDA) announced that Oldbury Nuclear power plant would continue operations into 2009 beyond its scheduled closure at the end of 2008 [4]. This action was taken to raise more decommissioning funds with the operating life extension considered more economically viable following investment in safety upgrades. This initial extension has been subsequently increased to 2012 [5].

The concept of lifetime extension depends critically upon the technology of Non Destructive Evaluation (NDE) also known as Non-Destructive Testing (NDT). The NDE of engineering structures is an important and challenging activity which can help to locate the presence and extent of structural defects before failure occurs. It is used to interrogate the structural integrity of engineering components without damaging the component itself. Inspection of critical components can thus reduce costly outages, negative environmental impact as well as potential loss of life. A range of non-invasive NDE techniques are available including ultrasonic, visual, electromagnetic and radiography which are used to detect and characterise flaws in terms of their nature, size and position [6]. Through identification of anomalies, NDE can be used to replace only those components determined to be defective and can thus contribute to the extension of the operational life of the structure. Additional complexity to the inspection task is brought about by the scale of the entity under consideration which varies by application domain: structures can range in size from, for example, an jet engine turbine blade to kilometres of pipeline in an oil refinery.

Industrial sectors for which NDE is of major importance include oil and gas, nuclear, automotive, aerospace and power generation. Such industries are a sources of particular difficulties, often presenting inspection sites occupying inaccessible areas or where environmental conditions are hazardous for human operators working at height, exposed to radioactivity and proximity to high temperature and/or pressure process plant. The financial impact of NDE inspections is also significant, arising from both the intrinsic inspection costs and the associated cost of taking plant offline to conduct inspections [7]. Given that inspection in industries such as automotive, aerospace and rail is considered imperative under strict

industry standards and regulations, it is no surprise that NDE is big business. Frost and Sullivan in a 2009 report [8] estimate that in 2008 the NDE market generated revenues of $1.1 billion with the major end users of NDE equipment. Indeed focus upon safety and the resultant legislation is driving the NDE market forward.

In service automated inspection where feasible, is highly attractive, and potentially allows inspection of operational plant. The safety, environmental and financial benefits for automating NDE measurements are clear, and applicable across a broad range of NDE technology. Automation is currently being addressed through deployment of sensor laden remotely controlled robotic devices, well established examples being Pipeline Inspection Gauge (PIG) systems for internal pipe inspections or unmanned aerial vehicles (UAV) for visual inspection. The use of such technology is very attractive in terms of safety, cost and the potential for minimal disruption to the inspection site especially if they allow plant operations to remain online.

1.2 State of the work at Strathclyde University

An established, industry driven, programme of work has been running within CUE which is concerned with the development of wireless miniature robotic vehicles or Remote Sensing Agent (RSA) for NDE. This work is a programme initiated by The Research Centre for Non-Destructive Evaluation (RCNDE) which is a collaboration between industry and academia within the UK. The objective of this organisation is to facilitate the transition from University research into industry applicable technology and has membership from large end user companies including Shell, Airbus, BP, Rolls Royce and the National Nuclear Laboratory. The RSA's offer a mechanically robust, flexible and expendable inspection solution designed specifically for NDE. They are able to operate in inspection scenarios for which human involvement may be too costly such as the requirement for scaffolding and climbing equipment or inhibited by access limitations and/or biological, radioactive or chemical contamination. Each robot in isolation can perform a

limited function, it is the combination of many such devices that essentially constitutes a reconfigurable 3D scanner capable of complex geometry inspection of structures with complex geometry.

Friedrich [9, 10, 11, 12, 13], developed a bespoke robotic vehicle, NDE sensors and data fusion algorithms for carrying out automated NDE. Following on, Dobie [14], [15] added greater processing capability to the platform, considered the positioning problem aspect of the reconfigurable scanner and investigated the use of guided wave sensors launched from transducers mounted onboard the vehicle. The evolution of the platform is shown in Figures 1.1 (a)-(c) where the latter two designs were used in this thesis. Figures 1.1 (d) displays an array of the previously developed NDE sensors the could be connected to an RSA unit to conduct an inspection.

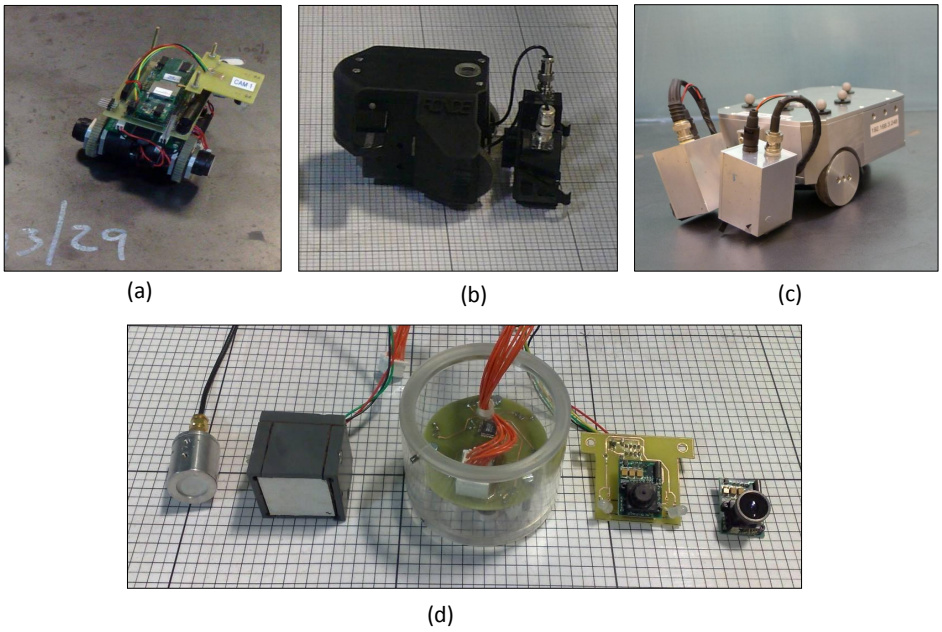


Figure 1.1: Evolution of the RSAs. (a) Mark I (b) Mark II (c) Mark III (current) (d) Sensors. Ultrasonic, Eddy Current and Cameras

1.3 Aims of this Thesis

Central to accomplishing the required degree of cooperating behaviour between multiple robots within the reconfigurable scanner concept, is the accurate positioning of the individual RSAs. Furthermore, the requirement for integrating NDE measurements onto the robots presents a significant challenge to the positioning problem. For useful NDE images to be assembled from the RSA scanning, there are a number of physical influences on the measurement process that can considerably degrade the quality of the NDE images and thus their usefulness. For example in air-coupled ultrasonic imaging applications, the separation and orientation of the transducers to the sample is critical [16]. This is in addition to the basic degradation of image quality from the gross positional uncertainty of the RSA. For example it may not be possible to assert defect presence or absence based upon comparison of expected time of flight and measured due to delays caused by error in location. For example, when scanning a surface accurate positioning is required to ensure full sensor coverage; deviating from the path leaves bands of missed surface area, while overlapping leads to longer inspection times. Positioning is also of importance for registering NDE measurements from different sensors acquired from multiple scans conducted at different times. In addition, it important that the robot is able to return to the same location on the structure repeatedly in order to monitor the time evolution of previously identified defects.

This thesis investigates a range of positioning strategies for a single RSA unit, the structure of the thesis is shown in the block diagram of Figure 1.2. Throughout this research a photogrammetry system is used to evaluate the performance of the developed algorithms. The accuracy and repeatability of this system for use as a ground truth system is investigated in Chapter 3. An onboard positioning system which extracts motion information from an image stream is developed and evaluated in Chapter 4. A low cost ultrasonic positioning system is characterised in terms of accuracy and repeatability for use in tracking applications in Chapter 5. In addition, several algorithms for ultrasonic ranging are investigated with a view to creating a high accuracy acoustic positioning system. In a real inspection scenario multiple positioning systems would need to be employed in order to

deal with the failure modes of each individual system. In Chapter 6, Bayesian filters are implemented for combining multiple sources of noisy positional data to produce an overall estimate of the position of an RSA.

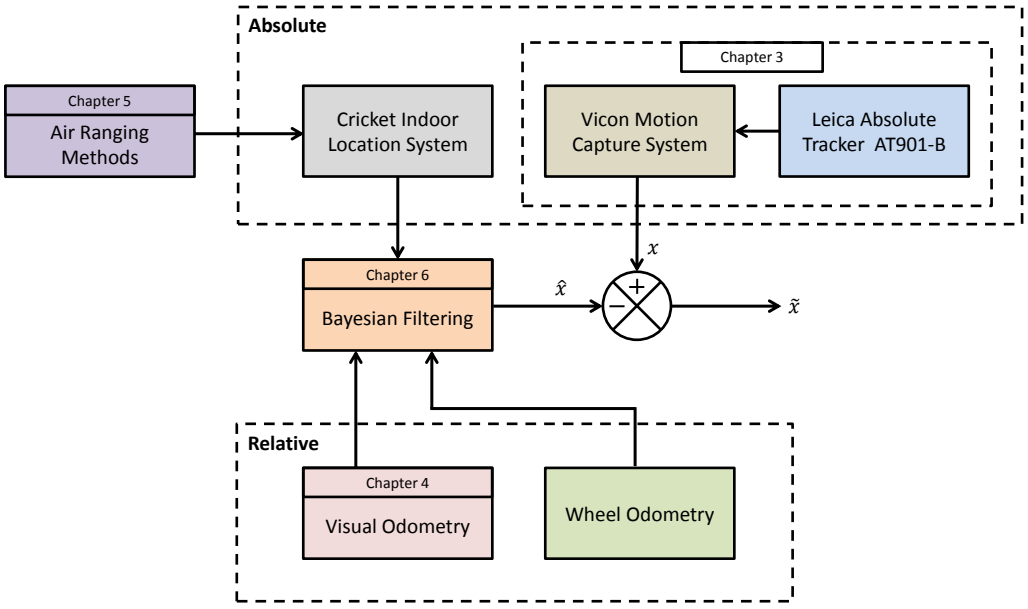


Figure 1.2: Block diagram showing the interconnections between the components of the thesis

1.4 Contributions to Knowledge

A high accuracy, high precision laser tracking metrology system was used to characterise the accuracy and repeatability of a commercial camera based motion capture system. The former provides single target tracking in 3 degrees of freedom while the latter has the capability of high target density, rapid tracking in 6 degrees of freedom and forms the ground truth system utilised throughout this thesis. The spatial error distribution of the camera system was evaluated following optimal calibration: such information may then be used to inform subsequent

experiments that utilise the equipment. Quantification of error was implemented through simultaneously tracking a test object scanned through a fixed measurement volume using both systems. A rigorous error analysis compares the accuracy of the camera based measurements to ground truth measurements as provided by the laser system. The presented analysis is the first detailed consideration of tracking error as a function of the measurement location in the system volume; this forms one of the main contributions of this thesis.

All sensors have limited perception capabilities imposed by the physical mechanism employed to perform measurements. This means that they have different accuracies, uncertainties and failure modes. In the case of robot tracking, data fusion is required to optimise continuous estimation of location. A detailed experimental evaluation of two Bayesian filtering algorithms using multiple positional sensing data to track the planar motion of a vehicle was carried out. A real time, on-robot implementation of an Extended Kalman and Particle filter was used to control a robot performing representative raster scanning of a sample. Both absolute and relative positioning were employed - the absolute being an indoor acoustic GPS system that required careful calibration. The performance of the tracking algorithms are compared in terms of computational cost and the accuracy of trajectory estimates.

A visual based relative positioning system was developed and evaluated for vehicles designed for challenging surface conditions, for which conventional optical encoder base positioning fails. An onboard machine vision camera is mounted perpendicularly to the heading vector of the robot such that it observes the surface over which the vehicle is travelling. The resultant orthographic images are transmitted over a network link to a base station PC for processing. The aim of the system is to provide the dual function of stitching successive surface images and simultaneously estimating the motion of the vehicle through consideration of the apparent motion of image features. The inspection domain determines the method of platform locomotion as result of surface terrain/material. Although the visual odometry by the nature of the estimation process will drift, it can provide a useful low level source of positioning data to be fed as an estimate into further processing.

The acoustic based positioning system is particularly attractive in comparison to the other systems employed in this thesis in terms of its relative low cost and small footprint; this property is of particular interest in access restricted inspections. Consider an inspection by a multi-robot system where access limitations preclude use of the laser tracker. In this scenario, the modularity such a system gives rise to the possibility of utilising, as the nodes of a positioning system, a subset of vehicles equipped with the associated acoustic positioning modules. The locations of such vehicles would become fixed upon entry into the inspection site and following the determination of inter module distances, these robots would act as transmitters to provide positional updates to vehicles conducting the inspection. his chapter provides a brief comparison of several methods for through-air ultrasonic ranging with a view to creating a high accuracy acoustic positioning system.

1.5 Publications Arising from this Thesis

Summan R, Pierce SG, Dobie G, Hensman J, MacLeod CN, "Practical Constraints on Real time Bayesian Filtering for NDE Applications", Mechanical Systems and Signal Processing (under review)

Dobie G, Summan R, MacLeod CN, Pierce SG, "Visual Odometry and Image Mosaicing for NDE", NDT International (under review)

MacLeod CN, Pierce SG, Sullivan J, Pipe A, Dobie G, Summan R, "An Active Whisking Based Remotely Deployable NDE Sensor", IEEE Sensors Journal (under review)

Dobie, G, Summan R, Pierce SG, Galbraith W, Hayward G, "A noncontact ultrasonic platform for structural inspection", IEEE Sensors Journal volume 11, issue 10, year 2011, pp. 2458 - 2468

J.C. Jackson, R. Summan, G. Dobie, S.M. Whiteley, S.G. Pierce and G. Hayward "Time-of-flight measurement techniques for airborne ultrasonic ranging" Transactions on Ultrasonics, Ferroelectrics, and Frequency Control

Summan R, Mackenzie L, Dobie G, Pierce SG and Hayward G, "Robotic non-destructive evaluation," Robotics and Vision, *One Day BMVA technical meeting at the British Computer Society*, 5 Southampton Street, London, UK on 24 March 2010

Summan R, Hensman J, Dobie G, Pierce SG, Worden K "A Probabilistic Approach to Robotic NDE Inspection", *36th Annual Review of Progress in Quantitative Nondestructive Evaluation*, Kingston, Rhode Island, USA, July 26-31, 2009

Pierce SG, Dobie G, Summan R, MacLeod CN, "Reconfigurable robotic platforms for structural non-destructive evaluation, *1st International Conference on Through-Life Engineering Services*, TESConf 2012, 5-6th November 2012, Cranfield University.

Macleod CN, Pierce SG, Sullivan JC, Pipe A, "Remotely Deployable Autonomous Surface Inspection and Characterisation using Active Whisker Sensors, *6th European Workshop on Structural Health Monitoring EWSHM*, Dresden, Germany, July 2012.

Pierce SG, Punzo G, Dobie G, Summan R, Macleod CN, McInnes C, Biggs J, MacDonald M, Bennet D, "Reconfigurable robotic platforms for structural health monitoring, *6th European Workshop on Structural Health Monitoring EWSHM*, Dresden, Germany, July 2012.

Pierce S.G., Worden K., Summan R., Dobie G., Hensman J.J. "Towards implementation of reconfigurable robotic strategies for structural health monitoring", *5th European Workshop on Structural Health Monitoring EWSHM*, Sorrento, Italy, July 2010.

Pierce SG, Dobie G, Summan R, Mackenzie L, Hensman J, Worden K, Hayward G, "Positioning Challenges in Reconfigurable Semi-autonomous Robotic NDE Inspection," SPIE, *Smart Structures and Materials + Nondestructive Evaluation and Health Monitoring 2010*, San Diego, CA, 7-11 March 2010.

Chapter 2

Problem Background

2.1 Introduction

In carrying out a task, a key aspect for any robotic system is determining its location with respect to a frame of reference. If a robot does not possess this function, its resultant behaviour consists entirely of reacting to sensory input within the field of view of its sensors and renders it incapable of planning actions beyond this vicinity. It is recognised that a fundamental requisite of autonomous mobile robotics is that the robot be aware of its location in space [17]. This chapter firstly provides a general overview of robotic systems and their applications. The concept of position in the context of robotics is then defined and an overview of the techniques and algorithms that can be used to compute position is provided. The techniques discussed in this chapter are employed by the algorithms and technology used in subsequent chapters, the reader can therefore use this chapter as reference for understanding a particular positioning methodology. A review of robotic inspection systems is carried out and compared to the idea of a RSA based scanning system.

2.2 Overview of Robotics

Robotic systems have uses in ever-expanding application domains including manufacturing, military, transportation, space exploration and hazardous environments driven by goals primarily based upon efficiency and safety. This technology has its origin in the manufacturing industry and has a key role therein by minimising costs whilst offering increased productivity, versatility and quality [18]: indeed robots are well established in large-scale manufacturing especially in the automotive industry and component assembly type tasks. Robotic technology has an important role in military scenarios in the form of ground and aerial vehicles for reconnaissance missions where it is too dangerous for soldiers to explore or the surroundings preclude their entry [19]. Intelligent vehicles defined as those with additional sensory perception, actuating devices and reasoning allow for such tasks as regulating speed based on traffic and safe lane following/changing is a growing area of research. Space exploration by orbital and surface robots in highly unstructured environments is an emerging field with the Mars Rovers [20] being a successful application. Remote handling manipulators are used in deactivation and decommissioning activities in the nuclear industry. Equipped with manipulator mounted power tools and multi-view camera capabilities such devices under human control are used to dismantle installations and remove debris [21]. Swarm robotics is a research field concerned with coordinating the control of large numbers of relatively simple, self-similar robots in order to achieve some goal as a collective. This area seeks to draw upon observations of the natural world taking inspiration from the interaction of insects such as ants, wasps. Examples of different types of robotic systems are shown in Figure 2.1.

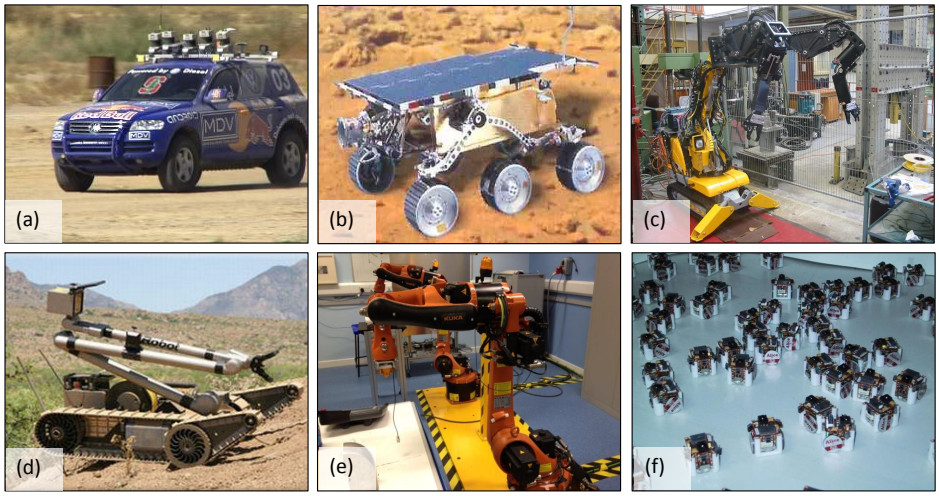


Figure 2.1: Robotic Systems used in different application domains (a) Self driving vehicle "Stanley" winner of the DARPA Grand Challenge (b) NASA Mars Rover (c) Nuclear Decommissioning Robot (d) Military Robot (e) KUKA industrial robotic arm (f) Swarming robots

The concept of autonomy with respect to robotics according to Gitalt et al [22] can be separated into two sub-concepts namely that of operating autonomy and decisional autonomy. The former refers to the ability of repetitively executing a set of actions that does not require human operator intervention: the authors consider that this a fundamental level of autonomy that in isolation does not render the robot autonomous. The latter sub-concept concerns the capability of the robot to make decisions to carry out a task for which the actuator actions have not been fully preprogrammed. The level of autonomy is a measure of the capacity of the machine to reason about how to execute a task by intelligently relating sensory observations to actuator actions. The level of autonomy embedded in a robotic system is a function of its application. In the automotive industry for example welding robots (e.g. KUKA robot in Figure 2.1 (e)) carry out well defined repetitive actions in a carefully engineered environment which is known *a priori*. The underlying instructions can be preprogrammed by a human operator in an offline phase before the robot is used online. Given the constrained nature of its operation such robots do not require sophisticated sensors to perceive the

environment. This contrasts with the application of a self-driving vehicle which is operating in an *a priori* unknown and therefore inherently uncertain environment. Such systems require various sensors such as cameras and LIDAR to sense the environment in order to detect and avoid obstacles, stay on the road etc - see for example Stanley winner of the DARPA challenge in Figure 2.1. A common thread linking these different types of robotic systems is the positioning task: the following sections present a review of positioning methodologies and discusses notable robotic inspection systems.

2.3 Review of Positioning Methodologies

In carrying out a task, a key aspect for any robotic system is determining its location with respect to a frame of reference. If a robot does not possess this function, its resultant behaviour consists entirely of reacting to sensory input within the field of view of its sensors and renders it incapable of planning actions beyond this vicinity. It is recognised that a fundamental requisite of autonomous mobile robotics is that the robot be aware of its location in 2D/3D space [17]. The positioning methodology employed depends both upon the mobility of the robot and the operating environment. This section presents an overview of positioning concepts/technologies that enable location to be calculated. The different forms of positioning can be divided into three categories;

1. *Relative Positioning*
2. *Absolute Positioning*
3. *Simultaneous Localisation and Mapping*

This following subsections firstly define the concept of position in the context of robotics and proceeds to consider each of the above approaches in turn.

2.3.1 Definition of Pose

The term *pose* refers to both the position and orientation of the robot relative to a world coordinate frame. Often robots are constrained to operate in a Two Dimensional (2D) plane, in this case, pose can be fully specified by the 3-vector:

$$\mathbf{x} = [x, y, \theta]^T \tag{2.1}$$

This descriptor defines the (x, y) position of the robot coordinate frame (r) in the plane as well as its orientation, θ, with respect to the x-axis of the World Coordinate System (WCS), this is illustrated in Figure 2.2. A description of this type is appropriate in applications such as navigation in a factory where the sensor platform is driving along a flat surface. If the robot is moving in a Three dimensional (3D) space then pose is represented by the 6-vector:

$$\mathbf{x} = [x, y, z, \theta_x, \theta_y, \theta_z]^T \tag{2.2}$$

composed of the Cartesian position (x, y, z) and orientation in terms of roll, pitch and yaw $(\theta_x, \theta_y, \theta_z)$ respectively, this is shown in Figure 2.3. Such a pose vector is applicable for describing the position and orientation of an aerial vehicle or multiple RSAs moving over a non-planar structure. The technologies and algorithms employed in this thesis estimate either a subset of the elements or all of the elements composing the pose vector of the robot. The process of estimating the pose vector of a robot is termed *positioning*.

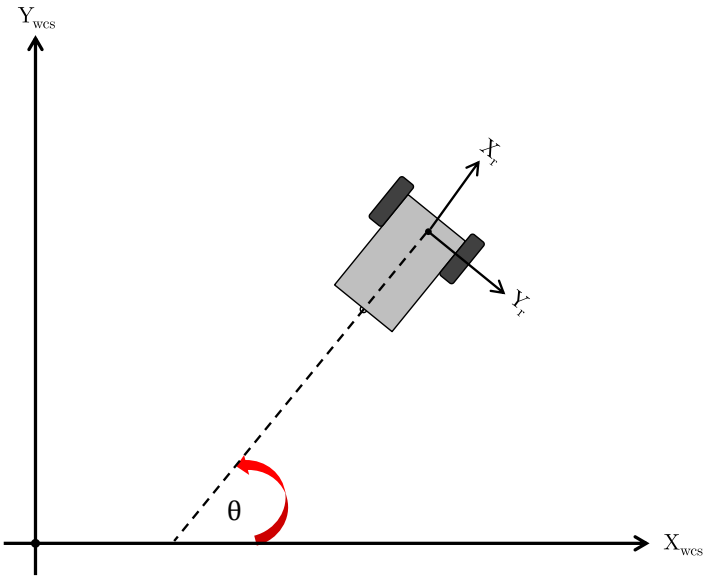


Figure 2.2: Variables comprising the pose vector for 2D motion

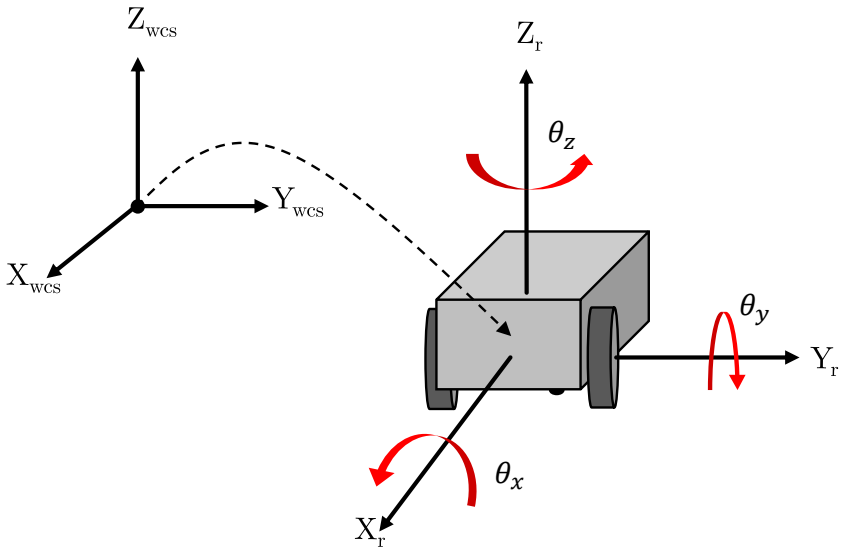


Figure 2.3: Variables comprising the pose vector for 3D motion

2.3.2 Relative Positioning

Dead Reckoning is the simplest form of position estimation. The technique recursively aggregates measurements of change in position to produce an estimate of the trajectory of the vehicle. Importantly it makes use of only on-board sensors to estimate change in location and therefore does not require the use of external positioning infrastructure. In this scheme the pose estimate at time step t is a function of the estimate at time $t-1$ and therefore any error in the prior estimate is carried forward to the current estimate: this problem is known as drift. Dead-reckoning can therefore produce an estimate that can deviate substantially from the true trajectory of the robot with distance traversed and so must not be used as the only source of positional information. The estimated change in position in general is not significantly erroneous and often serves as an input to be fused with some other form of positional estimate - this technique is investigated in Chapter 6. The effect of drift is illustrated very clearly in Figure 2.4 which shows GPS and wheel encoder estimates of a vehicle exploring the environment. The odometry estimate deviates substantially from the output of the GPS module due to many small errors integrated over the path length. Regardless of the underlying technology relative systems are subject to drift which grows unbounded with either distance or time and therefore such systems cannot be used as the sole positioning device.

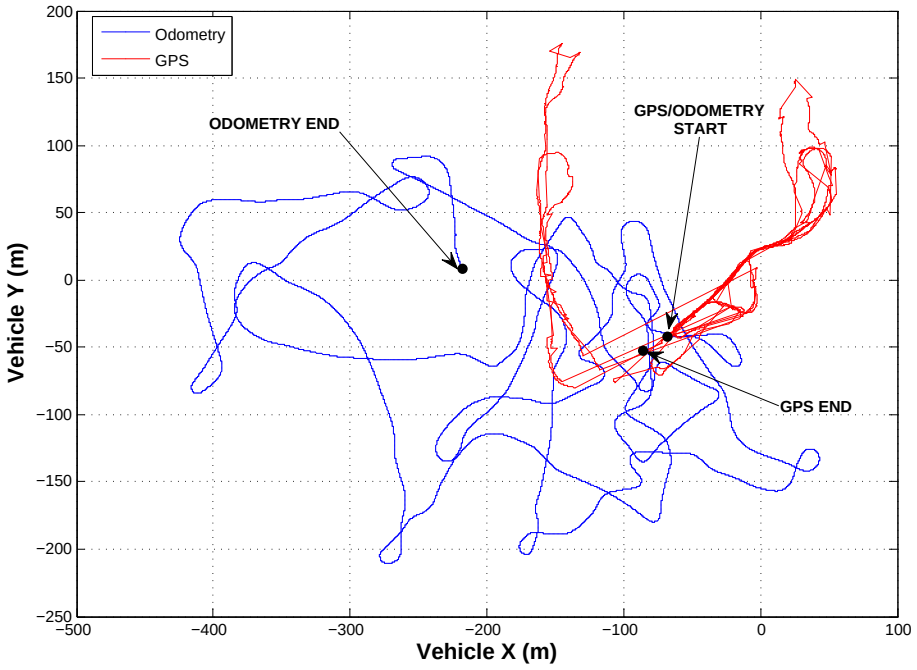


Figure 2.4: Extreme demonstration of encoder drift using the Victoria Park dataset [23].

2.3.2.1 Inertial Measurement Unit

An Inertial Measurement Unit (IMU) is a device that makes use of a number of measurement sub-systems to estimate the relative pose, velocity and acceleration of a mobile entity. These sub-systems typically comprise of gyroscopes and accelerometers. Gyroscopes measure rate of rotation by making use of physical laws that induce predictable results under rotations and are available in mechanical, optical and microelectromechanical systems based forms. The latter is based upon vibrating mechanical elements to sense rotation and is commonly used in robotics applications due to there being no rotating parts, compact dimensions and low power consumption. A single gyroscope measures rotation about one axis only, therefore, three such devices in an orthogonal configuration are used to measure 3D rotations. Accelerometers sense external forces acting upon robot, including the force of gravity, from which it determines acceleration. These devices

can be mechanical or piezoelectric based with the latter, which is more commonly used, measuring the voltage developed across a piezoelectric crystal. This voltage changes in response to a mass acting upon the crystal, which in turn is the result of external forces. As is the case for the gyroscopes, three accelerometers are typically used in an orthogonal arrangement.

The operation of the IMU utilises the integrated rotation rate of the gyroscope to transform the accelerations from the accelerometers into a local level navigation frame allowing the gravity vector to be cancelled out from the estimation: this results in an acceleration measurement of the vehicle. The acceleration estimate is once and twice integrated to produce velocity and position estimates respectively. The estimates produced by the IMU are very sensitive to errors in the raw data from the underlying sub-systems which are integrated over time resulting in drift and thus some form of absolute correction is necessary in practical use [18].

2.3.2.2 Optical Encoders

Optical Encoders are devices that can be attached to rotational joints in order to measure angle of rotation. In combination with a kinematic model of the drive mechanism employed by the robot, the output of the encoders can be used to estimate position. The principle of operation comprises of a focused beam of light being directed at a photodetector lying behind a rotating disk. This disk which is attached to the shaft whose rotation is to be measured contains an encoded pattern of transparent and opaque regions. When the light passes through the transparent areas on the disk, the photodetector senses the light beam resulting in a high output while opaque regions yield a low output. Once the factor of pulses per degree of rotation has been established, counting the number of pulses can be used in a straightforward calculation to obtain angle.

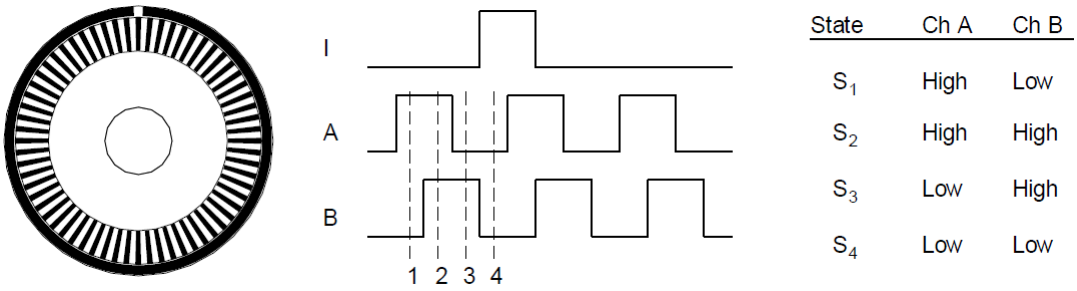


Figure 2.5: Incremental wheel encoder from Borenstein [24].

Encoders are available in two forms namely absolute and incremental form. The former divides the disk into N circular tracks that in part defines the resolution of the device. Multiple light beams and a detector array are then used to read clear/dark regions of the tracks where each angular position is encoded by a unique bit pattern as defined by these regions. Absolute encoders are only applicable for measuring slow and infrequent rotations. Incremental encoders, however, are utilised for measuring fast continuous rotation such as the drives wheels of a robot. These devices measure the motion of the shaft relative to its previous position. A particular type incremental encoder called a Phase-Quadrature encoder is widely used and is illustrated in Figure 2.5. This device make uses of two channels separated in phase by 90° and can be used to determine both the speed and direction of the rotating shaft and also increase the resolution of the device. The problem of encoder based drift is clearly illustrated in Figure 2.4.

2.3.3 Absolute Positioning

Estimation of the vehicle's position without drift and bounded uncertainty is only possible through use of an absolute positioning system. This type of scheme utilises infrastructure external to the robot and therefore requires the environment to be prepared prior to operation: this may not always be possible in real world scenarios. This infrastructure is composed of distinct landmarks the set of which is referred to as a *map*. The term *localisation* is used to describe the

scenario of estimating the robot's pose with respect to an *a priori* known map. The section presents the principles of operation for an absolute positioning system. Notable examples of this are the Global Positioning System (GPS) discussed in the following section. In addition to this the Facility for Innovation and Research in Structural Testing (FIRST) laboratory within CUE has a number of positioning technologies used during the course of this thesis which are discussed in detail below. A general overview of the fundamental measurements necessary to compute position are given.

2.3.3.1 Principle of Calculation

Fundamentally measurements of distances and angles to an object are required to calculate its position. There are several physical mechanisms that enable acquisition of these quantities which are discussed below but firstly the techniques that calculate location from this data are discussed. The two methods of position computation comprise of *Triangulation* and *Trilateration*; each shall be discussed in turn. Trilateration [25] is a method to calculate the position of an object given simultaneous range measurements from three stations located at known locations, given more stations multi-lateration can be used. Triangulation uses sensors separated by a known distance to measure the angle of the object with respect to each sensor. Given the fixed distance and measured angles the distance to the object can be calculated. These techniques have their origins in the surveying literature where they are used to compute accurate control points for creating topographic maps, charting lakes and for surveys of substantial public/private works [26].

2.3.3.2 Triangulation

Triangulation [27] is a technique that estimates distance to an object by using measurements of angles to the object acquired from sensors separated by a known distance or *baseline*. The technique is illustrated in Figure 2.6 where the sensors observe the target at bearings α and β. Using the angles in combination with the sensor baseline L, the object range, d, can be calculated as follows:

$$d = \frac{L}{\frac{1}{\tan(\alpha)} + \frac{1}{\tan(\beta)}} \tag{2.3}$$

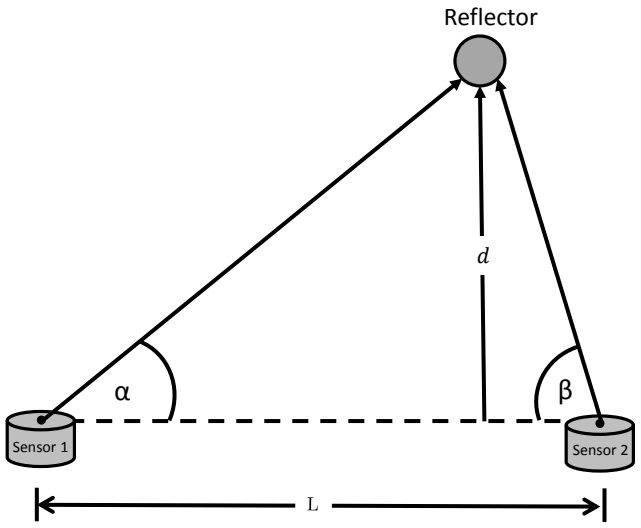


Figure 2.6: Triangulation diagram. The sensors measure the angle of the target with respect to the baseline of length L on which the sensors lie.

Triangulation is the method by which a stereo pair of cameras determine the 3D locations of imaged points. Each camera provides bearing measurements to points in the scene being observed. Assuming that the same point has been identified in both images, the pixel distance between these two points can be calculated and then triangulation can be used to to determined 3D position. Figure 2.7 which shows the pinhole model of camera illustrating the mapping from a real world

point to an image point. With reference to Figure 2.7, the following provides an overview of the principle of operation of camera triangulation. From a scene point $\mathbf{X}$ there emerges a ray of light that impinges upon the image plane of the camera at the point $\mathbf{x}$ and terminates at the projection centre (origin of the coordinate frame), C, of the camera coordinate frame. The camera coordinate system is related to the world coordinate system via the transformation matrix T. The image plane records only the direction of the ray. The mapping from the world or *scene* point, $\mathbf{X} = [X, Y, Z, 1]^T$ (expressed in homogeneous coordinates [28]), to the image point, $\mathbf{x}$, is expressed via the projection matrix, $P \in \mathbb{R}^{3\times4}$, as follows:

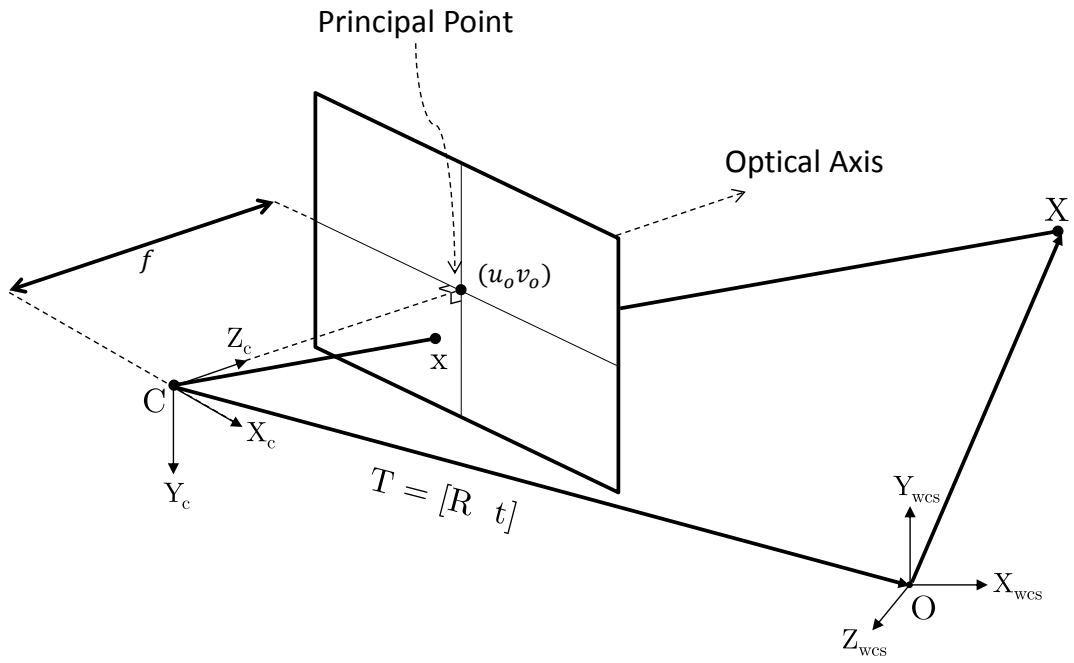


Figure 2.7: Pinhole camera model, diagram adapted from [29]. The scene point $\mathbf{X}$ is imaged at the point $\mathbf{x}$.

$$\mathbf{x} = P\mathbf{X} \tag{2.4}$$

where, $\mathbf{x}$ is given by:

$$\mathbf{x} = \left[f\frac{X}{Z}, f\frac{Y}{Z}, 1 \right]^T \tag{2.5}$$

which through the use of similar triangles may be observed from Figure 2.7. Importantly, the image coordinates are normalised by the depth, Z, of the world point. This means the from a single image 3D position cannot be recovered from the corresponding image point. The projection matrix can be decomposed into the the product of an intrinsic (also known as the camera calibration matrix), K, and an extrinsic matrix, T, as follows:

$$P = KT \tag{2.6}$$

The intrinsic matrix contains camera specific internal parameters and has an upper triangular form:

$$K = \begin{bmatrix} \alpha_u & s & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.7}$$

where $\alpha_u = fk_u$ and $\alpha_v = fk_v$ are scale factors in which, k_u and k_v, are the pixel sizes in u and v directions of the camera respectively. The factor, f, is the focal length of the camera, s, is a skew factor of significance for non-rectangular pixels and the coordinate, $\mathbf{u} = [u_o, v_o]$, is the principal point; the point at which the the principal axis intersects the image plane. The extrinsic matrix is composed of the rigid body parameters of rotation and translation relating the camera coordinate frame to the world coordinate system, it is expressed as follows:

$$T = \begin{bmatrix} R & \mathbf{t} \end{bmatrix} \tag{2.8}$$

where, $R \in \mathbb{R}^{3\times3}$, is a 3D rotation matrix and $\mathbf{t} = [t_x, t_y, t_z]^T$ is a translation vector.

The scenario in which the same scene point is imaged from two different views is illustrated in Figure 2.8 where it has been assumed that the camera has undergone both translation and rotation in moving from view 1 to view 2. The configuration yields specific geometric constraints which can be used to triangulate $\mathbf{X}$. The projection centres, C and C', together with the scene point $\mathbf{X}$ define a plane referred to as the *epipolar plane* which intersects the images planes observing the scene from two different viewpoints. The points of intersection of the baseline connecting the projection centres with the image planes are known as the *epipoles* denoted, e and e', in Figure 2.8. There are a number of geometric constraints that hold when the cameras observe the scene in the arrangement. If a ray is back-projected from the image point $\mathbf{x}$ then the resultant line is imaged in the sensor of view 2 as a line l known as an *epipolar line*. The image of $\mathbf{X}$ in the second view i.e $\mathbf{x}'$ is then constrained to lie along this line. The epipolar geometry is embedded in the fundamental matrix, F, expressed as:

$$F = K'^{-T}[\mathbf{t}]_x R K^{-1} \tag{2.9}$$

which is a function of intrinsic matrices K and K' associated with each camera and the translation, $\mathbf{t}$, and rotation, R, relating to the two views. Note, $\mathbf{t}$, in Equation 2.9 is in skew symmetric form:

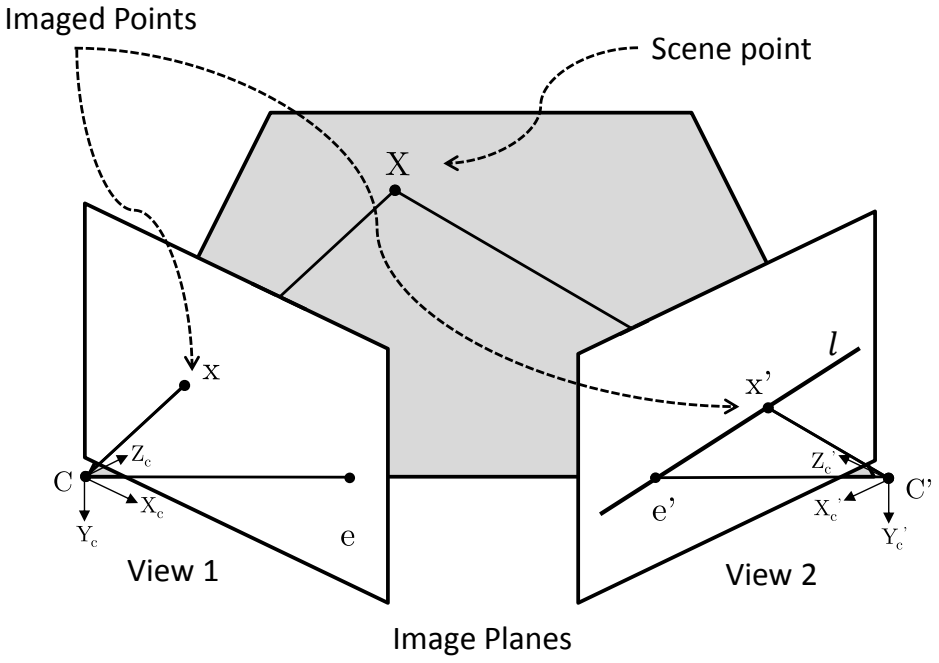


Figure 2.8: Illustration of the epipolar geometry that results from two views of the same scene.

$$[\mathbf{t}]_x = \begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix} \tag{2.10}$$

The objective is to extract the motion parameters from F. If at least eight image point correspondences between the two images are established then F can be calculated. A number of different techniques can then be employed to extract the rotation and translation factors: a review of numerically stable methods are described in [28]. Given a sequence of images acquired from a single camera the true metric motion of the camera cannot be determined due to the projective nature of the camera. The motion can be determined, instead, up to a projective

transform [28]. In a stereo configuration in which two cameras are separated by a fixed distance as in Figure 2.6 then it is possible to determined the 3D coordinate of an world point from its imaged points in the camera views.

2.3.3.3 Trilateration

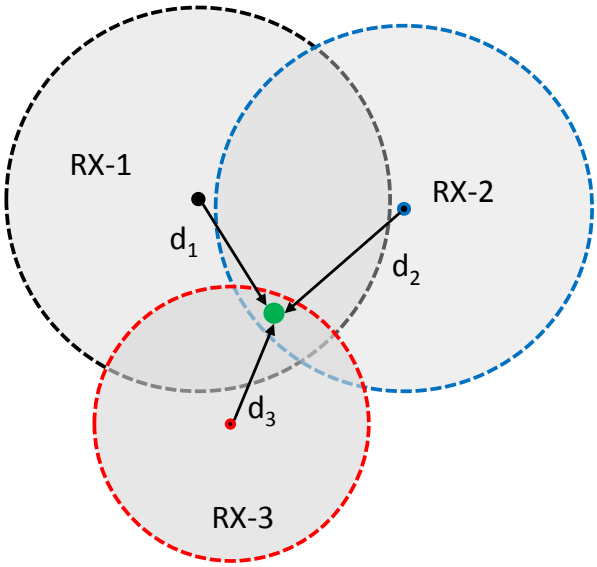


Figure 2.9: The transmitter is located at the intersection of the receiver centred circles resides at a distance d_i from the i^{th} beacon.

Trilateration is a technique for determining the location of an object through the use of simultaneous range measurements from three stations with known positions. This method is commonly applied in robotics as well as aeronautics and computer graphics [30]. The technique is based upon finding the intersection of three spheres which corresponds to the solution of the quadratic equations:

$$
\begin{aligned}
(x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 &= d_1^2 \\
(x - x_2)^2 + (y - y_2)^2 + (z - z_2)^2 &= d_2^2 \\
(x - x_3)^2 + (y - y_3)^2 + (z - z_3)^2 &= d_3^2
\end{aligned} \tag{2.11}
$$

where the receiver positions are given by the coordinates, (x_i, y_i, z_i), and the object-receiver distances are given by, d_i, for $i = 1...3$. The location of the object is (x, y, z), the principle of operation is illustrated in Figure 2.9. Note that the role of the transmitters and receivers may be exchanged. The solution to Equations 2.11 can be found either in closed form by assuming particular geometric arrangements or via numerical methods if the configuration is not constrained.

2.3.4 Ranging Methods

Time of Flight (ToF) is a method that employs pulsed waveforms to measure the time delay incurred by a signal in traversing the path from a transmitter to a receiver. Given the time delay, Δt, as measured by the receiver and the known signal velocity, v, the distance may be extracted in a straightforward manner through use of the following equation:

$$d = v\Delta t \tag{2.12}$$

In order to extract the time delay this method assumes that the transmitter and receiver are using synchronised clocks which may be complex to implement in practice - a diagram of the method is shown in Figure 2.10.

A sub-technique of ToF that avoids the synchronisation issue is the Time Difference of Arrival (TDoA) method which measures at the receiver the TDoA of the simultaneous emissions of two signals with different velocities, v_1 and v_2 from the

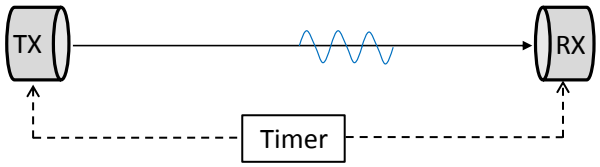


Figure 2.10: Time of flight. The transmitter and receiver have to be synchronised clocks (in effect the same clock in order) to measure the transit time of the pulse

transmitter e.g. acoustic and electromagnetic signals travel at different velocities. the technique is illustrated in Figure 2.11. The distance can then extracted using the equation:

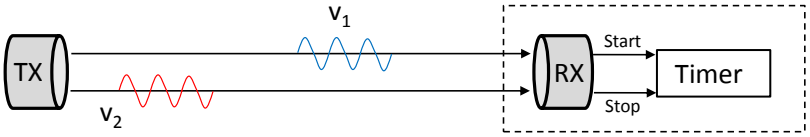


Figure 2.11: Time difference of arrival. The signal that reaches the receiver first starts the timer which is then stopped by the slower secondary signal resulting in the time difference of arrival.

$$d = \frac{\Delta t}{\frac{1}{v_1} - \frac{1}{v_2}} \tag{2.13}$$

If an electromagnetic wave is used as the physical basis of the signal, very high accuracy is required with respect to measurement of the transit time of the signal. Approximating the speed of light in a vacuum to 3×10^8 ms^{-1} and given that the estimated transit time has associated with it a standard deviation of 10^{-10} s, the standard deviation in distance is 15 mm using Equation 2.12. This requirement of high precision timing motivates the use of the following phase based technique.

The Phase Shift method utilises continuous wave emissions to derive distance through measurement of the phase difference, $\Delta\phi$, between the transmitted and received signals. The principle is illustrated in Figure 2.12. The distance measurement is obtained through summing the integer number of wavelengths and the fractional wavelength $\Delta\lambda = \frac{\Delta\phi}{2\pi}$ accrued in traversing the transmitter to receiver path [31]. The estimated distance is then given by:

$$d = M(\frac{\lambda}{2}) + \frac{\Delta\phi}{2\pi}(\frac{\lambda}{2}) \tag{2.14}$$

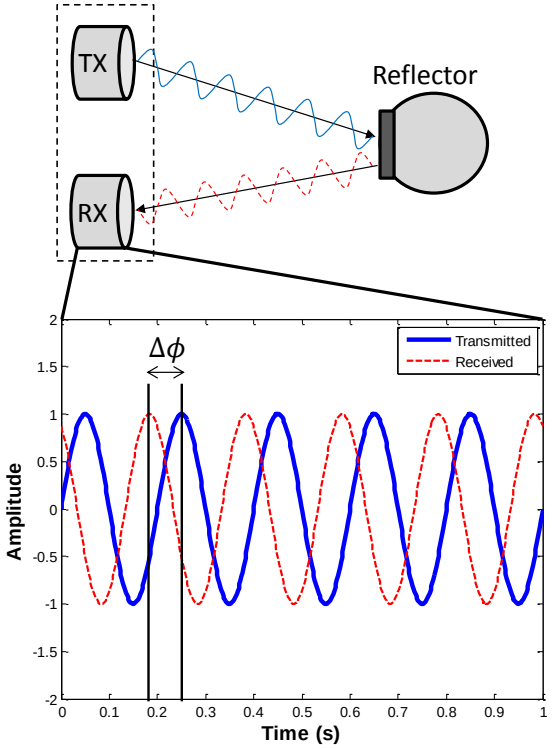


Figure 2.12: Principle of phase shift measurement. The phase shift $\Delta\phi$ measured through analogue or digital methods yields the fractional wavelength $\Delta\lambda$

where, M, is the number of complete wavelengths, $\Delta\phi$, is the phase angle difference and $\lambda/2$ is the fundamental wavelength of radiation. Evaluation of the phase difference only extracts the fractional wavelength. The integer number of whole wavelengths, M, is obtained through use of a secondary waveform with a larger wavelength. The high frequency wave serves to obtain the fractional part of the distance while the lower frequency secondary waveform is used to acquire a "rough" estimate of the distance. A potential issue with the phase ranging technique relates to $\Delta\phi$ going beyond 2π which give rise to the so called ambiguous range which does not permit the phase difference to be mapped to the correct distance. The ambiguous distance may be calculated by the formula [24]:

$$r_{ambiguous} = \frac{v}{2f} \tag{2.15}$$

2.3.4.1 Global Navigation Satellite Systems

Global navigation satellite systems systems such as the Global Positioning System (GPS) are capable of estimating the position of any point on or above the earth's surface assuming the receiver is visible by a minimum of four satellites [32]. The satellites comprising the system circulate the earth in predictable orbits allowing their positions to be estimated for the localisation task. One of the most commonly used systems is the NAVSTAR satellite system built and maintained by the United States. The RF signals transmitted by the satellites are synchronised by atomic clocks where the emissions comprise of encoded bit streams. The receiver solves for four unknowns in the form of (x, y, z) location with an accuracy of approximately 10 m and time given the time shifted bit streams from four satellites whose positions are known. This accuracy can be substantially improved through making use of fixed ground-based GPS stations with accurately surveyed coordinates. The error between the GPS estimated and known locations can be used to estimate signal noise and drifting bias which can be subsequently removed from the estimate of the target object's location. This

mode of operation is known as Differential GPS and can provide estimates with 1-2 m accuracy. A further variant called kinematic DGPS is also available and can provide an accuracy of 1-2 cm.

The main issue with GPS is the requirement for unobstructed line of sight - this requirement can very easily be violated particularly in nuclear inspection scenarios and also in other *in situ* inspections where the robot is surrounded by metallic structures acting to block the GPS signal. As well as the line of sight criterion, there is the problem of multipath where the GPS signal has not taken a direct path to the receiver and will subsequently result in a greater distance being reported and thus error in position.

2.3.5 Simultaneous Localisation and Mapping

There are many instances in which the robot must operate in environments for which there exists no form of positioning infrastructure. Absence of such infrastructure can result for a number reasons including access restrictions, dangerous operating conditions and high cost of the technology and associated installation. In addition to this the blueprints pertaining to a building often do not correspond with reality thus rendering the environment effectively unknown. This unknown factor means that the environment is inherently uncertain with respect to the self localisation task of the sensor platform. As well as environmental uncertainty, there is uncertainty associated with the robot it itself due to number of sources as described by Thrun et al [33]:

- Sensors are limited in what they can perceive
- Robot actuators do not execute the desired action perfectly
- Software models are approximations
- Algorithmic approximations used to bound processing time can potentially lead to reductions in accuracy

Asserting that the vehicle resides in a particular location without accounting for these sources of uncertainty in the data processing will results in divergence of the estimate from the true path: the uncertain nature of the data gives rise to more than one possible location. Uncertainty, therefore, needs to be explicitly modelled and accommodated in any algorithms utilising data from these sources where the variables involved are random variables with associated probability distributions.

The term *mapping* is defined as the task of determining the location of the landmarks in the area of operation given that the robot's true trajectory is known. *Simultaneous localisation and mapping* (SLAM) is the technique applied when neither the trajectory of the robot nor landmark locations are known: this is a more challenging problem than either localisation or mapping in isolation. The feature based SLAM technique builds up a dynamic, stochastic map of the environment in realtime using purely on-board sensors. Ideally the sensor provides both range and bearing estimates to targets in the environment. However, it is possible with greater difficulty to use only range [34] or only bearing measurements [35]. There are three forms of maps typically used in robotics, namely feature based maps, occupancy grids and topological maps. Each type of map will be considered in turn in the following sections.

2.3.6 Feature Based Maps

Feature based maps are a metric representation that model the environment by a sparse set of salient objects or *features* that naturally reside in situ. Features usually in the form of points and lines are extracted and parametrised from incoming sensory data. Man-made indoor environments typically have a rectilinear structure thus allowing extraction of such features from for example Light Detection And Ranging sensors (LIDAR). Feature extraction in general outdoor scenarios can be more difficult with respect to LIDAR's in terms of sparseness of reflectors in which case augmenting the robot's perception with vision sensors may be appropriate [36].

The operation of feature based SLAM comprises of two tasks namely that of adding features to the map whose parametrisations are functions of vehicle pose and then, when the robot moves, estimating pose from the landmarks in the map. These steps are illustrated in Figure 2.13. The true path of the robot is shown as the white sequence starting with pose $\mathbf{x}_{k-1}$ and ending at $\mathbf{x}_{k+2}$ driven by the controls $\mathbf{u}_k$ where k is the time index. The sequence of shaded vehicles results from the execution of noise corrupted controls leading to deviation from the true path. The point features L_1 and L_3 are observed from the robot at pose $\mathbf{x}_{k-1}$. The coordinates of these features with respect to the world coordinate system when initialised in the map are calculated from both the corresponding sensor reading and $\mathbf{x}_{k-1}$: any error in vehicle pose and the measurement are carried into the estimate of the point. When the robot subsequently moves to $\mathbf{x}_k$ driven by the control $\mathbf{u}_k$ and re-observes L_3, the estimate of the robots pose is determined from the control and importantly from the map and is thus a function of L_3. Uncertainty in sensor measurements, therefore, yields both uncertain estimates of feature locations and uncertain estimates of pose: the pose and landmark uncertainties are correlated. Due to this correlation in order to determine the true path of the robot, the map must also be estimated, therefore making feature based SLAM a problem of jointly estimating the map and the robots path.

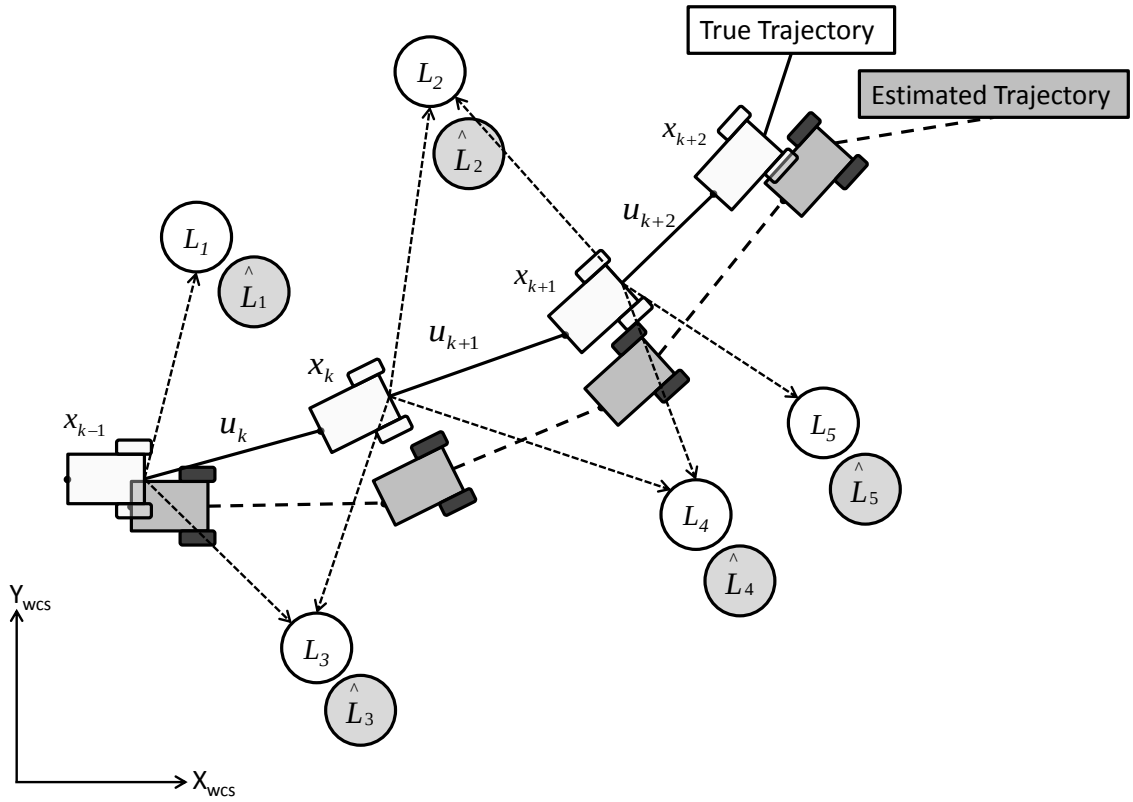


Figure 2.13: Representation of SLAM adapted from [17].

The technique works by tracking objects that are normally present within the workspace thus avoiding the need to install specially engineered landmarks to allow pose determination. The seminal paper by Smith et al [37] introduced the first mathematically convergent mapping technique. The proposed algorithm embraces directly the probabilistic nature of the problem through explicitly modelling the uncertainties pertaining to sensors and the actuators of the robot. The problem is cast into the framework of state estimation wherein the authors make use of an Extended Kalman Filter (EKF) as the basis of the estimation to incrementally update a map with sensor observations. The map comprises of a vector, $\hat{\mathbf{x}} \in \mathbb{R}^{(3N+1)}$, estimating the true state $\mathbf{x}$ holding the three elements constituting

robot pose (assuming planar motion) as well as the locations of N point landmarks in the environment parametrised by their 3D location, $\mathbf{L} = [L_{i,x}L_{i,y}, L_{i,z}]^T$, for $i = 1 \dots N$ as follows:

$$\hat{\mathbf{x}}_t = [x_t, y_t, \theta_t, L_{1,x}, L_{1,y}, L_{1,z}, \dots, L_{N,x}.L_{N,y}, L_{N,z}]^T \tag{2.16}$$

Both the robot and landmark coordinates above are referenced to a single global coordinate fame. Note that the time index, t, is applied to the pose of the robot only as the landmarks are assumed to be static in the environment over time, it is however possible to deal with objects that move over time and filter out their influence in the estimation process [38]. Associated with the estimate is a measure of its corresponding uncertainty and is encoded in the covariance matrix $\Sigma_t \in \mathbb{R}^{(3N+1)\times(3N+1)}$ as follows:

$$\Sigma_t = \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y & \sigma_x\sigma_\theta & \dots & \sigma_x\sigma_{L_{N,x}} \\ \sigma_y\sigma_x & \sigma_y^2 & \sigma_y\sigma_\theta & \dots & \sigma_y\sigma_{L_{N,x}} \\ \sigma_\theta\sigma_x & \sigma_\theta\sigma_y & \sigma_\theta^2 & \dots & \sigma_\theta\sigma_{L_{N,x}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sigma_{L_{N,x}}\sigma_x & \sigma_{L_{N,x}}\sigma_y & \sigma_{L_{N,x}}\sigma_\theta & \dots & \sigma_{L_{N,x}}^2 \end{bmatrix} \tag{2.17}$$

The diagonal terms are the variances associated with each variable in the state vector while the off-diagonal terms capture the cross-covariances between different elements of the vector. The level of dependency between elements shown in the matrix can be more easily shown via correlation coefficients calculated by dividing the covariance term by the standard deviations of the variables in isolation:

$$\rho_{ij} \triangleq \frac{\sigma_{ij}}{\sigma_i\sigma_j}, \qquad -1 \leq \rho_{ij} \leq 1 \tag{2.18}$$

where a coefficient of 1 (variables vary in the same direction) or -1 (variables vary in different direction) denotes full correlation and 0 indicates no linear correlation. As the sensor platform explores the environment and new landmarks are observed

the state vector grows in length and the covariance matrix expands with the square of the state vector size. This growth has implications with respect to the computation required to execute the algorithm: this is one of the main downfalls of the EKF approach. Implementation of SLAM requires solution to a number of sub-problems namely that of data association and loop closure discussed in the following sections.

2.3.6.1 Data Association

Data association is a core problem within SLAM and refers to the problem of correspondence which in turn is the problem of identifying the same entity in different sets of data e.g. recognise the same object in two different images. The parametrisation or *descriptor* associated with landmarks need to be sufficiently robust and information rich to enable reliable correspondence between incoming sensory observations and the current list of landmarks in $\hat{\mathbf{x}}$. The selected representation is also a function of the sensor being used for example image based descriptors would differ to those utilised with a LIDAR. The descriptor, additionally, must allow efficient comparisons to be made in order to achieve real time operation. Data association must determine if a new measurement originates 1) from one of the features already in the map 2) results from a new previously unobserved feature or 3) is spurious. Invalid data association is one of the primary causes of filter divergence in the EKF framework which can lead to complete failure of the algorithm [39]. Environments which contain repetitive structures can potentially lead to failed data association. There exist techniques which consider comparisons between batches of landmarks arranged in a particular configuration over simply comparing two points leading a higher percentage of valid matchings.

Data association can take the form of Nearest Neighbour (NN) gating which matches sensor measurements with map features by evaluating a distance metric: the pairing that produces the lowest metric and is lower than a preset threshold is flagged as a match. If the distance metric is higher than the threshold then it it likely that a previously unobserved feature has come into sensor view and can therefore be initialised as a new element in the map. The distance metric used is

often the Mahalanobis distance [40] which considers the positional uncertainties of the vehicle and feature in the distance calculation. The NN technique is effective in situations where the environment contains little clutter i.e. objects of no interest to the mapping problem, high sensor precision and moderate error in pose. If these conditions are violated then there exists more sophisticated techniques such as Joint Compatibility Branch and Bound [41] which takes into account constraints amongst multiple landmarks when searching for a pairing resulting in a lower rate of false positives. One of the main causes for failure of the feature based mapping technique is the problem of false positives in the data association process.

2.3.6.2 Loop Closure

Loop closure is the event in which the robot recognises that it has returned to a previously visited location after having travelled in a loop. Prior to loop closure the tracking error of SLAM is effectively the same as that of relative positioning. Loop closure constitutes one of the major sub-problems of SLAM and consequently has been a heavily researched area. This problem is highly related to data association but can be differentiated in that during loop closure there is much larger vehicle pose uncertainty in comparison to the uncertainty when determining correspondences over a short segment of motion. This increased level of uncertainty gives rise to a greater search space in terms of the required number of comparisons when trying to match current observations with landmarks in the map. When a loop closure has been found the error transform between the start and end points of the loop can be calculated and propagated back around the loop to correct the trajectory: this is the key to SLAM. As in the case of local data association, NN can be used in the process of determining loop closure: if the distance metric returns a value less than a threshold then loop closure can be asserted. The use of NN is only viable where the robot is not in gross error of its true location which is generally not the case during loop closure, therefore, more robust methods are required.

Detecting loop closure on the basis measured position is difficult due to the gross error in vehicle pose at the time of loop closure. Advances in processor technology, especially graphics processing units, have brought visual methods to the fore in recent years to carry out appearance-based approaches to closing loops. Newman and Ho [42] consider that the most difficult element of loop closure is detecting when it is even a possibility: it requires the vehicle to have the ability to decide where and when to look. It is not sufficient to search only within the local vicinity of the sensing platform given gross error in pose. The authors propose the use of a camera to provide a loop closing capability that is independent of the central SLAM estimation loop. The authors present a system that utilises a scanning laser to build the SLAM map and invokes camera data, which has been periodically collected, during loop closure events. Salient and wide-baseline stable image regions are selected from incoming images: image features are subsequently extracted and stored in a database. If a recent image is found to match an image acquired earlier then it is likely that loop closure is a possibility - this is predicated independently of the laser data.

Fast Appearance Based Mapping by Cummins and Newman [43] is one of the most successful vision based systems demonstrating stereo vision as the primary SLAM sensor. The system employs a sophisticated machine learning approach for performing data association. In an offline phase a large collection of images are acquired from the traversal of general outdoor scenery. These images are then input into a "bag-of-words algorithm" that effectively performs data reduction in generating a collection of v visual words that, in different combinations, can be used to approximate the original images - running this algorithm is computationally expensive. During online operation, the visual words enable places to be parametrised by binary vectors of length v, where the presence of a 1 in bit position N indicates that the N^{th} word is present in the image being considered - this operation is quick allowing for frame-rate operation. Probabilistic comparison of visual words is used to determine loop closures.

2.3.7 Occupancy Grid

An occupancy grid, introduced in [44], represents the environment as a set of cells each of which is assigned a probability of occupancy typically based upon laser or sonar sensor measurements. This form of map is widely utilized in robotics for dense environment modelling and is particularly effective where feature extraction is difficult to carry out [45]. They also have the advantage over feature maps in that both free and occupied space are represented which is beneficial for path planning and obstacle avoidance. Occupancy grids are commonly generated when the true path of the robot is known i.e. with *known poses*, it is however possible to map with *unknown poses* whereby the robot estimates its position online [46], [47]. In the case of known pose it is assumed that ground truth is available directly or SLAM is carried out to estimate the true path of the robot such that the occupancy grid can by generated in post processing offline. In the case that pose is unknown, it is a more difficult problem that aims to generate the occupancy grid online tending more towards SLAM. A brief summary of the theory for the former case is presented and then the latter technique is reviewed in the following section.

The scheme divides the environment into a regular grid which can be expressed as a set, $m = \{m_i\}$, with cardinality N consisting of the cells m_i each of which resides in a binary state: *unoccupied* $= 0$ or *occupied* $= 1$. The idea is shown in Figure 2.14 for a robot using a LIDAR.

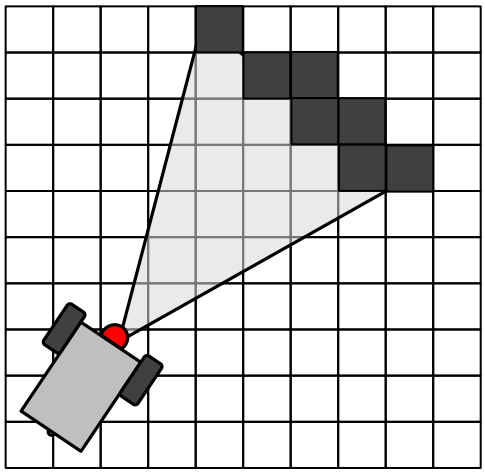


Figure 2.14: Occupancy grid generated by a robot carrying a LIDAR sensor. Detected obstacles are assigned a probability of occupancy in the corresponding grid cells

The occupancy grid ideally maintains a joint probability distribution across the the cells [33]:

$$p(m|z_{1:t}, x_{1:t}) \tag{2.19}$$

where, t, is time, $z_{1:t}$, is the history of sensor measurements and $x_{1:t}$ is the trajectory of the robot. The binary state of the cells gives rise to 2^N possible occupancy maps over which the distribution must be maintained, for large N i.e. a detailed representation, this is computationally intractable. An approximation of this distribution can be made by assuming the cells are independent thus allowing evaluation of Equation 2.19 as a product of the marginal cell probabilities:

$$p(m|z_{1:t}, x_{1:t}) = \prod_{i}^{N} p(m_i|z_{1:t}, x_{1:t}) \tag{2.20}$$

which drastically reduces the required computation at the expense of losing the ability to express dependencies between cells. It should be noted that the matrix representation could be wasteful for non-rectangular environments. Occupancy grids can take the form of planar maps acquired from robots operating on flat surfaces. Such a case is shown in Figure 2.15 in which an area has been divided into 10 cm x 10 cm cells that are assigned a probability of occupancy based on sonar sensor returns. White cells represent free space while the grayscale cells express probability of occupancy. Note that the grey background is due to a prior probability of cell occupancy. When the vehicle is not constrained to planar motion, occupancy grids can be represented by 3D grids synthesised using algorithms such as Octomap [48].

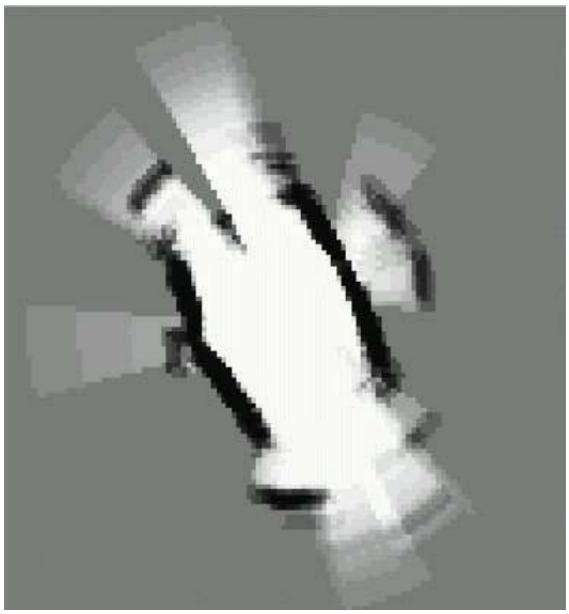


Figure 2.15: A 114×120 cell 2D occupancy grid from Thrun [45] generated by a robot carrying an array of 24 sonar sensors.

Occupancy grids facilitate fusion of measurements from different sensors by virtue of the grid structure. Stepan, Kulich and Preucil [49] demonstrate, under the assumption of planar motion, a system which fuses measurements from on-board laser, sonar and camera sensors toward improving robustness and accuracy of the generated map. Occupancy grids are firstly constructed for each sensor separately,

three methods are then presented for fusion. The first merges the maps on a cell by cell basis by combining measurements directly following filtration of false measurements due to for example specular reflections in the sonar. The second method casts the fusion process into a Bayesian framework employing Bayes rule to merge the probabilities of occupancy output from the sensors for each cell. The third method is an extension of the second with sensor precision being taken into account through comparison of a neighbourhood of cells resulting from sonar returns with higher precision camera measurements; if the two measurements don't agree then the sonar measurements are disregarded and the camera data is used exclusively.

Birk and Carpin [50] describe a system that merges the maps generated from multiple robots exploring the same environment with the objective of speeding up the mapping process. The authors detail a technique in which an optimisation algorithm seeks to maximise the overlap between acquired maps. Independent of pose information, a map m is rotated and translated with respect to a reference map m' until a common region is maximised as specified by an image processing based similarity metric. Results are shown in which maps from six robots are merged to produce a single global map. Once an occupancy grid exists navigation can be achieve by cross correlating sensor observations with the map. The computational effort is a function of the granularity of the grid, there is a trade-off between having a very detailed grid in which accurate localisation may be achieved at the expense of greater computational burden over the scenario of utilizing a coarse map offering fast but inaccurate localisation.

Schiele and Crowley [51] desribes a method of localisation using occupancy grids where it is assumed that the occupancy grid does not exist *a priori*. The approach taken involves maintaining two occupancy grids, one that models the local vicinity of the robot while the other is a global representation which evolves with exploration of the environment. Local grids are folded into the global grid via an update step which importantly takes into account the uncertainty associated with the robots position. The general technique firstly treats the grids as images through the application of an extended Hough transform developed within the paper to extract line segments i.e. areas where there are contiguous cells with

high probability of occupancy. Correspondence between line segments in the local and global grids are established from which rotation and translation parameters are calculated with slight variations of this being presented.

Yamauchi et al [46] make use of a similar concept in the generation of short term maps which are regularly registered with a long-term map of the environment: the process uses a constrained stochastic search to determine the optimal translation and rotation. The values calculated are then used to correct the path estimate of the robot such that is aligns with the long-term map. The occupancy grid is segmented using image processing based techniques into so-called *frontiers*, areas on the boundary between unoccupied and unexplored space. The objective of the described system is to find new areas to explore: frontiers serve as starting point to map previous unexplored areas of the environment.

2.3.8 Topological Maps

Topological maps are non-metric models of the environment. Such maps serve to represent the interconnections amongst points in the environment much like underground railway system maps. This kind of representation can be considered to be a graph in which the edges represent instructions for how to get from one node to another where nodes represent landmarks in the environment. Topological maps offer benefits in terms of fast path planning and a compact representation of the operating environment. One of the major issues with this method is that it does not allow accurate determination of pose of the vehicle and relies heavily upon the ability to recognise the places specified by the nodes [52].

2.3.9 Discussion

In a practical system the robot would explore the environment and in doing so build up a feature based map. Because of the computational complexity of the algorithms that implement feature based SLAM, the map tends to be a sparse representations of the environment. Given the feature based map which has

corrected for positional drift, a dense representation of the environment may be constructed through an occupancy grid. Occupancy grids divide the environment into cells and assign a value of occupancy to them depending up sensor observations; the number of cells that need to be updated is a function of the field of view of the sensor. Dependent upon the granularity of the grid, this means that updating the map can be achieved at potentially lower computational cost in comparison to the feature based map. Also, since the feature based map has corrected for drift, there should be minimal drift in the occupancy grid. Such a representation has the benefit of explicitly indicating free space which is required for path planning routines. The topological map abstracts away distance and scale information, rather it represents only connections between locations in a map and can be used for path planning purposes.

2.4 Review of Robotic Inspection Systems

The purpose of NDE is to detect internal and surface flaws such as cracks, voids, porosity, inclusions and corrosion in engineering structures without compromising the structure itself during the testing. Industrial sectors for which NDE is of major importance include aerospace, nuclear and petrochemical extraction and processing. Such industries are a source of particular challenges, often presenting inspection sites located in inaccessible locations or where environmental conditions are hazardous for human operators working at height, exposed to radioactivity, proximity to high temperature and/or pressure process plant. The financial impact of NDE inspections is also significant, arising from both the intrinsic inspection costs and the associated cost of taking plant offline to conduct inspections [7]. Application of NDE testing methodologies is often a labour intensive task. This means operator fatigue leading to higher potential of erroneous equipment usage and therefore "bad" data which is not representative of the structural condition of the entity under investigation. A number of NDE inspection tasks, currently conducted manually, can be automated with robotic devices to address these issues.

Development of inspection platforms is being driven by a combination of increasing safety legislation and the high economic and time costs associated with manual inspections. Robotic systems offer several advantages to inspection tasks over their human counterparts. The use of robots can decrease inspection times by operating in situ and thus avoid disassembling complex structures, scaffolding/excavation requirements and equipment recalibration. Zwicker and Zesch [53] report on the design of a robot for the NDE of low-pressure steam turbine rotors. The robot takes thirty minutes to complete the inspection whereas a manual approach takes an average of four hours, furthermore using two robots in parallel could reduce the time to fiften minutes: the time saving is substantial. This example also serves to illustrate the idea that multiple devices can potentially be used simultaneously and also that the number of robots can be scaled with the size of the inspection area. Robotic NDE can offer higher reliability than human operators as they do not fatigue and can lead to the generation of digital documentation of defects registered against position. Automated inspection where feasible, is highly attractive, and potentially allows inspection of operational plant. The safety, environmental and financial benefits for automating NDE measurements are clear, and applicable across a broad range of NDE technology.

Robotic NDE inspection platforms are an active area of research, there are numerous examples in the literature proposing devices for diverse application domains. The majority of the research in this area is funded from government agencies or arise from collaborative projects. The systems being utilised can be divided into two categories namely those that are designed for a specific task and those that are based around conventional robot work cells such as pick and place type robots that are being used to carry out repetitive scanning of components potentially prior to use : North Star Imaging Inc offers several real-time systems based on ABB or Fanuc robotic arms [54]. The former are highly specialised devices, with varying levels of autonomy, that must satisfy very specific constraints determined by the application. The design is a function of access restrictions, safety issues, structural geometry and the material composition pertaining to the entity that is to be interrogated. There exist a wide range of inspection conditions giving

rise to the development of inspection devices for a broad spread of engineering structures/components including storage tanks, pipe networks, ships' hulls and offshore structures. Independent of the application domain any inspection robot must address the following areas:

1. *Adhesion* - Due to the complex 3D shapes of structures encountered in the field, the adhesion mechanism must be capable of holding the robot in all inclinations. If the structure is ferromagnetic, permanent/electromagnets are typically used as the method of adhesion. In the case of non-ferromagnetic scenarios, vacuum suction is often employed.

2. *Locomotion* - The locomotion employed is one from the set comprising: fluid driven; wheel; caterpillar; wall pressed; inchworm; screw or some combination of these methods.

3. *Localisation* - Inertial measurement units, odometers, absolute positioning systems.

4. *System Integration* - Robots developed for NDT can be differentiated in terms of the level of system integration. Compact platforms require greater integration to satisfy size constraints. Larger robots operating in less restrictive environment can make use of commercially available sensors and control units.

With respect to the literature, the focus is mainly on the mechanical design required to negotiate the geometry of the specific structure under investigation. This section will concentrate on reviewing notable systems in the literature and consider in particular the positioning systems being employed.

Pipeline Inspection Gauge (PIG) systems are used to carry out internal pipe inspections of oil, water and gas pipelines and constitutes one of the most widely adopted methods for such applications. PIGs can be used in live operation whereby the fluid running through the pipe serves to propel the device along the pipeline. This type of platform often has a bullet like form, with diameter

matched to the pipe to be inspected. They were originally used for cleaning pipes but are now used to carry sensors to measure properties of the pipes: Magnetic Flux Leakage (MFL)and ultrasonic techniques are typically used to measure loss of wall thickness [55]. The position of the PIG is recorded using relative and absolute sensors. Odometers attached to arms that maintain contact with the pipe walls provide distance measurement along the pipe. IMUs are commonly carried onboard providing data to measure pipe curvature, orientation and location as well as PIG position. GPS tracked receivers are often attached to the pipeline allowing the passage of the PIG to be recorded providing an absolute measurement.

Czyz et al [56] present a pigging system which makes use of IMUs to provide dense measurements of bending strain along a 70 km pipe section. Regularly spaced control points comprised of distinctive features (such as valves, bends etc) along the pipeline are identified in the IMU data and registered with the differential GPS measured locations of these points to minimise positional drift. Given am inter control point spacing of 10 km and the characteristics of the IMU, the PIG has a worst case positional accuracy of 2.5 m.

Yu et al [57] describe a system which fuses data from on-board odometers with IMU data to combat drift. The data fusion is carried out using a filter based technique to yield a filtered trajectory error with respect to the absolute measurements of less than 1 m. Use of passive pigging is limited to inspection in which the pipe network consists of straight sections with mild bends and a diameter which can accommodate the PIG. Disadvantages include the inability to turn through 90° bends, failure to deal with couplings where the pipe diameter changes and the possibility of becoming stuck in the pipe.

A recent paper by Schempf et al [58] describes a pipeline robot developed to conduct inspections of natural gas distribution mains. The device is untethered with control being issued over a point-to-point communication system operating in the industrial, scientific and medical radio band using custom protocols and antennae. The robot is composed of interlocking modules each fulfilling certain functions: a set of modules contribute drive mechanisms necessary for controlled vertical

ascent/decent and active negotiation of 45°/90° bends while others contain NDE sensors, batteries (providing enough energy for a average single day mission) and camera/lighting capabilities. The robot has a diameter of 0.1 m while its length and weight can vary between $1.5 - 2.7$ m and $6 - 39$ kg respectively depending upon the number of connected modules. The camera subsystems are complimented with MFL sensors which together constitute the inspection sensor suite for the platform. Positioning is achieved through the use of encoders attached to the wheels of the modules and also through use of the camera counting the welds connecting pipe sections of known length. As a result of experimentation the authors conclude that the raw odometry provides an estimate of the robot and thus defect locations with an error range of $0.3 - 0.9$ m.

White et al [59] designed, developed and trialled a large ($0.5 \times 0.8 \times 0.5$ m^3, 20 kg with payload) mobile tethered crawler type robot for application in the aerospace industry. The system utilises vacuum suction as the traction mechanism for the rapid inspection of non-ferrous external surface structures such as wings enabling vertical and overhanging surfaces to be negotiated. An absolute positioning system in the form of a laser tracker (Leica LTD-800) providing (x, y, z) measurements is employed to track a pan-tilt mounted retroreflector: the retroreflector is actively driven to maintain line of sight with the tracker. In combination with angular feedback from the pan-tilt head and odometric data the Leica measurements are fused using an extended Kalman Filter to estimate the complete pose of the crawler. The authors report an accuracy on a flat vertical plate of ± 0.1 mm, the orientation accuracy, however, is not mentioned. The inspection system makes use of a probe that enables measurement of the full 6 degrees of freedom of the crawler and provides an accuracy of 0.05 mm in position and 0.04° in orientation.

The paper by Caprari et al [60] describes several prototype robotic devices for use in power plant inspections which are the result of industry-university collaboration between ALSTOM Ltd and the Swiss Federal Institutes of Technology in Zurich and Lausanne. One of the most impressive systems which has now been commercialised is *Magnebike* described by Tâche in [61]. Magnebike is a compact ($185 \times 143 \times 263$ mm^3, 3.5 kg), tethered, high mobility bicycle based

robot designed for inspection of complex shaped ferromagnetic structures such as the inner casing of pipes. The wheels are composed of permanent magnets that when coupled with magnetic force reducing lifters allow the vehicle to pass complex 90° convex and concave obstacles with almost any inclination with respect to gravity. Localisation is a detailed consideration in the system given that during operation in enclosed environments the device is likely to be out of view of the operator. Wheel encoders and a 3-axis accelerometer is used to provide an odometry estimate of the motion of the bike. A LIDAR in the form of a Hokuyo URG-04LX is used to periodically scan the environment producing a set of point measurements of the environment at time t that overlaps that obtained at $t-1$. A scan matching procedure is carried out to find the rotation and translation parameters relating successive scans and thus the 6 degree of freedom motion of the robot between scans. The scanning matching procedure is optimisation based in which the odometry estimate serves to provide a starting point. In order to minimise drift periodically non-successive scans are matched. The authors consider that a positional accuracy of ± 5 cm is necessary for safe remote operation. In the course of localisation experiments it found an that an accumulated error of 5 cm in position is true after a traversal of 0.33 m. The desired error can be maintained if scans are acquired at 0.33 ms intervals. In terms of the orientation estimate an accuracy of $\pm 10°$ is achievable by taking scans every 0.4 m of travel. This work focuses primarily on the locomotion and localisations aspects of the problem, NDE payloads have not been considered.

Unmanned aerial vehicle (UAV) based inspections have come to prominence as a result of advances in the underlying technology especially battery technology. Visual inspections from stabilised cameras conducted by UAV's offer cost reduction over the alternative of full size aircraft inspection. Such devices are being to carry out power line inspections [62]. There a number of companies offering inspection services by UAV's such as the UK based company Cyberhawk Inspections Systems Ltd [63].

The robotic inspection systems discussed tend to be purpose built to carry out a particular inspection job. This differs markedly from the concept of the RSA based scanning system composed of multiple general purpose robots that work

together to achieve the inspection goal. In all cases the above robots combine relative and absolute positioning systems available at different rates in order to estimate pose. In this arrangement, the lower rate absolute positioning system effectively corrects for the error accumulated by the higher rate relative positioning system. This method is demonstrated for a single RSA in Chapter 6 where an external acoustic based positioning is fused with onboard encoders to estimate pose in a planar environment.

2.5 Conclusion

This chapter has presented a review of the state of the art in robotic inspection systems. A common feature of such systems is that they are generally purpose built in order to address the specific inspection need. In addition, these devices often have constrained motion determined by the structure under investigation, for example, PIGs. This differs markedly from the concept of the RSA based scanning system which aims to provide a flexible and scalable inspection system that can potentially move in an unconstrained manner over complex geometry surfaces.

A central problem in the field of robotics is that of estimating the location of the robot. With respect to the RSA scanner, estimating the location of the constituent vehicles is very important in the NDE context for a number of reasons. Positional information is required to register a defect with the location at which it was found on the structure. However, the required accuracy of the positioning data is dependent upon the inspection scenario. For example, when inspecting a structure composed of panels such as an aircraft wing section, centimeter accuracy may be sufficient to specify where the defect is located - in reality an entire panel will be replaced. One of the primary advantages of a scanner composed of mobile robots is the ability to produce dense scans of an area. In such an inspection the quality of NDE images generated from positional information is directly related to the accuracy with which the position of the robot is known. Dependent upon the charactersitics of the sensor much greater than centimeter accuracy may be

required. When trying to return to a previously visited location on a structure to assess the condition of a known defect, the accuracy with which the robot can return to this location could determine if the structure is allowed to remain operational.

A review of techniques for estimating position was presented. These techniques may be divided into relative, absolute and SLAM. Relative techniques can be used onboard a robot and operate through sequentially composing estimates of relative motion to obtain an absolute trajectory. However, each individual estimate has an associated error and when many estimates are composed the resultant path can drift significantly over time and distance. Absolute positioning systems make use of sensors embedded in the operating environment that require line of sight to the robot. Such systems can provide positional estimates that do not drift. Clearly, for an in situ inspection absolute positioning systems may not be suitable due to access constraints and occlusions. SLAM based techniques offer the possibility of using features in the environment itself to navigate. Such features can be repeatedly identified and tracked by a sensor carried onboard the robot and used to build a sparse map of the environment. However, the technique relies upon the availability of features in the environment and also that the features are sufficiently distinctive such that they are identified correctly. SLAM has the potential to be the most applicable technique for the RSA based scanner, however, an absolute positioning system is likely to offer greater reliability.

This body of research makes use of several technologies that implement the discussed techniques. The reader may refer back to this chapter to gain an understanding of the underlying theory when these technologies are used in subsequent chapters.

Chapter 3

Spatial Error Characterisation of the Vicon Motion Capture System

3.1 Introduction

This chapter describes an experimental characterisation of the Vicon motion capture system, one of the main tracking systems employed during the course of this research. This system was used both to serve as the primary tracking system employed during experiments and was also used to validate the performance of algorithms developed to estimate vehicle pose. The study was motivated by empirical observations of Vicon measurement accuracy varying as a function of location within the system measurement volume. Due to the extensive use of the system, a systematic analysis was conducted to understand the error surface evaluated across the measurement space. Such positional dependence was thought to stem from the spatial distribution of measurements acquired during system calibration. The calibration phase of operation is a key step required before equipment usage often taking place twice daily in typical circumstances thus making the present evaluation an important consideration. The aim of this

chapter is to evaluate the spatial error distribution of the system following best-practice calibration: such information may then be used to inform subsequent experiments that utilise the equipment. Quantification of error was implemented through simultaneously tracking a test object volumetrically scanned through a measurement space using both the Vicon system and a Leica Absolute Tracker AT901-B. The accuracy of the Vicon measurements were compared to ground truth measurements as provided by the Leica Tracker which by virtue of its operating principle offers greater accuracy and precision. An overview of the two systems is firstly presented followed by a description of the experiment and the results of data analysis.

3.2 Overview of Systems

This section provides an overview of the positioning systems involved in the experiment describing their respective implementations of the techniques discussed in Chapter 2. The pros and cons of each system in the context of tracking mobile robots are also discussed.

3.2.1 Laser Tracker

The Leica Absolute Tracker AT901-B (LAT) is a metrology system intended for use in applications where high accuracy, precise tracking capabilities are required typically involving sparse target densities. This system is commonly used in the automotive and aerospace industries in tasks such as aligning and assembling parts and also where use of traditional fixed coordinate measuring machines are infeasible due to part size. The ranging method utilised is described in Section 2.3.4. The instrument measures the 3D location of passive Spherically Mounted Reflectors (SMR) through the projection of a laser beam from the instrument to the SMR. An image of the LAT and associated control equipment is shown in Figure 3.1 wherein the system dimensions equal 620 mm x 290 mm x 240 mm (ignoring the tripod) thus offering portability for field use. The head of the LAT

rotates around the vertical axis of the body, providing azimuthal scanning of the laser, while a rotating mirror provides elevational scanning. High resolution encoders with an angular resolution of 0.14 arc sec are attached to theses axes providing the angular data necessary to transform the native spherical coordinates into Cartesian form. The system is capable of detecting reflectors in a volume of radius 80 m at a rate of 1000 points/s. Control of the LAT is performed via the AT Controller 900 which runs the Leica EmScon server.

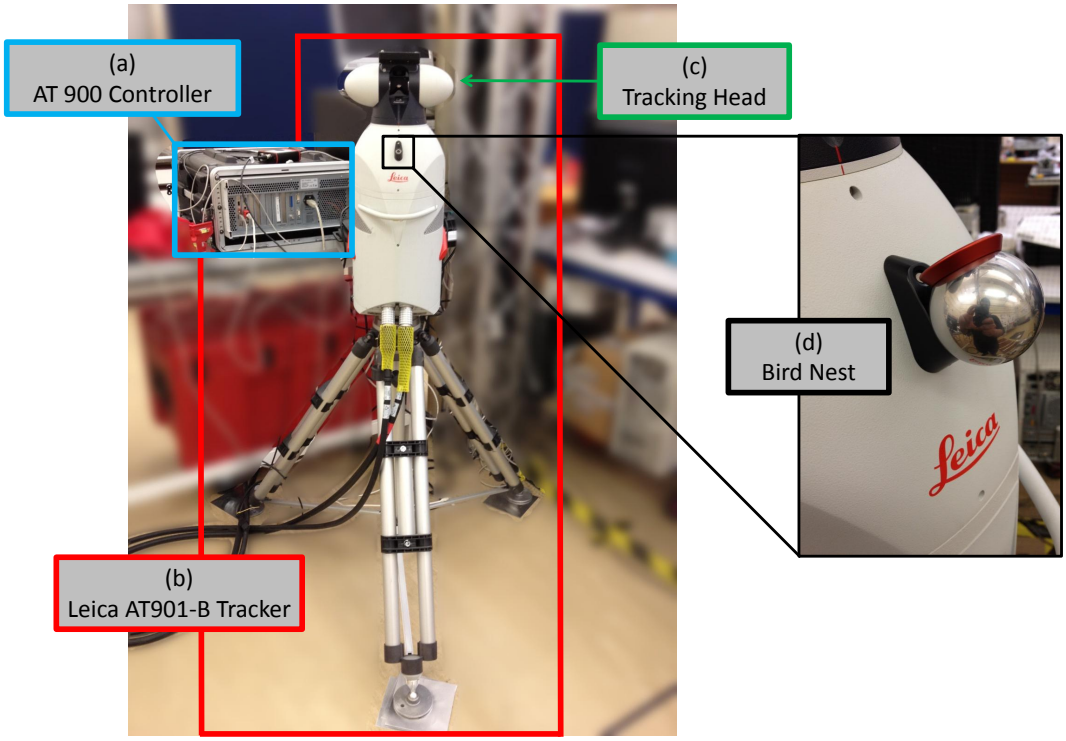


Figure 3.1: (a)AT Controller 900 running EmScon software (b) Leica Absolute Tracker AT901-B mounted to heavy duty tripod (c) Tracking head which moves during operation to track an SMR (d) Bird nest serving as a starting point for the IFM

A particular SMR called a Red Ring Reflector (RRR), which is attached to a magnetic mount known as a drift nest is shown in the image of Figure 3.2. The internal construction of the RRR consists of a corner cube reflector with an associ-

ated acceptance angle for the laser beam of $\leq \pm 30^{\circ}$ assuming the detachable ring is present. The ring serves to prevent the reflector from assuming orientations in the drift nest that would block entry of the laser beam: if removed, however, and used carefully this acceptance angle can be increased significantly. The spherical design of the SMR means that, at all times, its centre resides at a fixed distance offset with respect to any surface being measured. The major limitation of the LAT is the requirement for both line-of-sight between the tracker and the SMR *and* the reflector must be oriented such as to allow entry of the laser beam. The RRR offers the highest accuracy tracking available from Leica Geosystems and was the reflector used in the experiment.

The tracking system of the LAT is composed of two sub-systems namely an Absolute Distance Meter (ADM) and an Interferometer (IFM) which in combination is referred to as a Absolute Interferometer. The ADM provides *absolute* positional data by using time-of-flight to compute the location of a SMR with respect to the coordinate system of the LAT in a point-and-shoot type operation. The ADM provides a typical-value accuracy of ± 10 μm [64]. The IFM provides *relative* measurements of the SMR through monitoring the incoming and outgoing continuous wave signals. The interference of these waveforms is calculated wherein every peak in the resultant wave corresponds to a change in distance of half of a wavelength which in turn equates to a distance of $\Delta d = 0.32$ μm [65]. Through counting the number of peaks and multiplying by Δd the change in distance can be obtained: this operation can execute almost instantaneously. The IFM offers a typical-value accuracy of ± 0.2 μm + 0.15 μm/m [64].

The rate of operation of the IFM is far quicker than the ADM due to the integration time incurred by the ADM during distance calculation. The IFM is then well suited to dynamic measurements but as mentioned previously only provides relative measurements therefore some means of providing a datum point is necessary for IFM only based operation. Two options are available. The first option makes use of a known position attached to the body of the LAT referred to as the *bird nest*, see Figure 3.1(d), which serves as a starting point for IFM measurements: this imposes the limitation that the SMR be initially located in the nest and moved to the measurement location. The datum point may also

be supplied by the ADM if the laser beam is broken during system operation. Aided by an overhead camera which can find the lost SMR in incoming images and subsequently provide data to drive the tracker head motors, the ADM can be used to obtain the absolute position of the reflector at which point the IFM takes over.

(a) (b)

Figure 3.2: RRR, ø = 1.5", 170 g, constructed from surface-hardened stainless steel and has a removable ring. (b) ø = 0.5" fixed installation reflector intended for use with robotic platforms

The LAT outputs only the 3D coordinates of points using the default software running on the laser controller: it does not compute orientation of an object which has multiple reflector mounted onto it. There do exist specific products which can provide full pose information such at the T-Probe/T-MAC also available from Leica Geosystems. However, these devices are too unwieldy to be mounted onto the CUE robotic platforms. Another possibility for extraction of full pose data is to mount at least 3 reflectors on the robot and multiplex to measure each reflector in turn. A kinematic model of robot motion based on wheel encoder data could be used to guide the LAT in searching for the reflectors.

3.2.2 Vicon Motion Capture System

The Vicon Motion Capture System (VMCS) is a digital motion capture system that tracks the 3D motion of passive spherical/hemispherical targets from multiple camera images acquired from different viewpoints. The system is predominantly employed in the life sciences and animation production and also engineering in the form of virtual prototyping where dense object tracking is required. It also has specific application in robotic tracking with notable users being the Bristol Robotics Lab for tracking wheeled vehicles and the GRASP Laboratory at the University of Pennsylvania and ETH Zurich for tracking UAV's. The VMCS employs multi-view stereo image processing to locate a single target in space as described in Chapter 2: at least two cameras are required to observe a marker in order for its 3D position to be determined through triangulation. The CUE laboratory setup is as illustrated in the schematic of Figure 3.3 which shows six T160, 16 megapixel cameras symmetrically arranged at two ends of a rigid frame enclosing a large measurement volume. Through attachment of at least three markers to the object to be tracked, its orientation in terms of yaw, pitch and roll may also be estimated alongside the 3D location. If there are multiple objects to be tracked, then arranging markers uniquely and asymmetrically for each object allows simultaneous tracking. The VMCS is also capable of dealing with partial observation of the tracked objects.

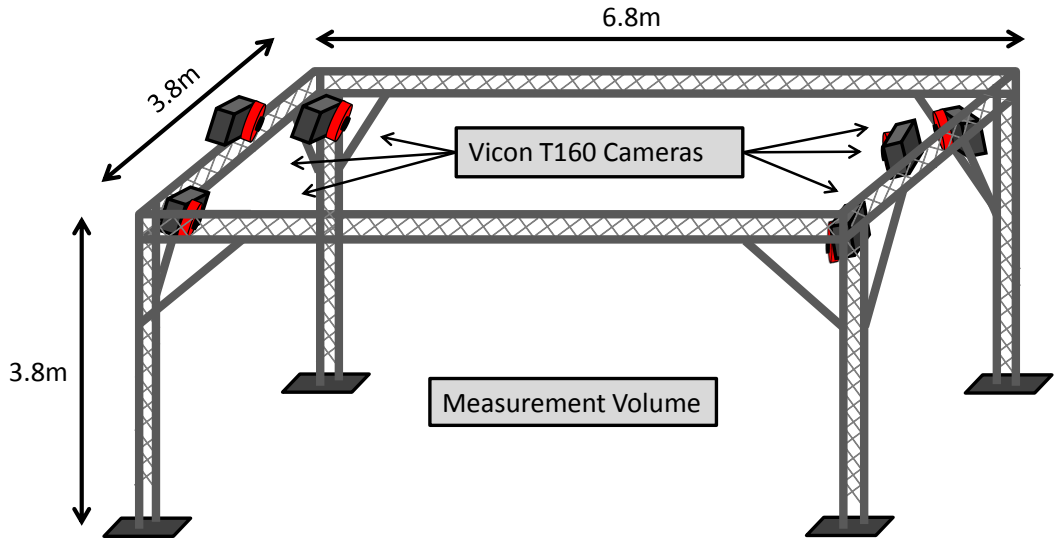


Figure 3.3: The cameras were clamped using professional grade camera mounts to the smallest sides of a rigid aluminium frame (pipes ø = 50 mm).

Each camera has mounted onto it a strobe unit which consists of a ring arrangement of LED's that serve to flood the measurement volume in high intensity infrared light: an image of a T160 camera is shown in Figure 3.4. A strobing unit containing a ring of LED's comprising 320 visible red (623 nm) and 252 Near IR (780 nm) is attached to the front of the camera. The image sensor consists of custom 4704 × 3456 CMOS 10 bit sensor while frame rate can be can vary between 30-2000 fps. The processing unit contains three processors carrying out tasks such as masking, thresholding and 2D tracking [66] Under such illumination, the Vicon targets which individually comprise of a spherical construction of wrapped retroflective tape, glow very brightly in the images captured by the cameras. A 12 mm diameter Vicon target is shown in Figure 3.5(a) illustrating the wrapped tape construction. When illuminated the marker produces a very strong reflection as shown in Figure 3.5(b).

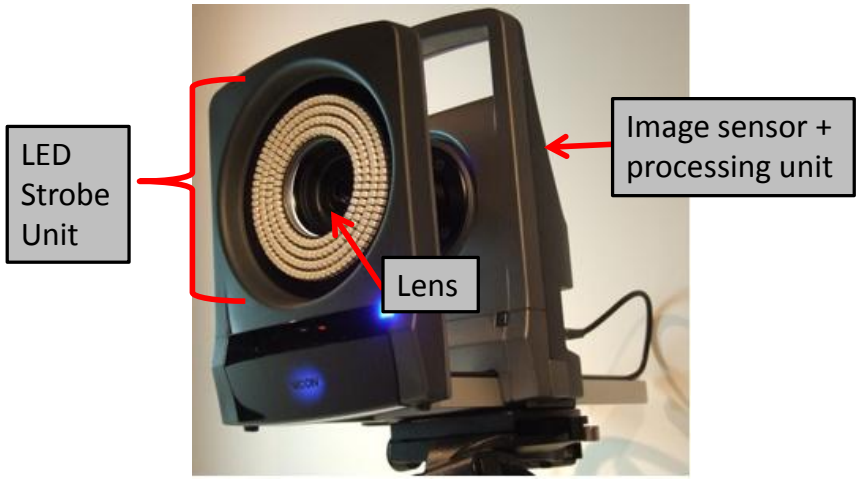


Figure 3.4: Vicon T160 camera

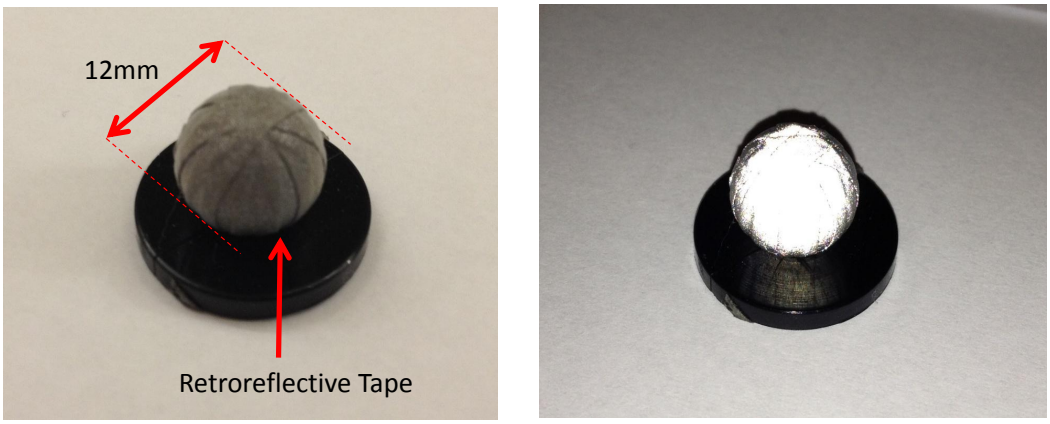


Figure 3.5: (a) 12 mm diameter spherical Vicon target consisting of wrapped retroreflective tape mounted upon a plastic base. (b) Vicon target under IR illumination from a camera flash illustrating strong reflective properties of the tape.

The system is controlled by a closed-source software package called Vicon Tracker. This particular software allows the motion of rigid bodies to be tracked and assuming full resolution capture this can be achieved at a rate of around 120 Hz. The Tracker representation of the camera configuration of Figure 3.3 is

shown in Figure 3.6(a) where each frustum represents a camera. The intersection of the bold white cross in the grid below the cameras represents the origin of the system coordinate frame. Figure 3.6(b) shows the virtual representation of an object composed of six retroreflectors shown as the connected network of spheres. The object has an associated coordinate frame with the red, green and blue perpendicular axes above corresponding to the x, y and z object frame axes respectively. The tracked point on the object corresponds to the centroid of the markers which in turn resides at the centre of the cube in the image, the tracked point can also be user defined. Note that this view follows from carrying out a calibration procedure discussed in the proceeding section.

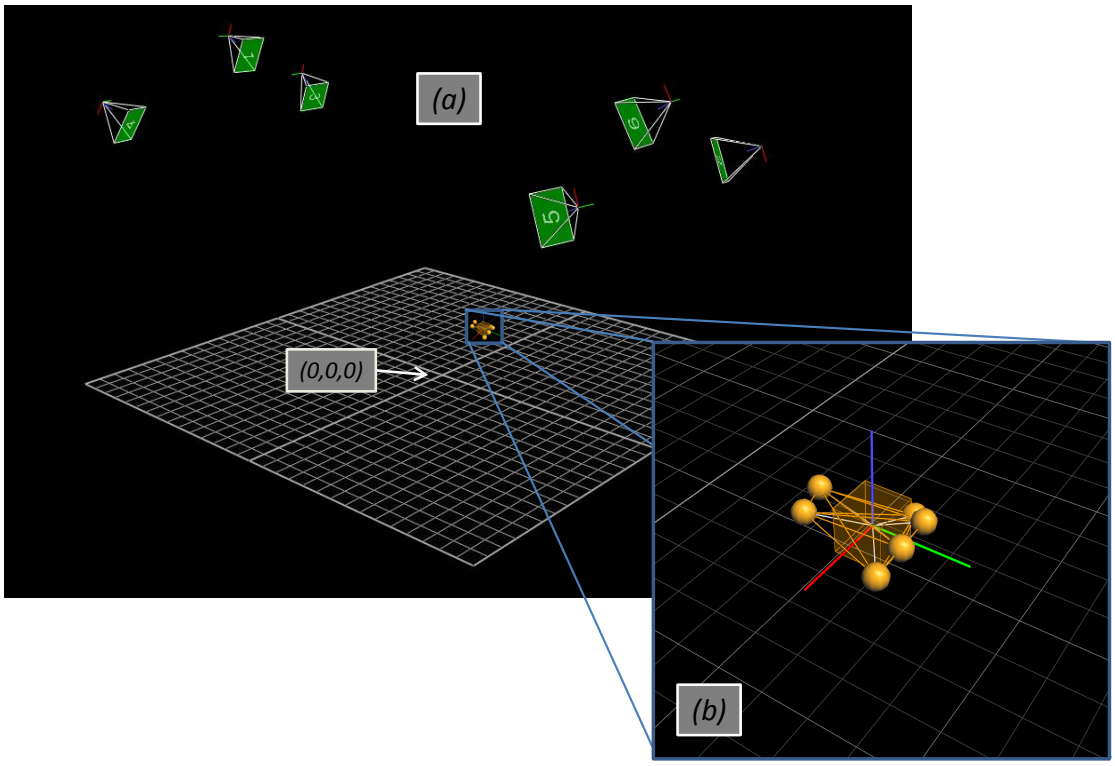


Figure 3.6: (a) Virtual representation of Vicon cameras in Tracker software. (b) Virtual representation of object.

3.2.3 Tracking Mobile Robots

The section compares the VMCS and LAT in the context of tracking mobile robots. The VMCS is capable of tracking multiple objects over large volumes where the size of the volume can be increased by using additional cameras. Importantly, the system provides both position and orientation directly at 120 Hz. However, it requires the cameras to be rigidly fixed in the environment and also arranged around the workspace. Furthermore, the system requires calibration which requires access to the workspace. Deploying this system in the field, therefore, would be problematic for access restricted areas. In addition to this, shiny materials in the environment could be confused for retroreflector targets. For these reason the VMCS is particularly suited for controlled laboratory settings.

The LAT is a able to track the position of a single target at any one time over a large volume from a single viewpoint. The tracking volume can be increased through the use of multiple trackers. The system consists of the instrument itself and a control unit and therefore offers great mobility allowing for transportation to site locations in the boot of car. The requirement for only a single view would be advantageous in a number of environments. For example, in nuclear scenarios access to areas in some cases can be limited to a small hatch. However, the SMR beam acceptance limits could be challenging to meet in complex industrial environments. The ability to track only a single object at a time is limiting if using multiple robots. This could be alleviated to some extent by searching in areas determined though predicting the positions of the robots from other sensors. The system is only able to provide position information directly at 1000 Hz. However, through attachment of multiple reflectors orientation could also be determined. The LAT is well suited to in the field use, however, integration with other sensors is required to maximise its potential.

In combination, the two systems provide tracking capability over a wide range of inspection scenarios.

3.3 Experiment

The objective of the experiment was to carry out a volumetric scan of the measurement space in order to characterise the accuracy of the VMCS as a function of position with respect to ground truth LAT data. In order to conduct the comparison, the systems had to measure in the same coordinate frame. Section 3.3.5 describes the procedure used to create a global coordinate frame shared by both systems. In addition to this, the object coordinate frames had to be made equal to ensure that each system was measuring the same point in space - the method used is described in Section 3.3.9.

Prior to carrying out this task the LAT was levelled in the pitch and roll axes of the system and the necessary calibration procedures were carried out. It should be noted that the exact numerical results obtained from the data set acquired during the course of this experiment is calibration dependent and thus may vary between calibrations. The calibration procedure was carried out in a repeatable fashion in order to try to minimise the effect of this variable in the experiment. The objective was then to identify trends in the spatial distribution of the data. The following sections describe the steps taken to optimise the performance of the VMCS, the design of the experiment and the results of data analysis.

3.3.1 Volume Coverage

A key step in the setup procedure for Vicon is that of positioning the cameras so as to maximise coverage within the desired measurement volume. Each camera has an associated Field of View (FOV) in the horizontal and vertical planes which is a function of the camera lens and takes the form of a cone. Figure 3.7 (a) illustrates an approximation of this cone for a particular camera that may be observed in the Tracker software. The problem of measurement coverage reduces to maximising the use of the FOV of each camera in terms of covering the desired measurement volume whilst simultaneously ensuring that at least two cameras can observe any Vicon target within this volume to allow its location to be tri-

angulated. The arrangement of the cameras was determined through an iterative process of traversing the volume with Vicon targets seeking out blind spots and making appropriate adjustments to the camera positions. The final arrangement of the cameras is shown in Figure 3.8 where the cameras were rotated to lie in a horizontal position in order capture markers at the edge of the volume and the central camera was mounted higher to maximise the height of the measurement volume. The intersection of the cones for one side of the volume is shown in Figure 3.7 (b), note that this configuration is mirrored on the opposite side of the volume.

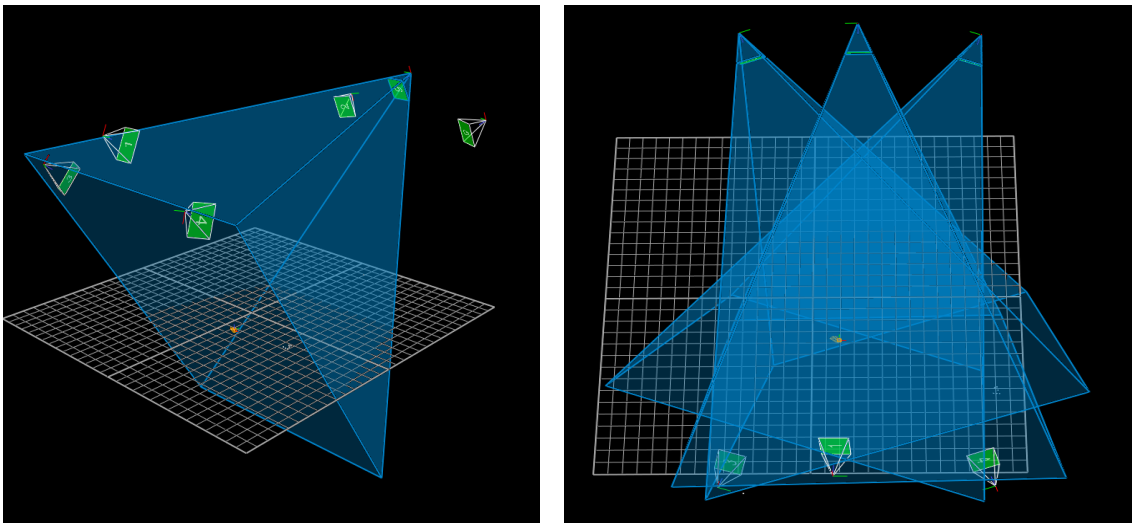


Figure 3.7: (a) Approximate FOV cone for a particular camera used for guidance in positioning the cameras. (b) The intersection of the FOV's of the cameras located at the right hand side of the volume. This arrangement is repeated on the opposite side but has been left out for clarity.

3.3.2 Vicon Calibration

The accuracy of Vicon is a function of the data gathered during the system calibration phase of operation. For the calibration a "wand dance" is carried out whereby a calibration artifact or *wand* with *a priori* known dimensions and

Figure 3.8: Camera arrangement on the right hand side of the frame in Figure 3.3. This arrangement was found to maximise coverage within the measurement volume.

marker arrangement is swept through the volume whilst being moved in a figure of eight motion. During this time each camera captures a user defined number of frames of the wand in different positions and orientations. The resultant point cloud data gathered by each camera is then used in an optimisation process to determine the intrinsic camera matrices and relative positions of the cameras. The calibration wand can then be used to define the location of the global coordinate frame through placement of the wand in a user defined location in the volume. The quality of the calibration and thus accuracy of the system is a function of the point cloud data and therefore the wand motion as executed by the operator. The calibration quality is measured by image error, a value which is associated with each camera and is expressed in terms of pixels. The image error also depends upon the settings of lens aperture and focal length which need to be adjusted before operations.

3.3.3 Optimal Lens Adjustment

Prior to conducting the experiment each camera lens was adjusted to ensure optimal system operation. The procedure consisted of placing the calibration wand into the centre of the volume followed by adjusting the aperture and focal length of each lens according to the characteristics of the imaged wand. A thresholded grayscale image of the wand in Figure 3.9 for a particular camera is shown in the Figure 3.10. This shows an ideal image of the wand resulting from application of the below adjustments. The specific marker characteristics that are of concern in this process are the *intensity* and *circularity* of the pixel arrays representing markers in the image. As part of the processing used in the system, circles are fitted to the marker pixel arrays and thus suboptimal lens optics can be a source of tracking error.

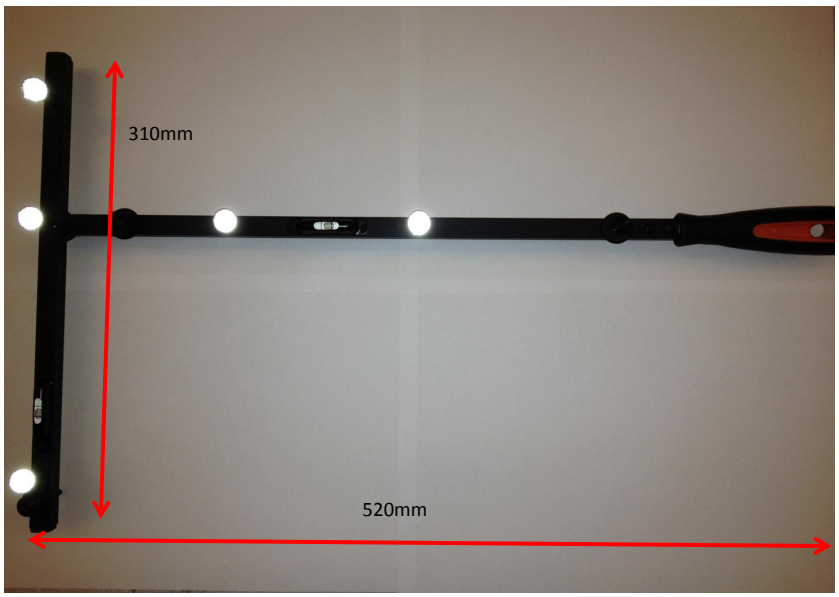


Figure 3.9: Calibration Wand

The aperture controls the amount of light entering the camera and thus the intensity of the pixels corresponding to the markers. Figures 3.11(b) and 3.11(b) illustrate the effect of the aperture for semi and fully open settings. The pixel intensity of the former is of an intermediate level thus making it more error prone

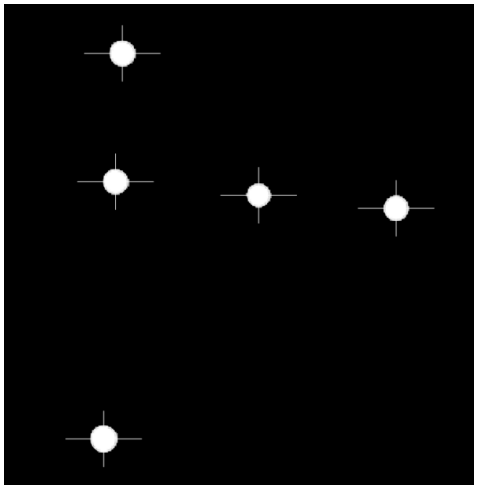


Figure 3.10: Example camera view of the calibration wand used to adjust lens properties for a particular camera.

while the latter is the ideal consisting of a high intensity array less likely to be affected by noise. The fully open setting can only be realised in environments containing little noise contributing anomalous reflectors. The aperture of each camera was set to fully open or nearly fully open as the measurement volume was clear of reflective materials.

Lens focus controls the circularity of the imaged marker which ideally should be a circular array of pixels. Figure 3.11(a) shows this ideal shape while Figure 3.12 displays a suboptimal focal length setting. The focal length of each camera was adjusted to yield circular pixel arrays such that the final wand image was of the form of Figure 3.10.

3.3.4 Custom Vicon Target

In order to compare the output of the VMCS with the LAT it was required that both systems measured the same point in space. This problem could be solved in a direct manner through making use of markers of equal dimension for both systems. The LAT RRR offered the most accurate optical centring accuracy of

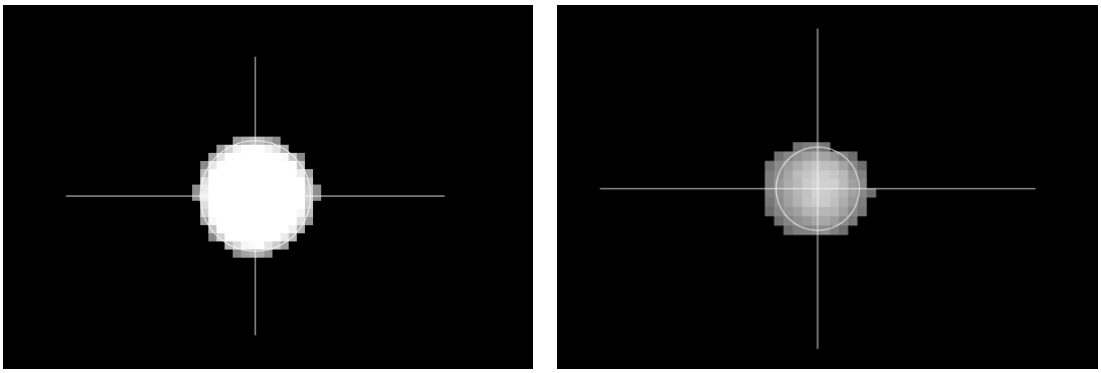


Figure 3.11: (a) Aperture fully open showing high intensity white pixels representing a marker (b) Aperture semi open resulting in low intensity pixels for a marker

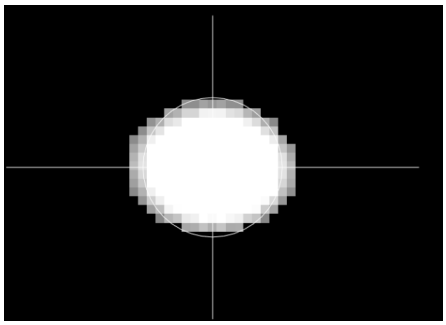


Figure 3.12: Non-circular array of pixels for marker resulting from sub-optimal setting of focusing dial on lens

the available SMRs, therefore, it was chosen for use in the experiment. Because a VMCS target of equal size did not exist in the set available targets, a custom spherical VMCS target with a diameter equal to the LAT RRR was created.

A 1.5" steel ball bearing was used as the basis of the target thus allowing it to be held in a magnetic drift nest during measurements. The bearings were given a retroreflective capability through bonding to the surface glass micro-beads (obtained from Polytec Ltd) with diameters ranging from 45 μm to 63 μm. The use of glass beads offered the advantage of greater marker sphericity in comparison to a wrapped tape version. The method used to wrap the tape leads to an elliptical marker with a non-smooth surface that would result in unwanted offsets. The micro beads were bonded using a 3M PhotoMount spray which allowed a thin layer of adhesive to be sprayed onto the surface of the bearing. The optimal photomount layer thickness with respect to the strength of retroreflection is half the mean diameter of the micro-bead: this thickness leaves exposed half of a micro-bead to fulfil the reflection function of the target.

A technique was developed to produce an even surface coverage of both the photomount and micro-beads, in total five targets were created. The 1.5" bearing used as the basis of the custom target is shown in Figure 3.13 (a) in its original condition. Application of the glass micro-beads is shown in Figure 3.13(b) in which the surface has become dulled with respect to Figure 3.13(a). The target retroreflection capability is shown in Figure 3.13(c) under illumination from a camera flash. Shot blasting the ball bearing prior to application of the micro-beads was also investigated but it was found that the original shiny surface of the bearing worked best.

It was found in practice that the ball bearing based markers displayed a bright spot in the centre of the surface and that the beads did not reflect the IR light as strongly as the taped based markers. As a result the threshold for each camera had to be adjusted carefully to ensure marker visibility at each measurement location during the experiment.

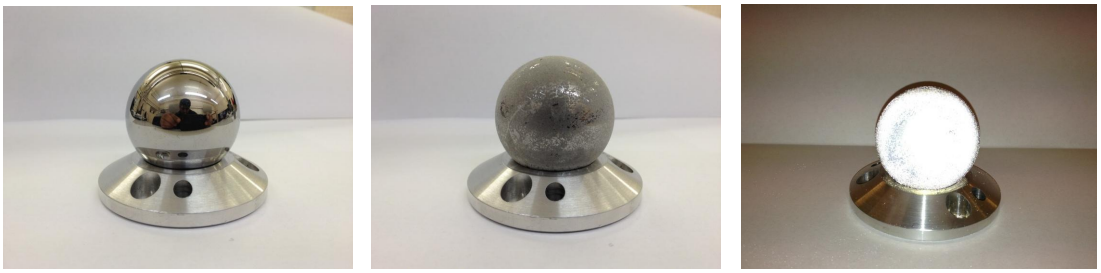


Figure 3.13: Custom ball bearing markers (a) Clean 1.5" ball bearing (b) Glass micro beads applied to ball bearing (c) Retroreflective behaviour of micro beads responding to a camera flash

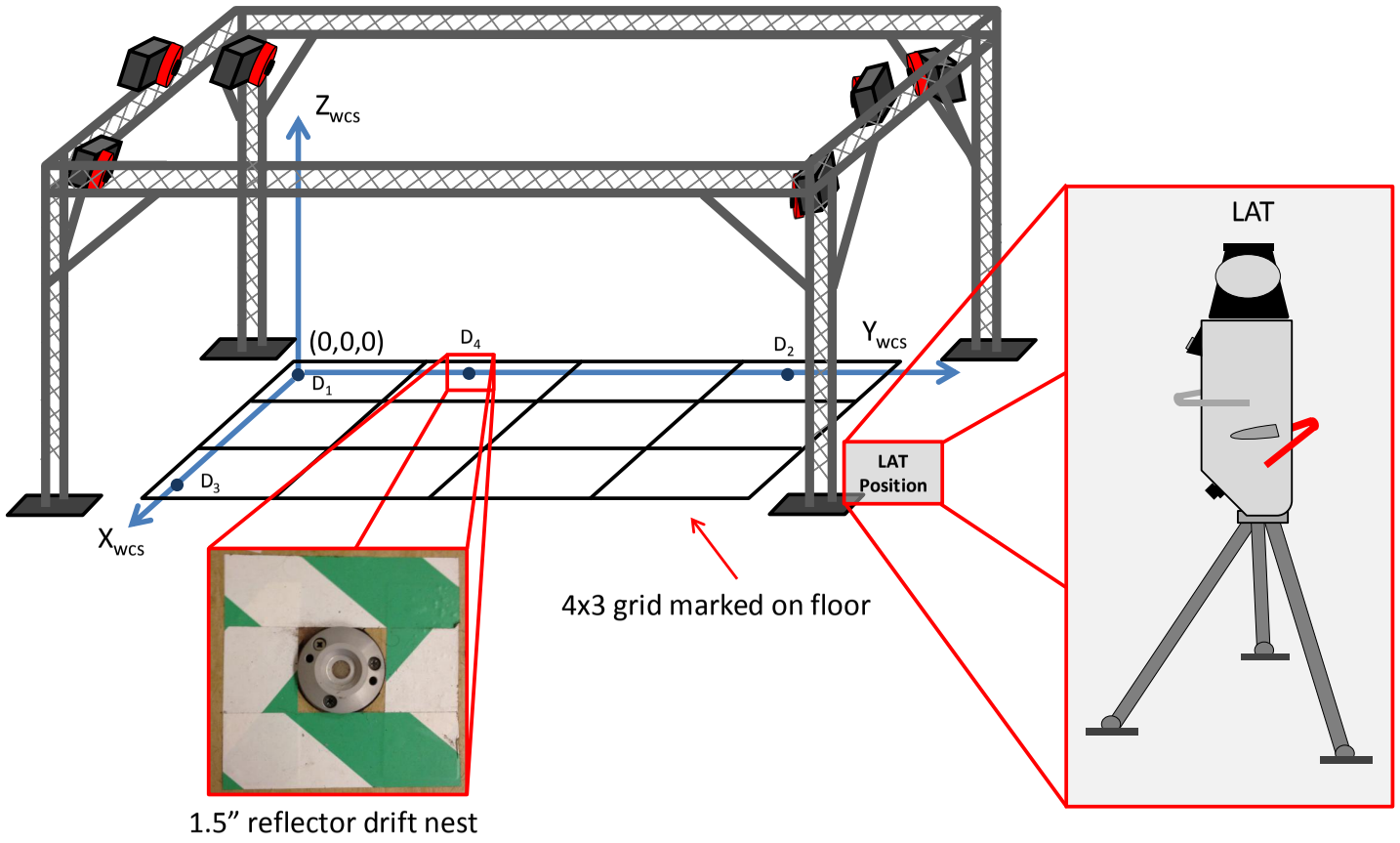


Figure 3.14: Experimental Setup. A 4×3 grid was marked out on the floor serving to guide acquisition to take the form of an approximate grid. Drift nests $\mathbf{D}_i$ for $i = 1 \ldots 4$ were securely mounted to the floor in an L-shape to create the new joint coordinate frame of which $\mathbf{D}_1$ was the origin.

3.3.5 Defining a Joint Coordinate Frame

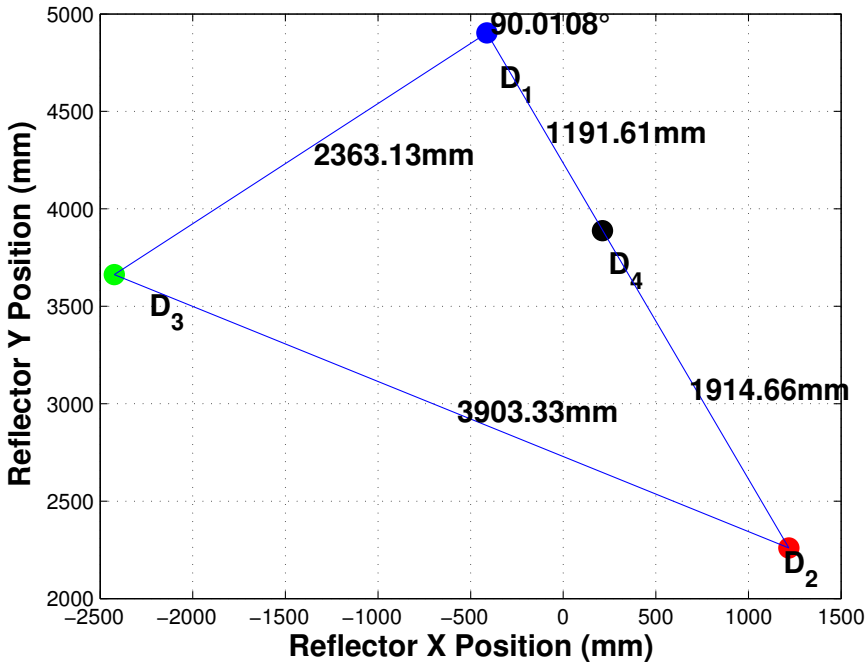


Figure 3.15: Arrangement of ground mounted drift nests forming the joint coordinate frame as measured by the LAT.

Since the VMCS and LAT measure with respect to their native frames of reference, the raw point cloud data resulting from the experiment would require transformation by the fixed rotation, R, and translation, t, relating the two frames. This operation could be carried out post experiment through a least squares procedure to minimise the error between the point clouds as a function of R and t. Alternatively the systems could be calibrated prior to the experiment such that they measure in the same coordinate frame. The latter approach was adopted in order facilitate data collection by making the measurements directly comparable during the experiment thus allowing invalid measurements to be reacquired systematically.

The coordinate frames for the two systems were made coincident through defining a new joint World Coordinate System (WCS), F_{WCS}, with dimensions comparable to the measurement volume. This new frame was defined by a set of drift nests rigidly affixed to specific locations on the floor within the volume: a schematic of the arrangement is shown in Figure 3.14. Use of the drift nests importantly enabled the *same* calibration artifact to be used with targets of the *same* dimensions for both systems. The arrangement of drift nests had to adhere to two rules observed by the VMCS specifically with respect to the number of targets used to define the frame and the inter-target separation. The minimum number of targets that can be used to define a VMCS frame is four while the inter-target distance within this frame must differ. Under guidance of the LAT, four drift nests were mounted onto the floor in an L-shape configuration, arranged so as to maximise inter-drift nest distances across the volume whilst keeping them in view of the cameras.

The procedure consisted of firstly mounting the drift nests, $\mathbf{D}_1$ and $\mathbf{D}_2$, to the floor in order to define the y-axis of the frame. A third nest, $\mathbf{D}_3$, was then introduced to define the x-axis and positioned so as to create an approximate right angle which was subsequently fine tuned using feedback from the LAT. Through placement of the RRR into each drift nest, the location $\mathbf{D}_i = (D_{i,x}, D_{i,y}, D_{i,z})$ of each nest was measured and used to compute the set of relative distances $\{d_{ij}|i,j = 1...4, i \neq j\}$ as shown in Figure 3.15. These distances were input into Equation 3.1 as follows:

$$\alpha = \cos^{-1}\left(\frac{d_{23}^2 - d_{12}^2 - d_{13}^2}{-2d_{12}d_{13}}\right) \tag{3.1}$$

to yield the angle between x and y axes. Following this, the angular error $\alpha_{err} = (90^\circ - \alpha)$ was calculated and used to pivot the y-axis around the origin $\mathbf{D}_1$ via Equation 3.2:

$$\mathbf{D}_{90^\circ} = \begin{bmatrix} \cos(\alpha_{err}) & -\sin(\alpha_{err}) & D_{1,x}(1-\cos(\alpha_{err})) + \sin(\alpha_{err})D_{1,y} \\ \sin(\alpha_{err}) & \cos(\alpha_{err}) & D_{1,y}(1-\cos(\alpha_{err})) - \sin(\alpha_{err})D_{1,x} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} D_{2,x} \\ D_{2,y} \\ 1 \end{bmatrix} \tag{3.2}$$

Corrections offsets in the LAT coordinate system were then calculated using Equation 3.3:

$$\mathbf{D}_{offsets} = \mathbf{D}_{2,(x,y)} - \mathbf{D}_{90^\circ,(x,y)} \tag{3.3}$$

where the notation $\mathbf{D}_{i,(x,y)}$ indicates use of only the (x, y) components of the coordinate. The angular error was minimised iteratively using the above procedure following multiple measurements. The final arrangement of the drift nests is shown in Figure 3.15 where an inter-line angle of 90.01° was achieved. The new frame was completed by the addition of a fourth drift nest, $\mathbf{D}_4$, that was situated colinearly between $\mathbf{D}_1$ and $\mathbf{D}_2$. Colinearity was measured through calculating the area of the triangle resulting from the preceding drift nests using Equation 3.4 as follows:

$$\lim_{\mathbf{D}'_4 \to \mathbf{D}_4} \frac{|D_{1,x}(D_{2,y} - D'_{4,y}) + D_{2,x}(D'_{4,y} - D_{1,y}) + D'_{4,x}(D_{1,y} - D_{2,y})|}{2} = 0 \tag{3.4}$$

where the area vanishes to zero when the current position of the fourth drift nest, $\mathbf{D}'_4$, assumes the ideal position, $\mathbf{D}_4$. The final area calculated following iterative refinement of the position was $0.0017m^2$. The following sections describe the transformation procedure from the system native frames into the new joint frame F_{WCS}.

Aligning the VMCS with F_{WCS}

After precisely aligning the drifts nest to form the new joint world coordinate system the nest locations measured in the frame of the LAT, F_{LAT}, were transformed into F_{WCS} wherein $\mathbf{D}_1$ was designated as the origin; these locations are shown in Table 3.1 alongside the original coordinates in F_{LAT}. In order for the VMCS to make use of this custom calibration artifact in the Vicon Tracker software, an Extensible Markup Language (XML) based Vicon Skeleton (VSK) file was created. A MATLAB script was written to generate the necessary XML which defined the positions and connectivity amongst the targets in addition to parameters such as target radius, intensity thresholds and positional variance. Given the VSK file, setting F_{WCS} as the world coordinate frame reduced to selecting the new calibration artifact in the calibration phase of system operation. The image errors resulting from the calibration are as shown in Table 3.2.

	F_{LAT}			F_{WCS}		
Target	X (mm)	Y (mm)	Z (mm)	X (mm)	Y (mm)	Z (mm)
1	-412.49	4902.88	-1378.75	0	0	0
2	1218.46	2259.26	-1390.38	-0.57	3106.24	-11.63
3	-2423.91	3662.47	-1376.40	2363.13	0	2.36
4	212.26	3888.18	-1379.54	0.85	1191.61	-0.78

Table 3.1: RRR coordinates as measured in F_{LAT} and specifed in F_{WCS}

Camera	Image Error (Pixels)
1	0.34
2	0.30
3	0.36
4	0.33
5	0.30
6	0.27

Table 3.2: Camera Image Errors Following Calibration

Aligning the LAT with F_{WCS}

The transformation procedure from F_{LAT} into F_{WCS} consisted of a least squares minimisation of the rigid body mapping of the set of RRR positions measured in F_{LAT} onto the set of positions in the frame of F_{WCS} shown in Table 3.1. The mapping from F_{LAT} to F_{WCS} may be expressed as follows:

$$T(\mathbf{x}) = \mathbf{t} + \frac{1}{s}R\mathbf{x} \tag{3.5}$$

where $T(\mathbf{x})$ is the transformed point $\mathbf{x}$, R and $\mathbf{t}$ are the 3×3 rotation matrix and translation vector relating the two frames respectively while, s, is a temperature related scale parameter in this case set to unity. Point measurements acquired by the LAT are the result of several measurements acquired internally thus allowing an associated covariance matrix to be calculated: the covariance matrix is available through enabling the extended statistics mode in EmScon and has the form:

$$C_p = \begin{bmatrix} C(x,x) & C(x,y) & C(x,z) \\ C(y,x) & C(y,y) & C(y,z) \\ C(z,x) & C(z,y) & C(z,z) \end{bmatrix} \tag{3.6}$$

where $C(x, y) = E[(x - \mu_x)(y - \mu_y)]$ and E is the expectation operator. The algorithm used to determine the R and $\mathbf{t}$ parameters was implemented onboard the AT Controller 900 which took as input the two point sets of Table 3.1 alongside the covariance matrices associated with the LAT measured points. The algorithm seeks to minimise the Residual Square Sum (RSS) defined by the following function:

$$RSS = \sum_{i=1}^{n} r_i^T W_i r_i \tag{3.7}$$

where r_i and W are the i^{th} residual and weight matrix respectively. The weight matrix is the inverse of the covariance matrix of the residual. The residual is calculated as follows:

$$residual = T(actual) - nominal \tag{3.8}$$

where actual is the point measured in F_{LAT} and F_{WCS}. The parameters calculated by the procedure are shown in Table 3.3.

Parameter	Unit	Optimised Value	σ
t_x		-412.49	0
t_y	(mm)	4902.88	0
t_z		-1378.75	0
r_x		0	0
r_y	∘	0	0
r_z		-148.34	0
s	N/A	1	0
RMS	(mm)	0	N/A
Max Deviation	(mm)	0	N/A
Variance Factor	N/A	0	N/A

Table 3.3: Summary of transformation parameters shown to 6 significant figures

3.3.6 Acquisition Software

A custom software package was implemented using the C# programming language to facilitate synchronous/asynchronous measurement acquisition from the tracking systems. A screenshot of the developed Graphical User Interface (GUI) is shown in Figure 3.16. The control software pertaining to each system offered server functionality to allow positional data to be broadcast over network connections. The GUI then acted as a client reading streams containing object position data via the Vicon Tracker and Leica EmScon Dynamic-link library's (DLL) associated with the respective systems: a block diagram illustrating the underlying structure of the software is shown in Figure 3.17. The program was written to allow control of some basic LAT features such as initialisation and measurement mode and reflector type selection as well enabling coordinate frame transforms. A grid mode capability was included to assist the user in spatially sampling the measurement volume at regular intervals: this functionality was used during the experiment.

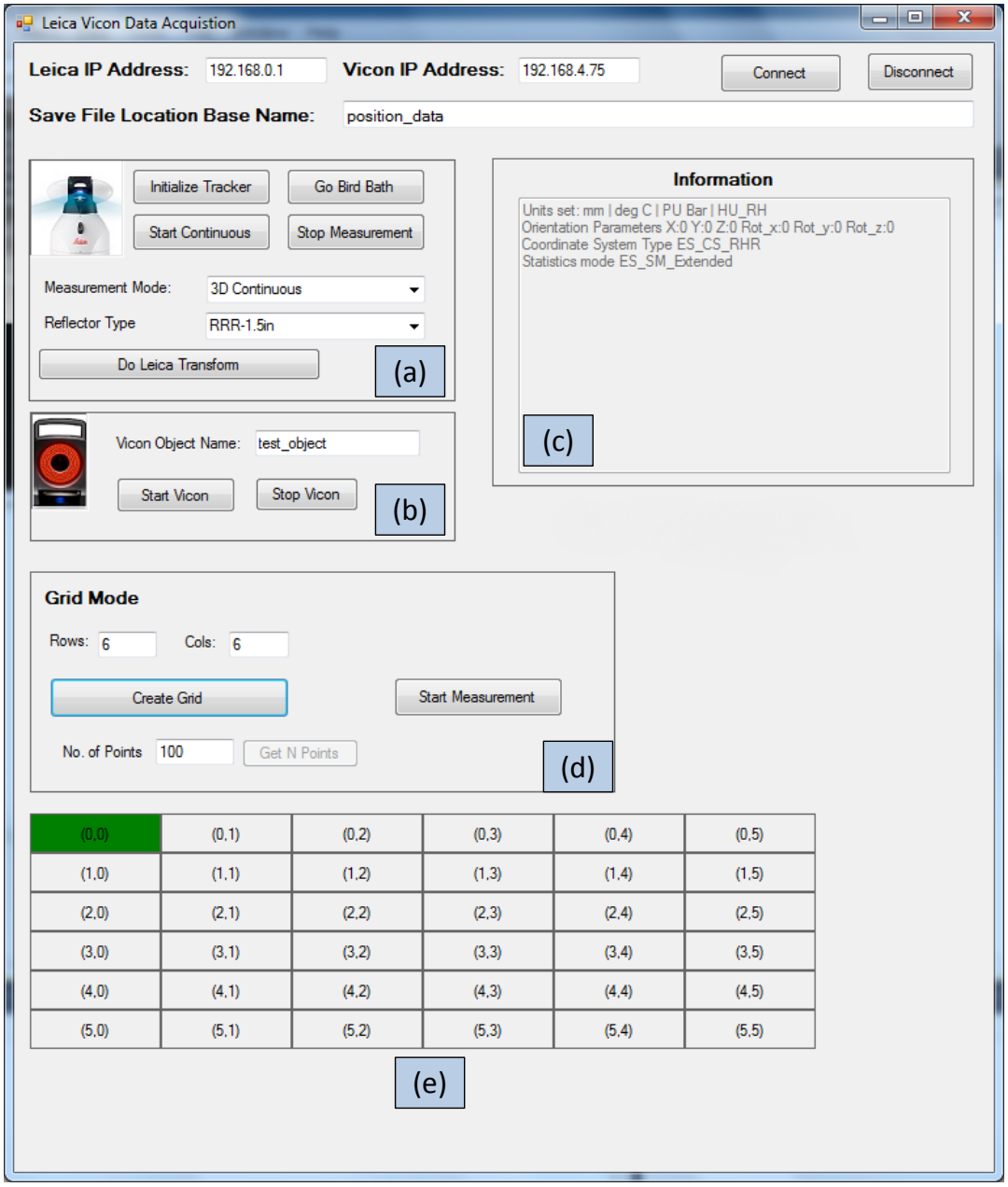


Figure 3.16: Screenshot of Leica Vicon Measurement Acquisition GUI (a) LAT controls allowing control of basic features and the coordinate transform procedure (b) Vicon controls allowing the virtual object name to be set and filtered from the input stream (c) LAT information window (d) Grid mode parameter settings where the number of points/location can be specified (e) Visual indication of current grid cell in which acquisition is taking place

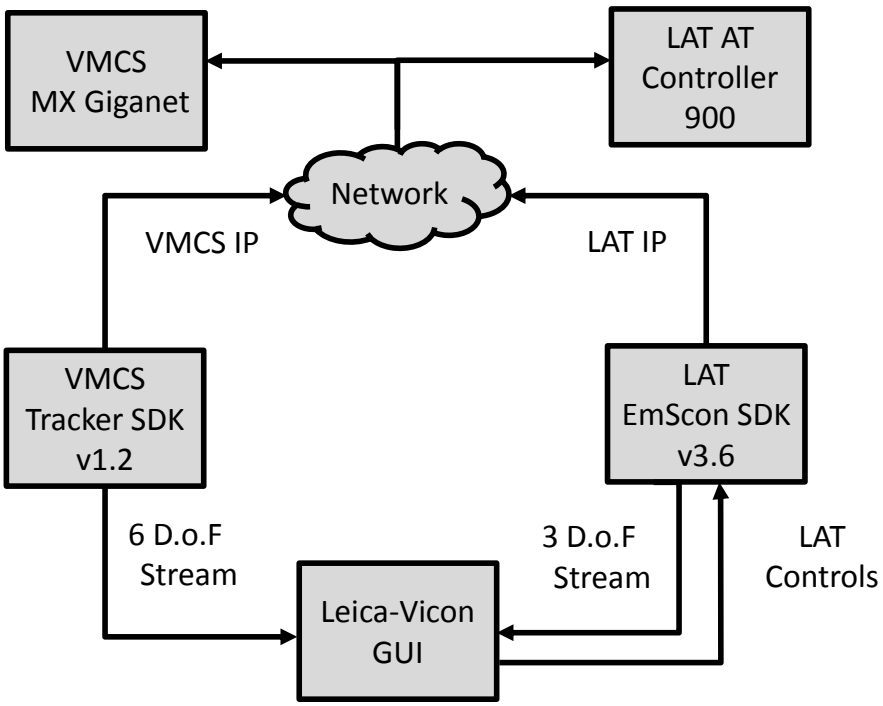


Figure 3.17: Block diagram showing underlying structure of GUI

3.3.7 Scanning System

A scanning mechanism was required to sweep the test object through the measurement space. A 3-axis XYZ scanner of dimensions comparable to the measurement volume would ideally fulfil this function, however, implementation of this was infeasible due to the associated scale and cost, therefore alternative approaches were considered. The criterion of critical importance in making this selection was concerned with the stability of the scanning system during measurement acquisition: ideally it would be subject to little or no movement. The LAT would provide the ground truth location of the test object in the volume therefore movement was not constrained to be recti1linear thus increasing the number of possible scanning methods.

The approach taken was to use a tripod to provide the scanning function. A professional grade Manfrotto 161MK2 tripod was selected as this particular tripod is an option offered by Vicon for mounting cameras due to its high rigidity. This approach would offer single user operation and rapid scanning. The stability of the tripod for carrying out measurements was considered with respect to the fixed frame through measuring the location of a reflector attached to the frame and the tripod in turn. The measurement data in the XY plane for the tripod is shown in Figure 3.18 while Figure 3.19 shows the measurements taken on the frame. The measurements in X and Y for the frame show a negative correlation while the tripod displays no correlation. The correlation can be attributed to the ambient vibrations in the laboratory causing vibration of the frame which is directly bolted to the floor. The tripod, however, has rubber feet providing isolation from the vibration.

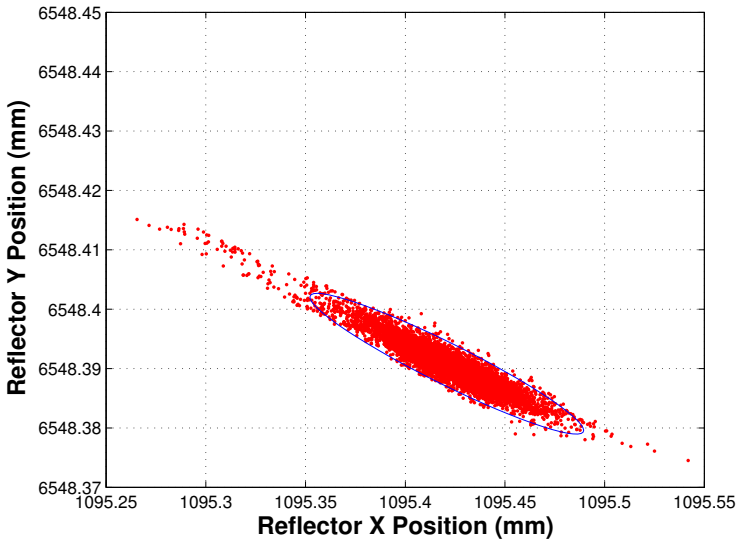


Figure 3.18: 10000 measurements of a reflector attached to the fixed frame.

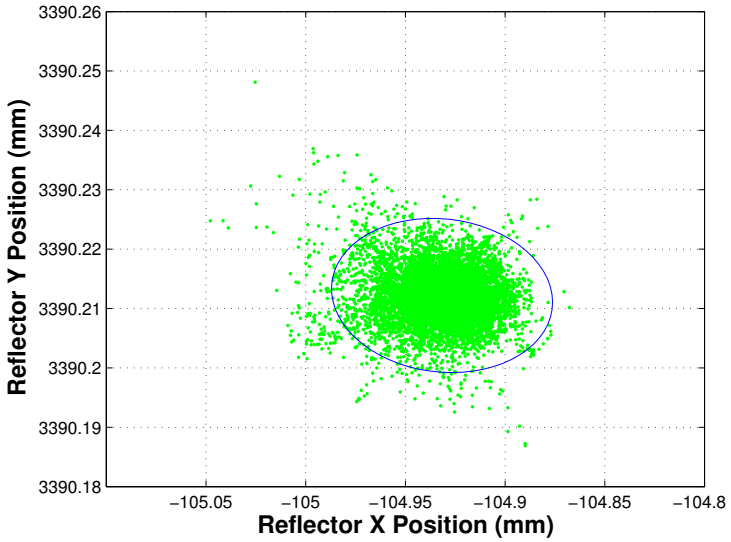


Figure 3.19: 10000 measurements of a reflector attached to the tripod.

In both cases the data are bounded by a three standard deviation uncertainty ellipse. The covariance matrices corresponding to the data are as follows:

$$C_{frame} = \begin{bmatrix} 523.7334 & -85.5730 & -8.4355 \\ -85.5730 & 15.6139 & -22.6894 \\ -8.4355 & -22.6894 & 393.5643 \end{bmatrix}$$

$$C_{tripod} = \begin{bmatrix} 341.5980 & -7.0513 & 47.7879 \\ -7.0513 & 18.7355 & -51.5729 \\ 47.7879 & -51.5729 & 144.5779 \end{bmatrix},$$

where the units are in μm^2. From the diagonal entries of the covariance matrices which correspond to the positional variance in each axis, the tripod is subject to smaller perturbations than the frame therefore was considered to be the best option. Note that the reflector is subject to less variation in the x-axis in comparison to the y-axis in both cases. The graphs have been plotted to show the distribution of the measurements where the frame data displays a negative cor-

relation between the x and y axes while the tripod data yields little correlation between the same axes. The tripod was used at three different height settings as illustrated in Figure 3.20. The tripod was selected as the scanning mechanism due to flexibility it offered.

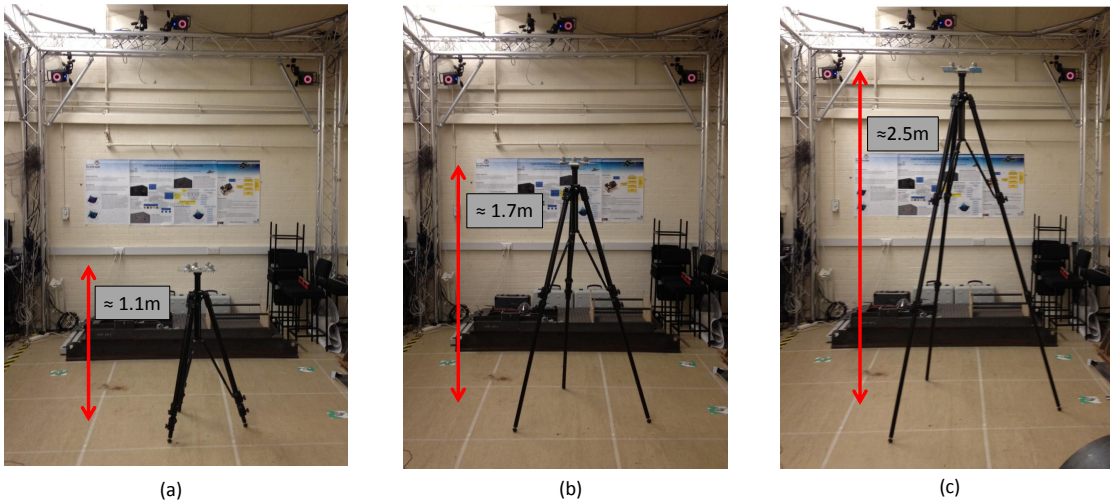


Figure 3.20: Different tripod height settings

3.3.8 Drift Nest Stability Analysis

The stability of the floor mounted drift nests forming the joint coordinate frame was also considered. The variances of each drift nest in the X, Y and Z axes are summarised in Table 3.4. The values of variance show a high level of stability for the drift nests.

3.3.9 Test Object Design

A test object had to be designed to sweep through the measurement space. The design had to be such that a common point could be tracked by both systems in order to compare accuracy. An arrangement of drift nests mounted onto an

Drift Nest	$\sigma_x^2(\mu m^2)$	$\sigma_y^2(\mu m^2)$	$\sigma_x^2(\mu m^2)$
1	70.71	15.87	104.54
2	161.89	40.60	71.56
3	91.02	29.79	102.25
4	105.51	8.03	64.73

Table 3.4: Summary of drift nest variances

aluminium plate was used as the test object. The locations of each nest could be measured precisely using the LAT therefore accurate alignment during manufacture was not necessary, these position are shown in Figure 3.22. The test object was mounted on top of the tripod in order to minimise target occlusion.

In total five drift nests were mounted onto the plate, where four would be utilised by VMCS and the fifth would be used by the laser - the test object is shown in Figure 3.21. The centroid of the object defined in the Vicon software was set equal to the location of the RRR occupying the central location on the plate. The virtual object is shown in Figure 3.23. Just as with the new world coordinate frame, the relative locations of the markers were calculated.

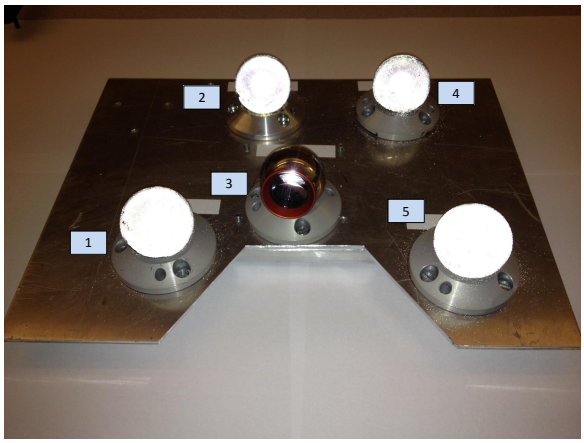


Figure 3.21: Test object containing coplanar targets. The centroid of the tracked object coincides with the location of the RRR.

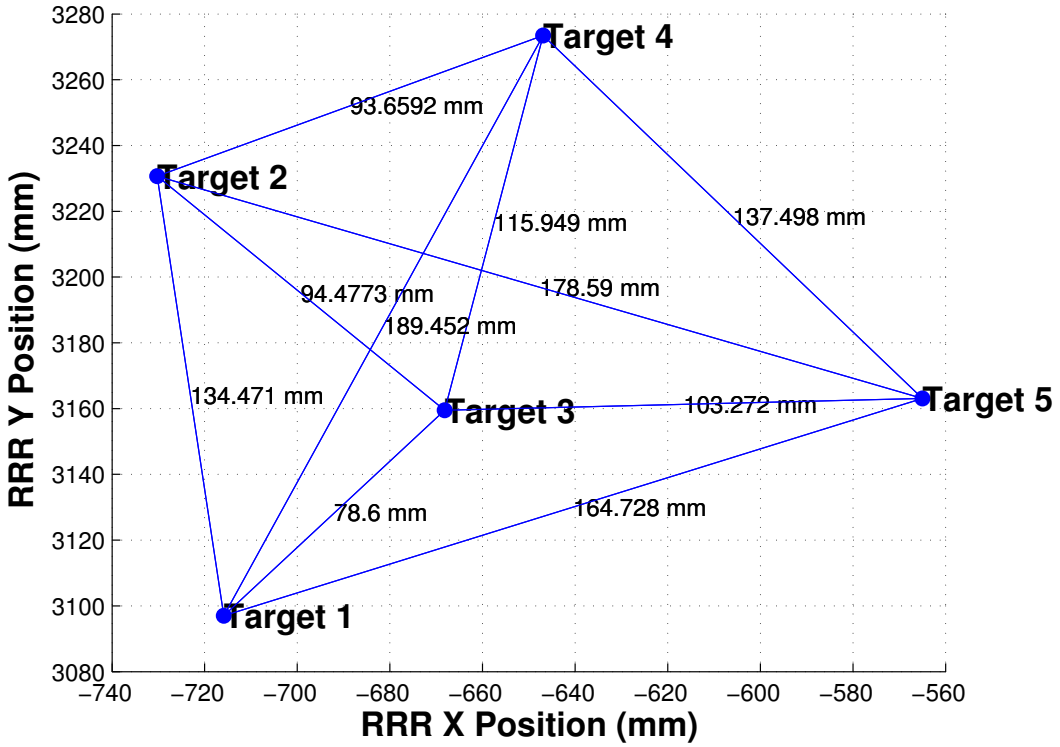


Figure 3.22: Relative measurements of target positions used to create Vicon object XML file

The location of the RRR was set to the tracked point in the Vicon Tracker software via the custom defined XML based upon the LAT measurements.

The strength of reflection from the custom retroreflectors is shown in Figure 3.24 using a binary image from one of the cameras.

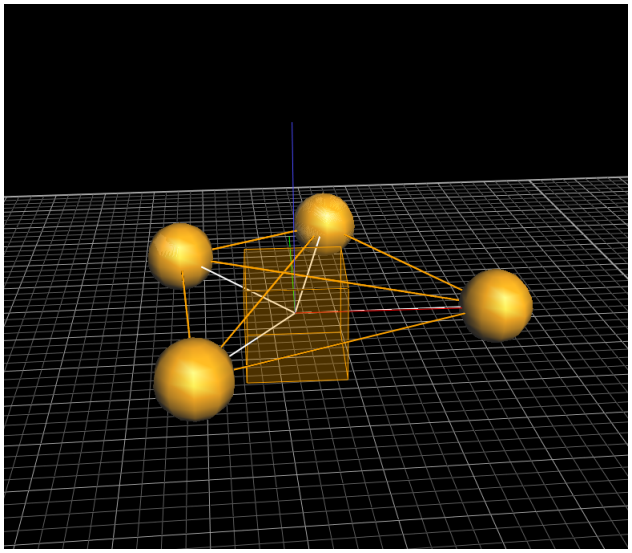


Figure 3.23: Virtual representation of the test object in which the underlying XML file has been defined by the LAT

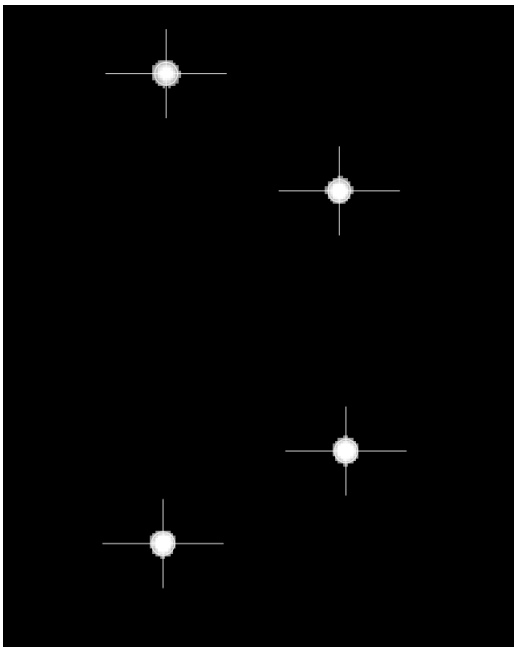


Figure 3.24: Recognition of custom markers by VMCS cameras

3.3.10 Method

A 4×3 grid was marked on the floor serving to help spatially sample the measurement space at approximately regular intervals on each level as specified by the tripod height setting. In total data in four layers were captured labelled $0-4$ at heights of approximately 0.1 m, 1.1 m, 1.7 m and 2.5 m respectively. The grid, whose perimeter was comparable to the footprint of the fixed frame, is shown in Figure 3.25 wherein each large grid cell was further divided into 4 cells giving rise to 48 sampling points/layer, in addition the grid was aligned with the axes of the world coordinate frame. The test object was placed in row 0, col 0 at the beginning of the experiment and was subsequently scanned column-wise through the grid. Upon completion of a layer the tripod was set to the next level and the test object was returned to row 0, col 0 to commence collection of the next set of measurements. A plot of the resultant data which takes the form of a grid is shown in Figure 3.26, where the measurements have been acquired in the joint coordinate frame. Note that due to line-of-sight issues with measurements of the test object close to the LAT, some data on layers 0 and 3 are absent. Layers 1 and 2 contain 48 measurement locations while layers 0 and 3 contain 42 and 30 measurement locations respectively. At each acquisition location about 100 points were acquired.

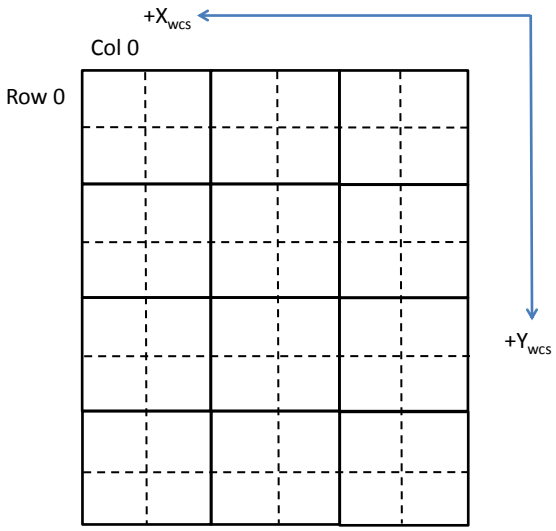


Figure 3.25: Measurements were acquired according to the grid marked on the floor giving rise to 48 samples/layer.

The threshold of each camera was adapted as a function of the height of the test object in the volume: this was necessary due to the strength of the reflection of the targets failing the circle fitting process carried out on-board each camera.

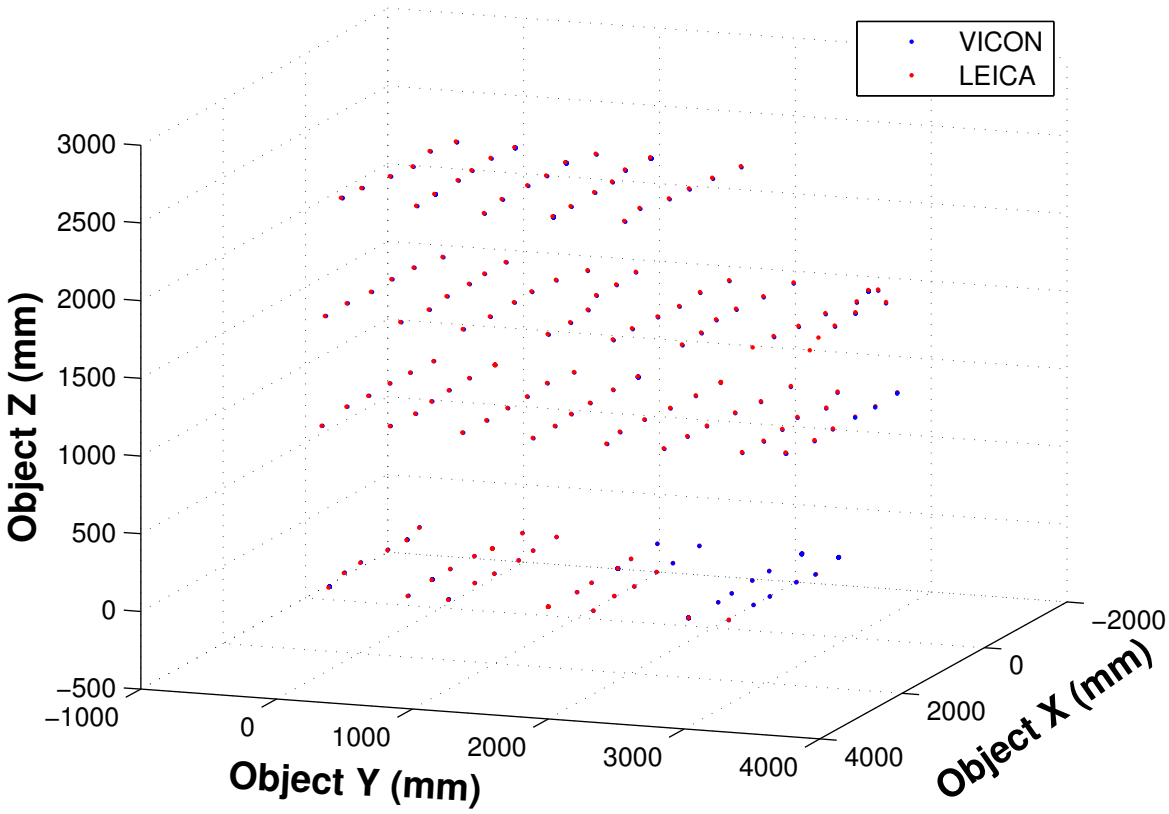


Figure 3.26: Plot of the raw data showing four layers in which measurements were acquired.

For each measurement location the error vector between the mean of the VMCS and LAT data is plotted in Figure 3.27. At each location the vector points to the LAT data. Interestingly a non-random behaviour is displayed in the data and indicates that the VMCS measurements are contracted with respect to the LAT data.

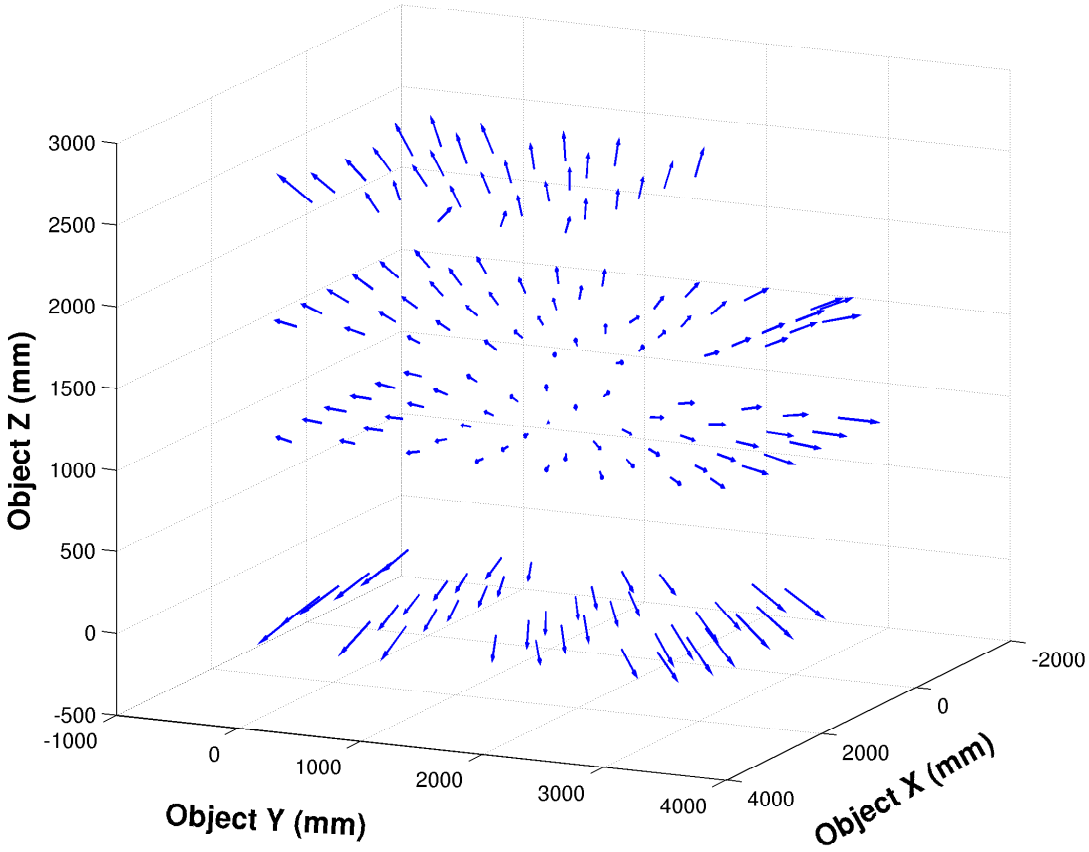


Figure 3.27: Plot of the error vectors

3.3.11 Results

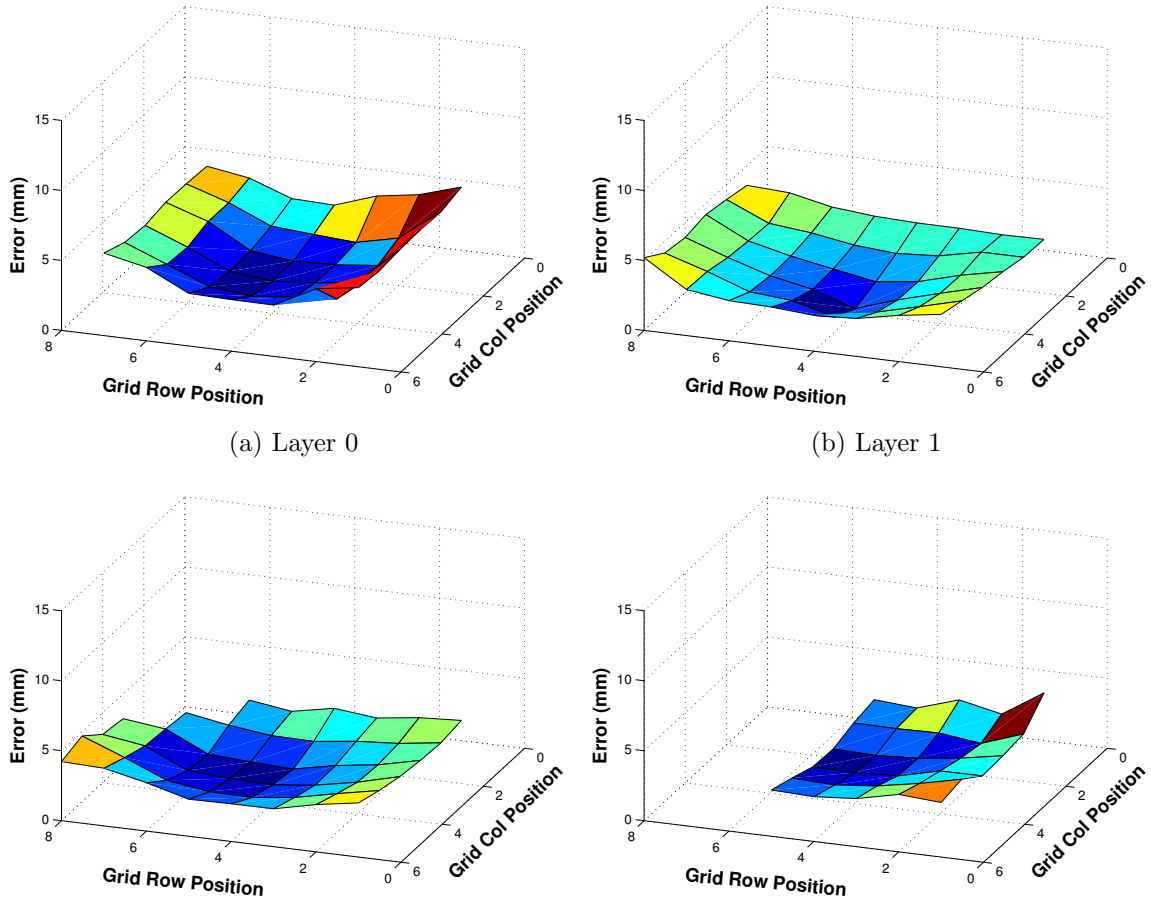
3.3.11.1 Accuracy of Vicon Measurements

The Euclidean distance error of the VMCS with respect to the LAT is calculated as follows:

$$e_{(r,c)} = ||\bar{\mathbf{p}}^{\mathrm{VMCS}} - \bar{\mathbf{p}}^{\mathrm{LAT}}|| = \sqrt{(\bar{\mathbf{p}}^{\mathrm{VMCS}} - \bar{\mathbf{p}}^{\mathrm{LAT}}).(\bar{\mathbf{p}}^{\mathrm{VMCS}} - \bar{\mathbf{p}}^{\mathrm{LAT}})} \qquad (3.9)$$

where $\bar{\mathbf{p}}^{\mathrm{VMCS}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{p}_i^{\mathrm{VMCS}}$ and $\bar{\mathbf{p}}^{\mathrm{LAT}} = \frac{1}{M}\sum_{i=1}^{M}\mathbf{p}_i^{\mathrm{LAT}}$ are the mean values of each measurement cluster containing N and M points respectively. The plots of Figures 3.28a through 3.28d show positional error for layers 0 through 3 respectively. Empirical observations of error as a function of location within the measurement space indicated greater error at the edges of the volume in comparison to its centre suggesting the error function would assume a parabolic form. This shape is borne out in Figure 3.28a showing the error for the test object as placed on the floor of the volume and to a lesser extent in the plots of Figures 3.28b through 3.28d. The layer error maximum for each graph was as follows 6.06 mm, 5.19 mm, 4.69 mm and 4.95 mm.

If the grid is considered to be composed of concentric rings, R_i, for $i = 0 \ldots 2$ as shown by Figure 3.31 the error at the edges of the volume may be compared to the inner regions. Fixing the layer under consideration L_j and plotting the error attributed to each R_i results in the graphs of Figures 3.29a through 3.29d for $j = 0 \ldots 3$. Note that the different curves lengths result from the descending number of contributing cells in going from the outer to the inner ring of the grid. Figure 3.29c contains missing data as a consequence of the test object being in a dead space spot with respect to the VMCS. If corresponding R_i's are plotted through each layer, Figures 3.30a through 3.30c result. The plots suggest that the upper and lower most layers exhibit higher error than the intermediate layers 1 and 2 and thus the centre of the measurement volume has better tracking accuracy than the surrounding regions.



(a) Layer 0

(b) Layer 1

(c) Layer 2

(d) Layer 3

Figure 3.28: Surface plots of error as a function of grid location in measurement space

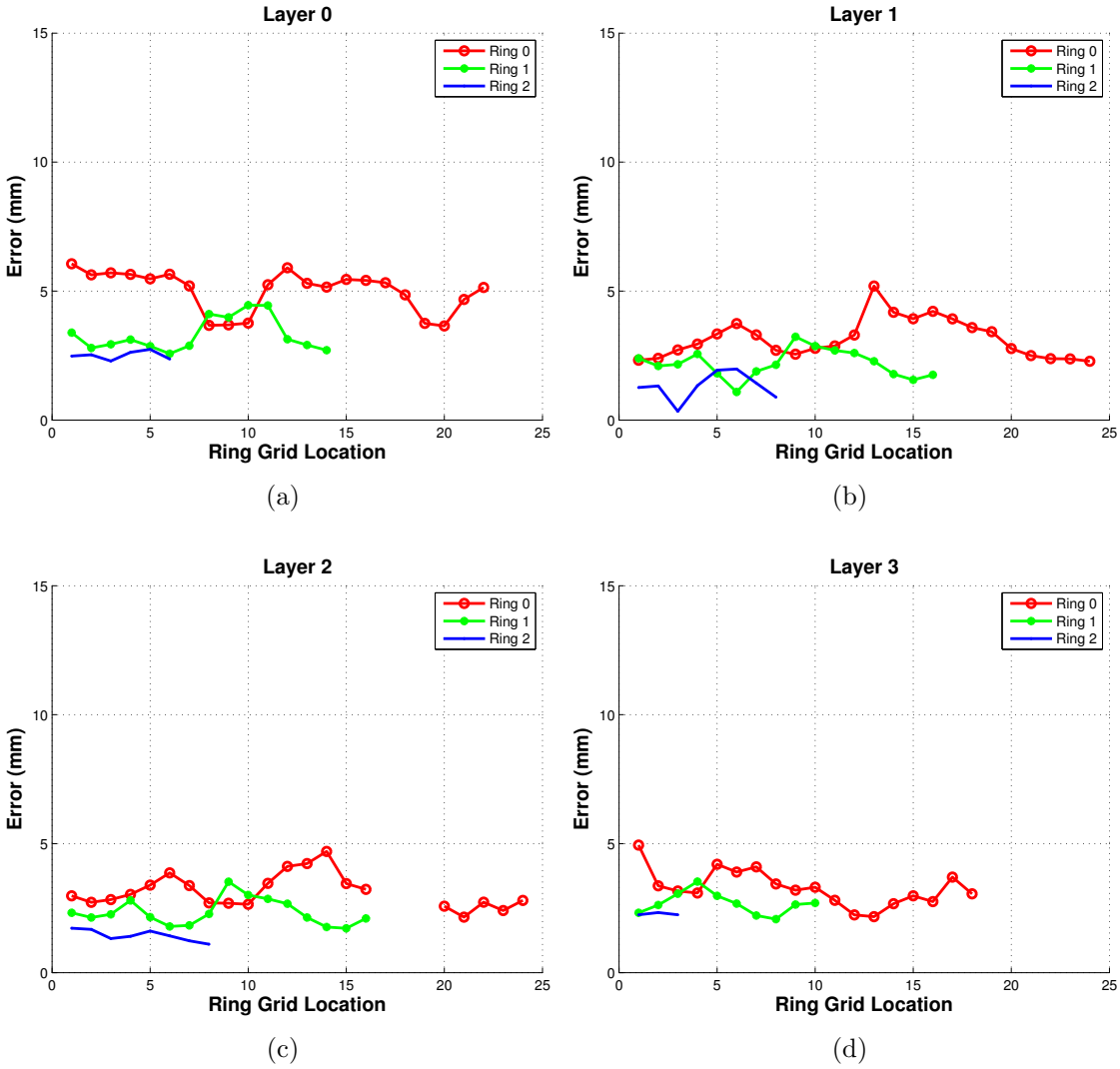


Figure 3.29: Surface plots of error as a function of grid location in measurement space (a) Layer 0 (b) Layer 1 (c) Layer 2 (d) Layer 3

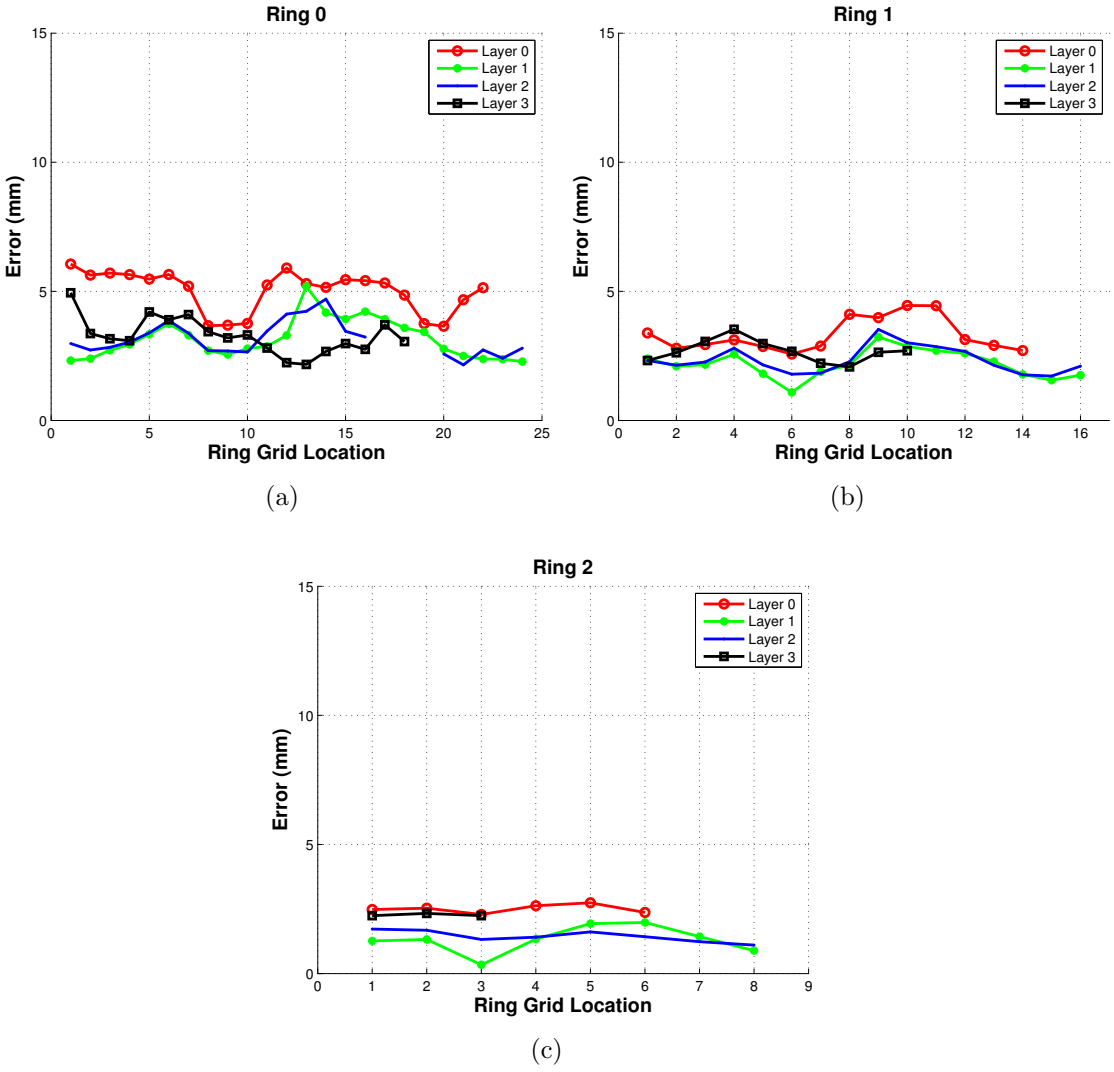


Figure 3.30: Plots of the same ring data for as a function of the layers (a) Ring 0 (b) Ring 1 (c) Ring 2

3.3.11.2 Uncertainty in Vicon Measurements

The standard deviation associated with the measurement clusters acquired at each grid location is considered in this section where the standard deviation, σ_x, is calculated as follows:

$$\sigma_x = \sqrt{\frac{1}{(n-1)} \sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{3.10}$$

where x_i is a sample and $\bar{x}$ is the mean resulting from n samples. Evaluation of the preceding equation through holding constant the ring, R_i, under consideration the standard deviation data corresponding to layers L_j for $j = 0 \ldots 3$ is plotted for $i = 0 \ldots 2$ in Figures 3.32, 3.33 and 3.34 for the X, Y and Z axes respectively. Viewing the data in this form allows consideration of the variance as a function of height and it can been seen that the mean levels appear similar discounting the the outlier data. The outlier data was due to the VMCS failing to track consistently in some locations due low target reflectance. In order to compute the mean standard deviation for each curve the data associated with each curve was filtered to remove these outliers through retaining those points, s_i, which satisfy the condition

$$(s_i - \bar{s}) \leq 2\sigma_s, \tag{3.11}$$

where $\bar{s}$ is the mean level of the curve and σ_s is the standard deviation of the curve. Repeating the above procedure but plotting the mean of the filtered data yields the graphs of Figure 3.35. The maximum and minimum standard deviations are 74.39 μm and 4.97 μm respectively over the entire volume, this information is summarised in a per axis basis in $\bar{\sigma}_X$, $\bar{\sigma}_Y$ and $\bar{\sigma}_Z$ in Figure 3.36.

If each layer L_j is considered in isolation and the mean filtered data pertaining to rings, R_i for $i = 0 \ldots 2$ is plotted then Figure 3.37 results where the rows iterate through $j = 0 \ldots 3$. Analysing the data in this way allows the deviations

in the outer ring to be compared to those of the inner rings. Given the empirical observations of greater tracking error at the fringes of the measurement space, it was expected that the deviations would follow a similar parabolic mapping with respect to the rings. Figure 3.37 (a), (c), (e), (f), (g), (h), (i), (k) and (l) validate these observations in showing that the size of the mean standard deviations descend in going from rings R_0 to R_2. The graphs in (d) and (j) display higher deviations in the Y components of the measurements in comparison to X while (b) shows higher deviation for the Z component over the Y component.

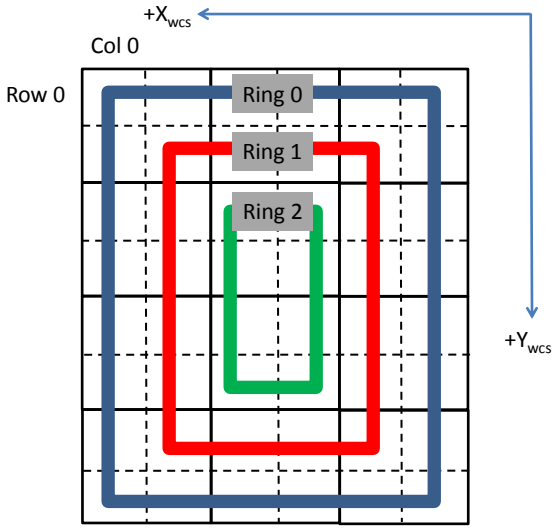


Figure 3.31: Viewing the measurement space as concentric rings facilitating comparison of the outer regions to the inner regions of the measurement volume.

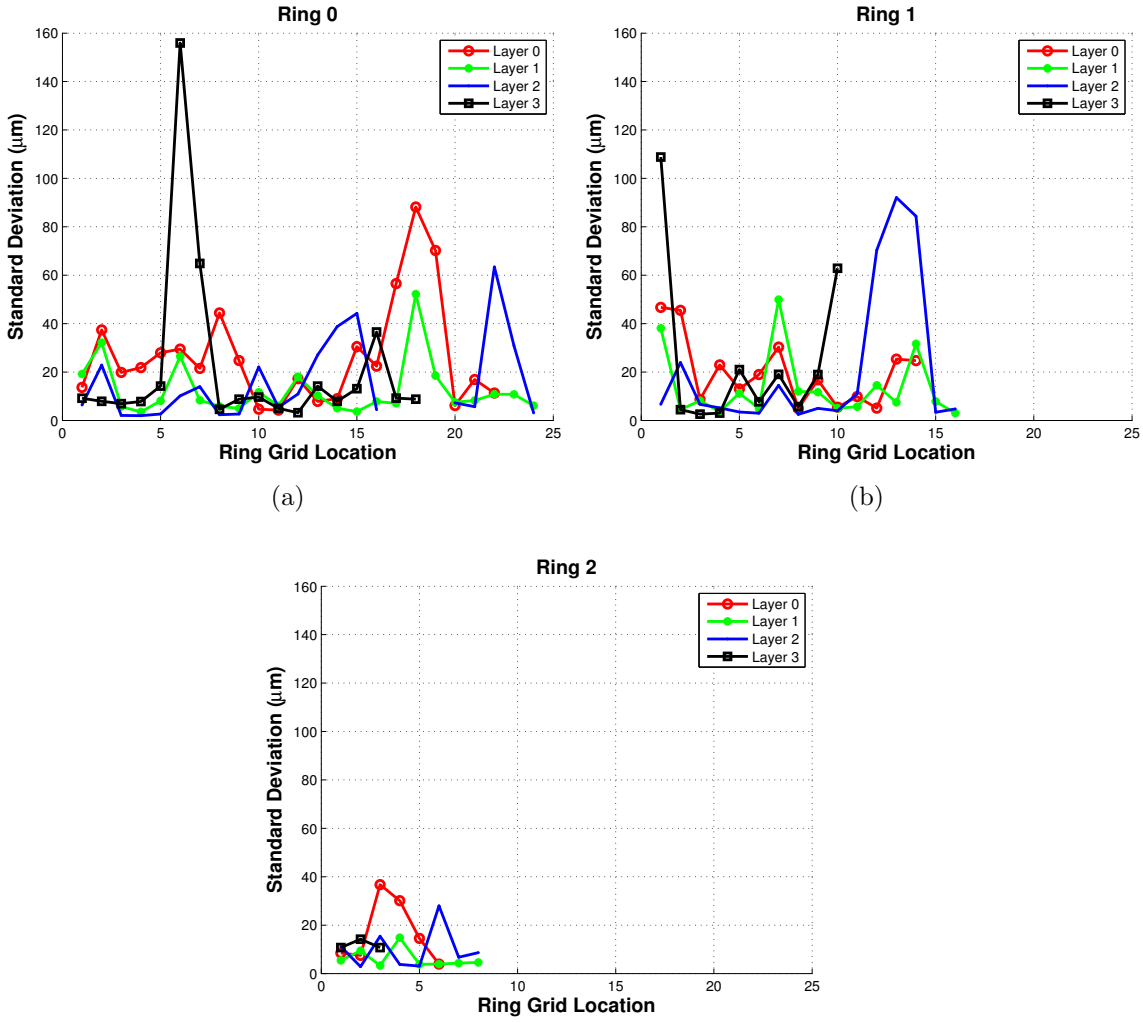


Figure 3.32: σ in the X-axis expressed in terms of the concentric ring composing the layers (a) Ring 0 (b) Ring 1 (c) Ring 2

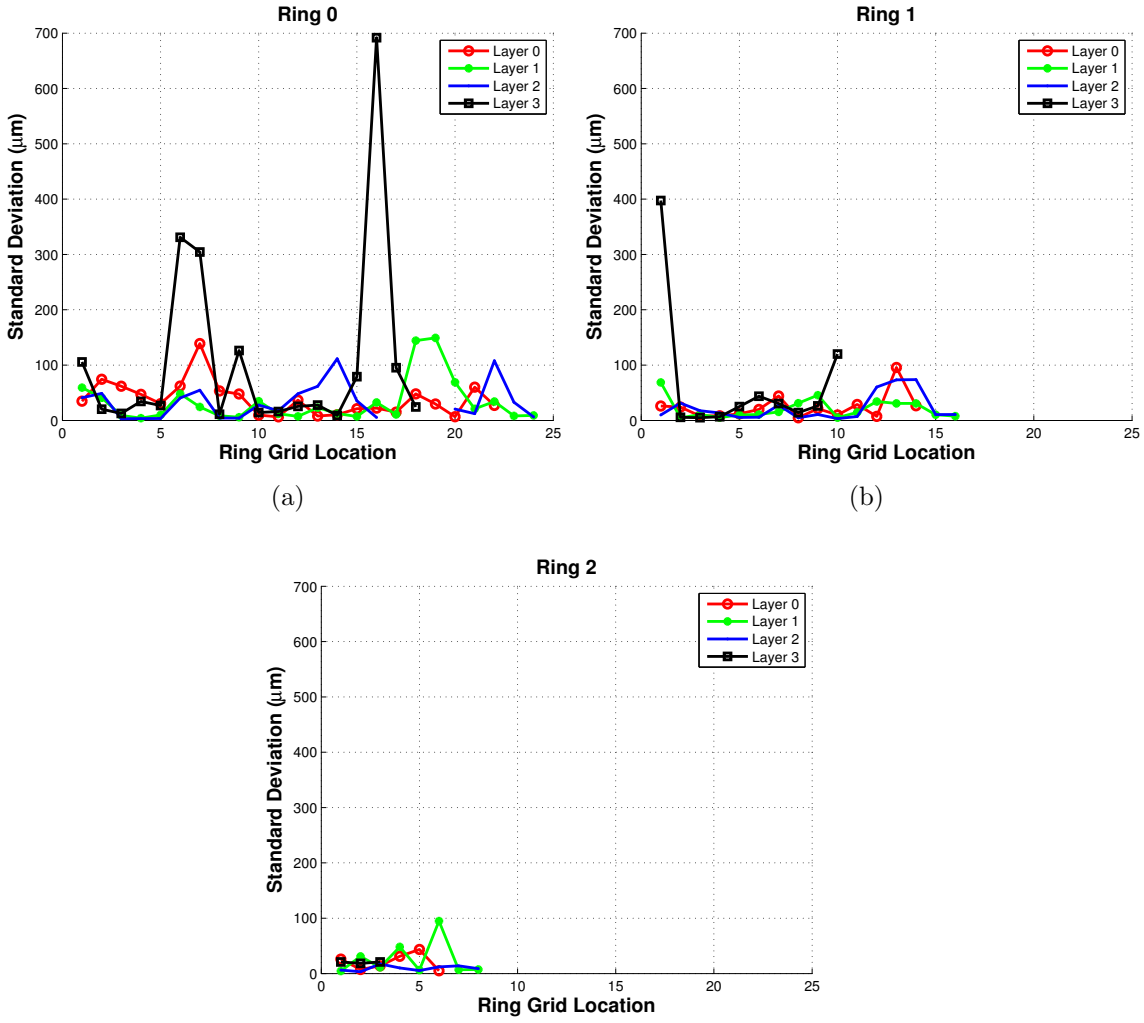


Figure 3.33: σ in the Y-axis expressed in terms of the concentric ring composing the layers (a) Ring 0 (b) Ring 1 (c) Ring 2

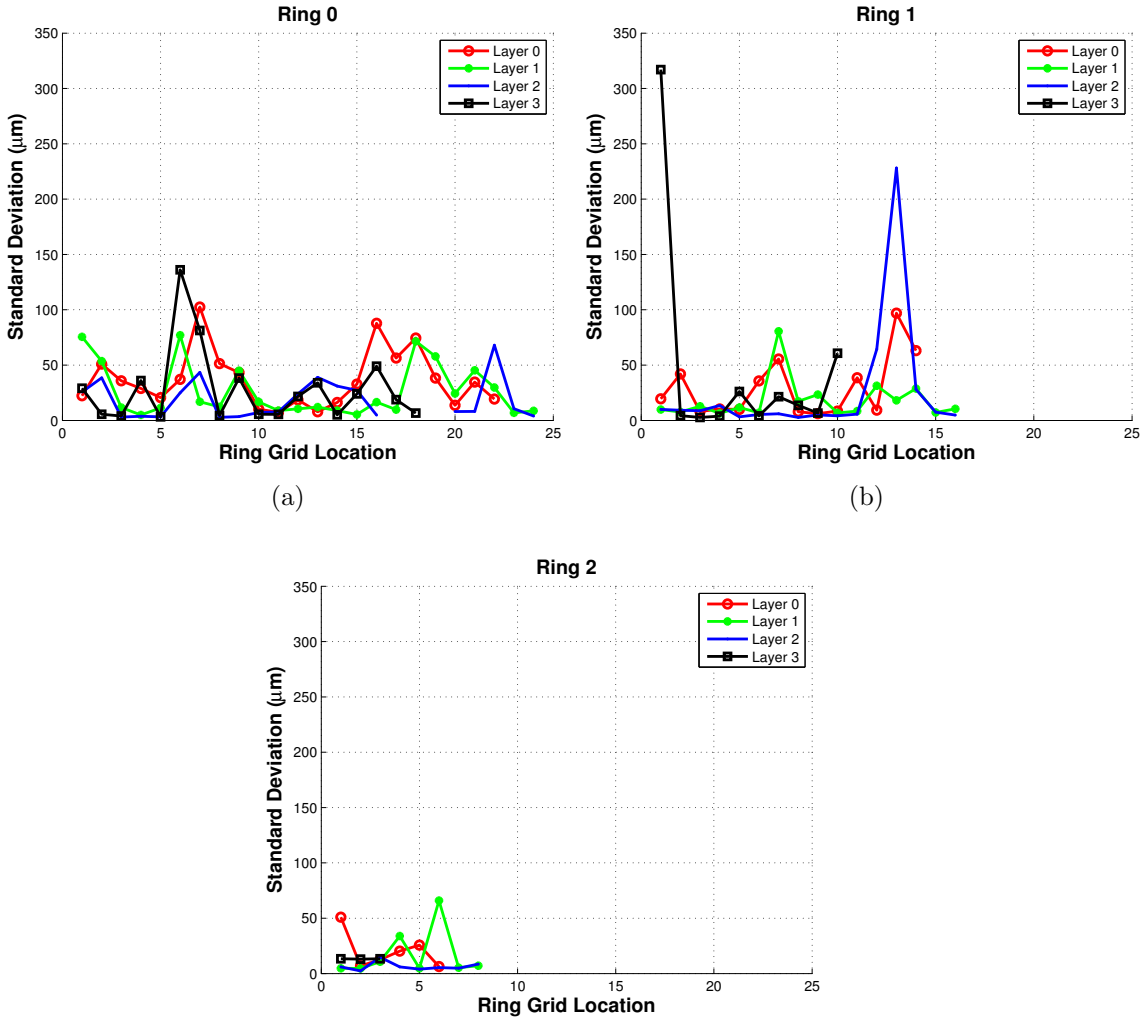
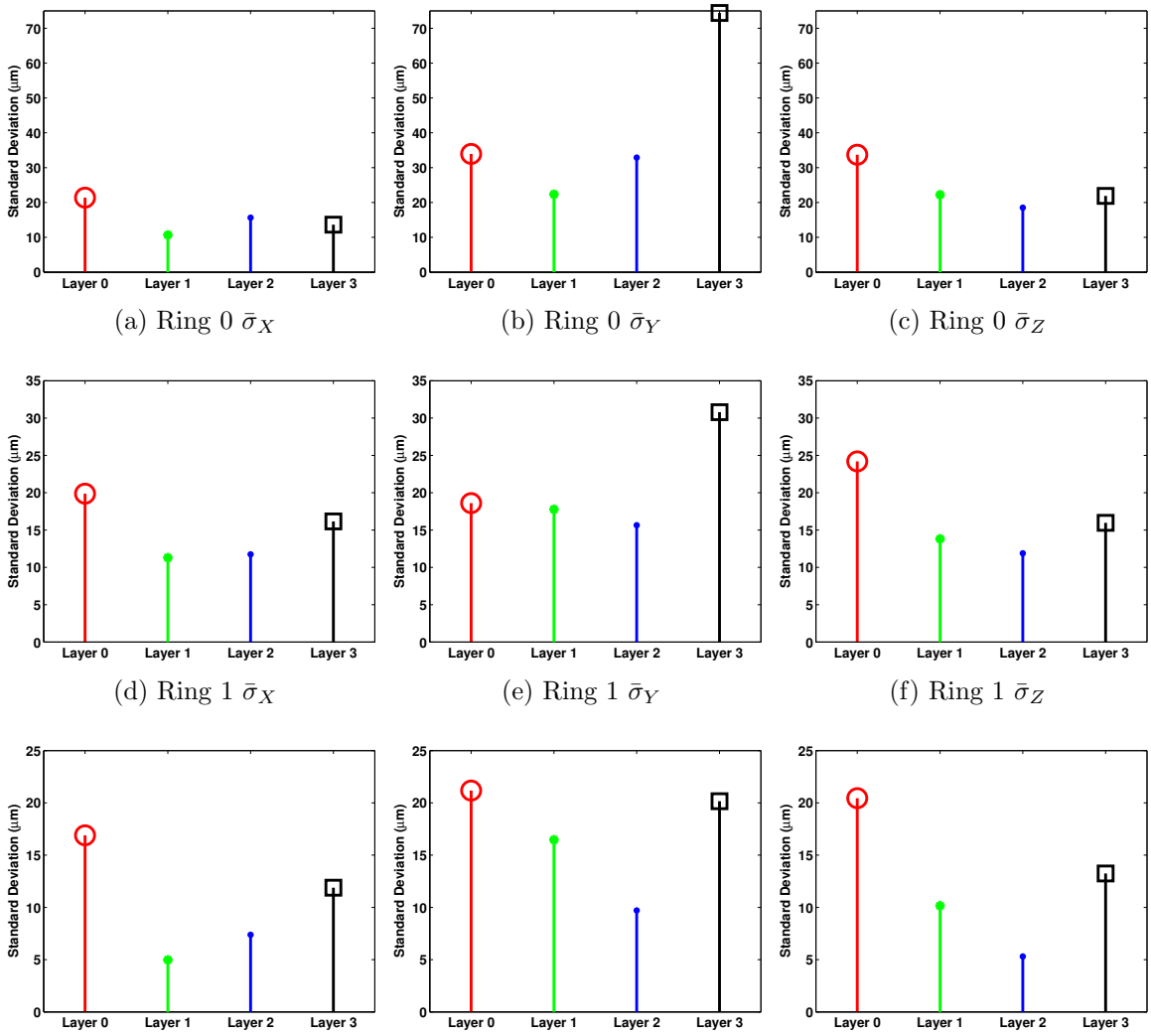


Figure 3.34: σ in the Z-axis expressed in terms of the concentric ring composing the layers (a) Ring 0 (b) Ring 1 (c) Ring 2

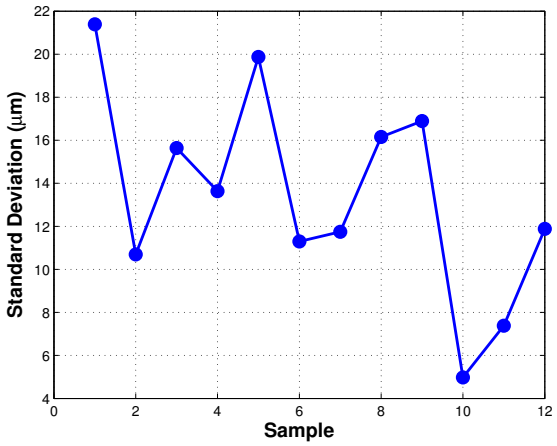


(a) Ring 0 $\bar{\sigma}_X$ (b) Ring 0 $\bar{\sigma}_Y$ (c) Ring 0 $\bar{\sigma}_Z$

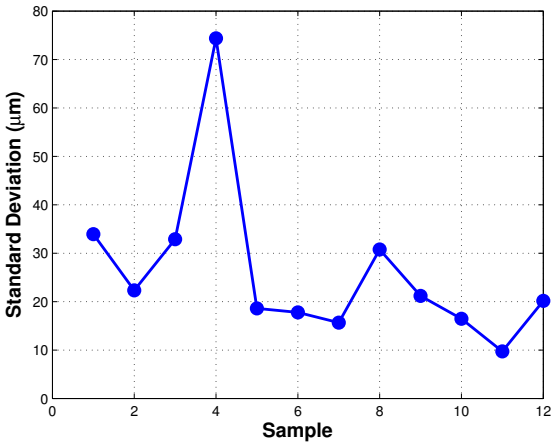
(d) Ring 1 $\bar{\sigma}_X$ (e) Ring 1 $\bar{\sigma}_Y$ (f) Ring 1 $\bar{\sigma}_Z$

(g) Ring 2 $\bar{\sigma}_X$ (h) Ring 2 $\bar{\sigma}_Y$ (i) Ring 2 $\bar{\sigma}_Z$

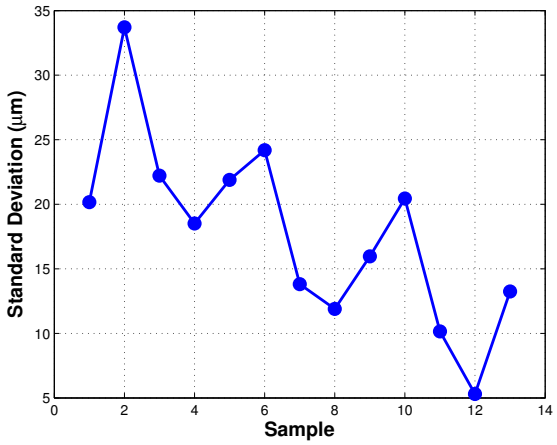
Figure 3.35: Mean Ring variance for each layer



(a) $\bar{\sigma}_X$

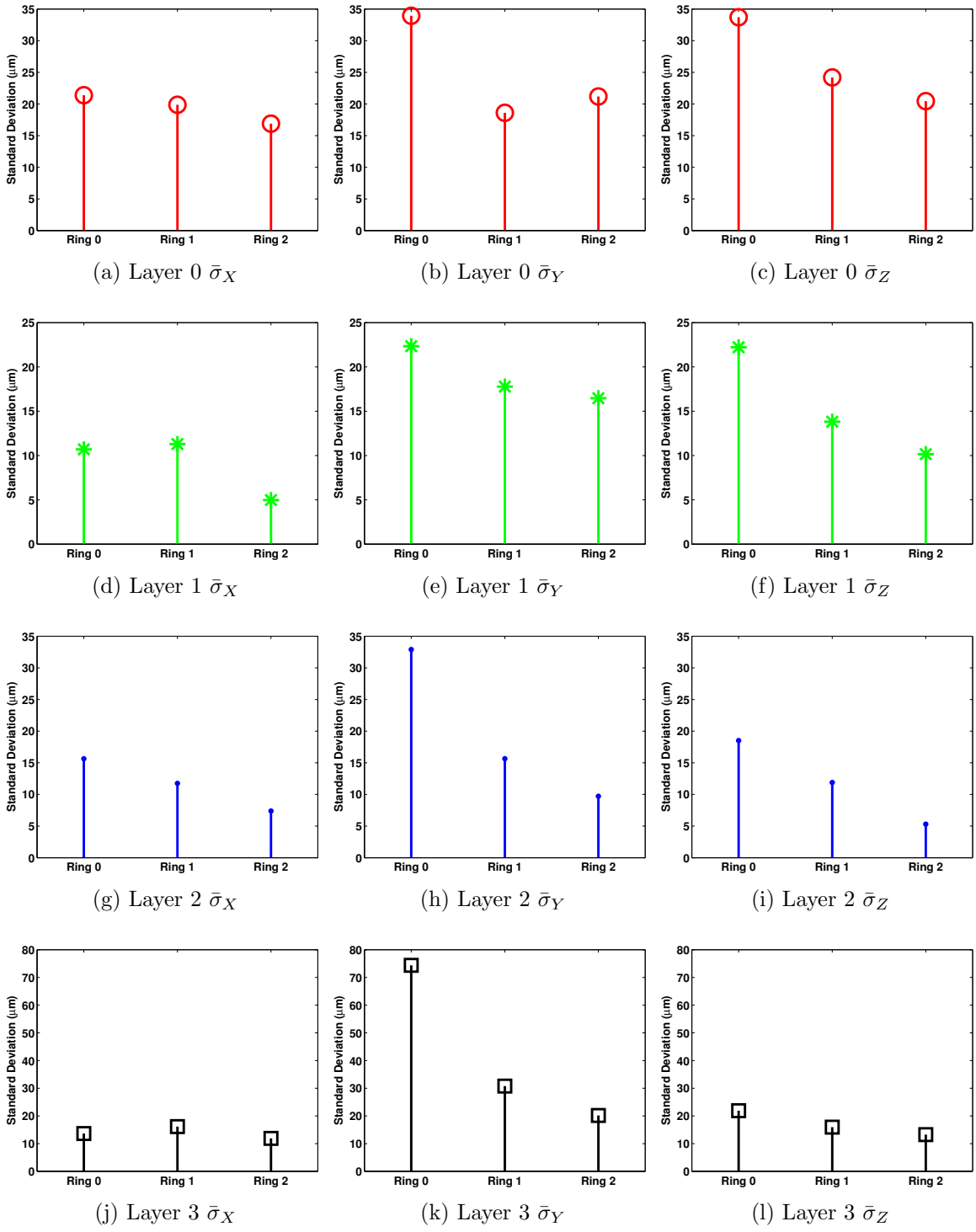


(b) $\bar{\sigma}_Y$



(c) $\bar{\sigma}_Z$

Figure 3.36: Average σ in each axes over volume



(a) Layer 0 $\bar{\sigma}_X$ (b) Layer 0 $\bar{\sigma}_Y$ (c) Layer 0 $\bar{\sigma}_Z$

(d) Layer 1 $\bar{\sigma}_X$ (e) Layer 1 $\bar{\sigma}_Y$ (f) Layer 1 $\bar{\sigma}_Z$

(g) Layer 2 $\bar{\sigma}_X$ (h) Layer 2 $\bar{\sigma}_Y$ (i) Layer 2 $\bar{\sigma}_Z$

(j) Layer 3 $\bar{\sigma}_X$ (k) Layer 3 $\bar{\sigma}_Y$ (l) Layer 3 $\bar{\sigma}_Z$

Figure 3.37: Mean Ring variance for each layer

3.4 Conclusion

An experimental characterisation of the Vicon photogrammetry system using a high accuracy laser tracker was conducted. The study was motivated by empirical observations of measurement accuracy varying as a function of location within the system measurement volume. Error quantification was implemented through simultaneously tracking a test object volumetrically scanned through a measurement space using both systems. The test object, held by a tripod, was measured at each point of a grid of dimension 8×6 and moved through four height settings. A parabolic surface was obtained when considering error as a function of position for a fixed height setting confirming the empirical observations. It was found that the maximum observed error (in terms of Euclidean distance) was approximately 6 mm while the maximum standard deviation was approximately 74 μm. The results show that the system is precise but inaccurate. Importantly this inaccuracy is a systematic error that may be mitigated through improving the calibration of the system. The underlying reason is thought to arise from the number of cameras contributing to the reconstruction of marker positions in the volume centre being higher than the edges.

This study has been carried out under the net effect of a number of error contributing variables that should be considered in isolation in future work. For example, the relationship of estimation accuracy and the number of camera rays involved in reconstruction is a important consideration. Temperature variation can cause expansion of the metal comprising the fixed frame thus changing the relative positions of the cameras and in turn reducing the accuracy of the calibration. Each camera is fundamentally focused in only one plane, the effect of depth of field on estimation accuracy is another factor that should be investigated. All measurement systems have limitations, of course, and the validity of the measurements taken by Vicon is dependent upon the nature of the experiment being conducted in the volume. The measurements are only valid when the measured quantity is not masked by the error in the measurement system itself.

The results obtained in this chapter are extremely important in the context of this thesis and a wider context for end users of Vicon systems. In terms of this thesis it may be concluded from the results that the centre of the volume yields the lowest error and variance in measurements with respect to the laser tracker. This implies that experiments utilising the Vicon system should be conducted within this central area. Vicon systems are used pervasively in the life sciences, animation and engineering applications. The procedure developed in this chapter and future work has the potential to improve the performance of the system. For example, the errors calculatd from the procedure could be added as an extra term to be minimised in the optimisation used in the calibration of the system.

Chapter 4

Visual Odometry

4.1 Introduction

This chapter describes the design, implementation and evaluation of a Visual Odometry (VO) system for use in planar operating environments. An onboard machine vision camera is mounted perpendicularly to the heading vector of the robot such that it observes the surface over which the vehicle is travelling. The resultant orthographic images are transmitted over a network link to a workstation for processing. The aim of the system is to provide the dual function of stitching successive surface images and simultaneously estimating the motion of the vehicle through consideration of the apparent motion of image features. The former is of direct benefit in the context of NDE in that the resultant image mosaic may be used for visual NDE.

The motivation for this work is based on the fact that the inspection environment determines the method of platform locomotion as result of, for example, surface terrain/material. The position of this platform may not always be amenable to estimation by wheel encoders which is the case for a four wheel drive robot (see Figure 4.1) which depends upon wheel slippage in order for the vehicle to turn: encoders attached to the wheels would drift very quickly. Although visual

odometry by the nature of the estimation process will drift, it can provide a useful low level source of positioning data to be fed as an estimate into further processing.

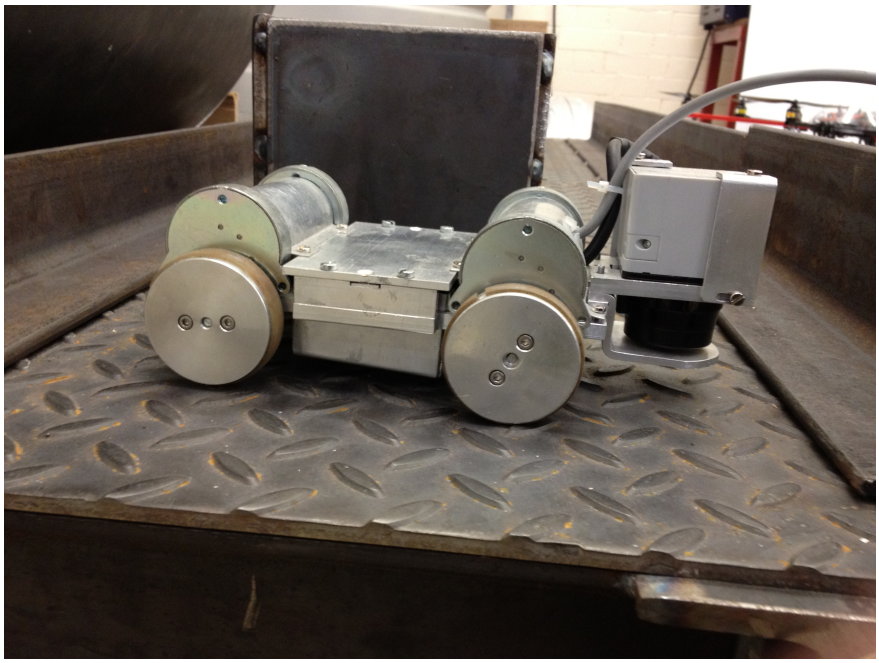


Figure 4.1: Four wheel drive NDE robot designed to operate on a standard non-slip metal chequerboard for industrial inspection application in nuclear plant.

4.1.1 Literature Review

Continual advancements in processing power particularly with Graphical Processing Unit (GPU) technology have meant that realtime image processing techniques are being used increasingly in robotics. Camera sensors can play several roles due to the information rich nature of an image including object/shape recognition, motion estimation and ranging. In the context of mobile robotics, the use of a multi-purpose sensor could be beneficial in terms of weight, power and cost. An additional advantage of a single sensor is that potentially complicated inter-sensor calibration can be avoided. In NDE inspections, visual inspection is extremely

important in its use as a first pass in identifying anomalous features on the structure under investigation, as such visual sensors form one of the major sensors carried on-board the RSAs.

4.1.1.1 Camera as a Metric Device

A single camera receives a projection of the scene onto its image sensor. This data by itself cannot provide metric information about the content of the image. If *constraints* are applied, however, such as knowing a priori the dimensions of an artefact in the scene then metric data can be obtained through a scaling from pixels to metric length. Monocular systems such as the one presented in this chapter must make use of such constraints. Single cameras augmented with sensors such as IMU's can provide metric data where for example the camera provides a refinement in the positional estimate output by the IMU. Metric data can be estimated directly if more than one camera is utilised in for example a stereo configuration as described in Section 2.3.3.2.

4.1.1.2 Image based Motion Estimation

The premise of all image based motion estimation systems is the same: identify common areas in different images and then use these areas to infer how the camera has moved between images or conversely how the object has moved between cameras. The common points can be a priori known markers or *fiducials*; this is the case for the Vicon system which uses fixed spherical markers. Because the system in effect knows what to look for in the incoming images, processing is inexpensive and the system can be made robust. There also exists the concept of markerless systems where the common areas comprise pixel arrays derived from the content of the image - this type of system is much more difficult to implement in practice. This is the type of system implemented in this chapter because usually in most cases the inspection environment cannot be prepared *a priori* and therefore virtual fiducials must be created.

Image stitching (also known as *mosaicing*) without the use of fiducials is an extensively researched area in the computer vision literature. Szeliski [67] provides a detailed overview of the techniques available for aligning images: the categories reduce to direct alignment or feature based alignment. The direct approach attempts to align images through minimisation of an error metric describing how well the pixels in both images agree - every pixel in both images are taken into consideration. For the method to work well it needs to be initialised close to the global minimum. The feature based approach seeks to align the images through consideration of a set of points extracted from both images - only a subset of pixels are used. With advancements in feature extraction algorithms, the feature based approach has gained favour in the literature [68] and as such the feature based approach was adopted for the system developed in this chapter.

Estimating vehicle motion from images was firstly considered by Moravec [69] for the case of a stereo pair of images and more recently in [70]. Nister [71] reports on a robust real time visual odometry system for both the stereo and monocular case executing on consumer grade hardware in outdoor environments. These systems employ the same fundamental steps: stable features are extracted from incoming images; an outlier rejection scheme is applied and the resultant set of points are used to estimate inter-frame vehicle motion. The problems of stitching and motion estimation are closely related. The estimate of motion in this system results as by product from determining the optimal rigid body parameters that minimise the positional error between features extracted in both images.

MacKenzie [72] demonstrated the use of a low cost imaging sensor integrated into a pipe crawling platform for automated defect recognition and classification. One of the objectives in this chapter is to stitch successive overlapping images, that the resultant image could then be supplied to the system developed in [72] for post inspection defect analysis.

A system conceptually similar to that developed here is reported in the paper by Ferrer et al [73]. The system creates an image composite from multiple images acquired from a subsea vehicle using a monocular camera observing the seafloor. The approach firstly arranges the images according to the position of the vehicle

estimated by an onboard IMU. The positions of the images are then refined through a global optimisation to minimise a cost function expressing positional error between the extracted features in the images. This system falls in the category of camera and sensor for providing metric data.

Hansen et al [74] describe a monocular visual odometry system for application in NDE that has several similarities to the system developed here. High resolution visual metric maps of internal pipe walls are generated by a wheeled robotic crawler (travelling axially in both directions) to enable the inspection of pipes carrying liquefied natural gas. Lighting is provided by Light Emitting Diode's (LED) coupled with an application specific calibration procedure to minimise the adverse effect of non-uniform illumination in the images. To extract metric data from a single camera the system relies upon two important constraints namely that the pipe is a straight section and is of constant radius. If these assumptions hold, the system is capable of estimating the 3D orientation and position of the camera through optimising the alignment of *all* (overlap from both passes) overlapping images. The authors implement both direct and feature based alignment methods and show that the positional error compared to ground truth is within 1% when traversing 4 m and 6 m long pipes.

In a follow up paper Hansen et al [75] substitute the single camera with a stereo configuration that enables direct 3D measurements and thus allows the previous pipe geometry constraints to be relaxed. A similar optimisation scheme is then used to best align all overlapping images, again the authors show an error of less than 1% compared to ground truth on the datasets used to demonstrate the system.

The feature based techniques discussed above depends upon the availability of features in the operating environment. If features are not present motion cannot be estimated from image data. This condition is usually met in natural environments such as in the outdoor and subsea environments in [71] and [73]. However, in artificial environments scenes composed of repeating structures and low texture materials could be problematic for feature extraction. As demonstrated in [74], the NDE operating environment is usually rich in texture from mechanical fix-

tures and surface defects such as rust, cracks and scratches. Given the likelihood of features being available in real world scenarios, the feature based approach was chosen.

4.2 Hardware Implementation

This section describes the embedded hardware and software used to implement the visual odometry system. The high level embedded system carried by each RSA is described followed by details of the imaging sensor that was interfaced with this embedded controller.

4.2.1 Embedded System

During the course of this research the embedded system used by the RSA was phased out of production, therefore, it was necessary to migrate to a new hardware platform. The Overo Fire embedded computer produced by Gumstix Inc. was selected as the basis of new the hardware system. The Gumstix would serve as the high level processor communicating with a server controlling one or more RSA units. The selection was based upon form factor, wireless communications and extensive processing capability provided by a 720 MHz processor coupled with graphics and digital signal processors. The system runs an embedded Linux operating system which supports a number of programming languages, the C/C++ programming language was chosen for efficiency. The Gumstix provides more computational resource than required on a robot, however, it results in great flexibility for development.

4.2.2 Image Sensor

Figure 4.2: Point Grey Research Chameleon. Sony ICX445 1/3" CCD based camera of dimension 25.5 mm x 41 mm x 44 mm

A Charge Coupled Device (CCD) based *Chameleon* camera from Point Grey Research Inc [76] was chosen for image acquisition, the sensor is shown in Figure 4.2. A Universal Serial Bus (USB) 2.0 connection allowed it to be interfaced with the embedded computer in a straightforward and robust manner. The CCD of size 1296 x 964 pixels with 8/16 bits of resolution and can produce color/grayscale images. The camera also has an external trigger to allow synchronisation with data captured from other sensors. A key selection criterion for the camera was the availability of a *global* shutter which is extremely important for imaging fast moving objects. Such a shutter exposes all pixels for the same time period ensuring that the scene being observed is captured without image smearing; this is an effect that can occur with *rolling* shutters. Use of a rolling shutter would be detrimental to the feature matching process as features could be smeared leading failed matches or false positives. A C-mount Fujinon [77] low distortion lens with

a focal length of 6 mm, variable F-number ranging from $F1.2$-$F16$ and horizontal and vertical fields of view of 43° and 33° respectively was selected to pair with the imaging sensor.

In the RSA implementation, 640 x 480 pixel images with 8 bits/pixel were acquired at a rate of approximately 30 Hz into the embedded system. The camera application programmers interface embodied the IEEE 1394 standard allowing manipulation of several important parameters controlling image acquisition including the shutter time and and gain factors. The libUSB [78] library was used to convert the IEEE 1394 packets into USB packets. Following acquisition the images were relayed across the active network connection (Wi-Fi/Ethernet) to a central controlling server for image processing. The images were transmitted without compression in order to avoid compression artefacts that could modify image features. A block diagram of the processing pipeline is shown in Figure 4.3.

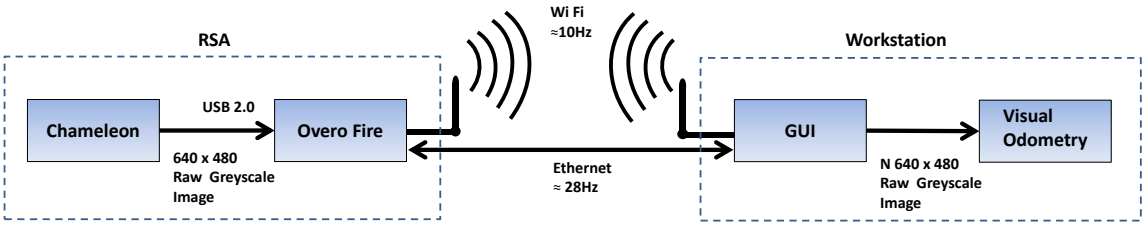


Figure 4.3: System block diagram. The acquisition rate over Wi-Fi was on average 10 Hz while 28 Hz was achievable on Ethernet.

4.2.3 Image Distortion

Prior to camera operation, the camera was calibrated in order to determine the parameters defining the mapping from world to camera frame coordinates. This mapping is a combination of a linear component resulting from the idealized pin hole model of the camera discussed in Section 2.3.3.2 and a non-linear component arising from the lens whose operation deviates from this ideal. The extent of the

non-linearity is a function of the focal length of the lens: the smaller this value then the greater the ability of the lens to bend incoming light rays and as result the greater is the non-linearity [28]. The distortion introduced by the lens may be modelled by *radial* and *tangential* distortion of which the former typically dominates in contribution. The image plane point to which a world coordinate maps is denoted, (u, v), and given by the following [79]:

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} c_x \\ c_y \end{pmatrix} + \frac{f}{z_c} \begin{pmatrix} x_c \\ y_c \end{pmatrix}, \tag{4.1}$$

where (x_c, y_c, z_c) is the 3D world coordinate in the frame of reference of the camera and f (assuming square pixels) is the camera constant denoting the distance in pixels between the image plane and the projection centre. If the distorted version of this coordinate is denoted (u_d, v_d) then the problem becomes one of finding the functional mapping between the two coordinates. In the case of radial distortion, the distortion can be modelled as follows:

$$\mathbf{x}_d = \mathbf{x}_n(1 + k_1 r^2 + k_2 r^4), \tag{4.2}$$

where $r = |\mathbf{x}_n|$ i.e. the radius upon which the undistorted point lies, k_1 and k_2 are distortion parameters and the points $\mathbf{x}_n$ and $\mathbf{x}_d$ result from rearranging Equation 4.1 and evaluating for $z_c = 1$. The tangential distortion can be modelled as follows:

$$\mathbf{x}_t = \mathbf{x}_n + \begin{pmatrix} 2k_3 x_n y_n + k_4(r^2 + 2x_n^2) \\ k_3(r^2 + 2y_n^2) + 2k_4 x_n y_n \end{pmatrix} \tag{4.3}$$

which is a function of distortion parameters k_3 and k_4. The full lens distortion model is a linear combination of the radial and tangential contributions as follows:

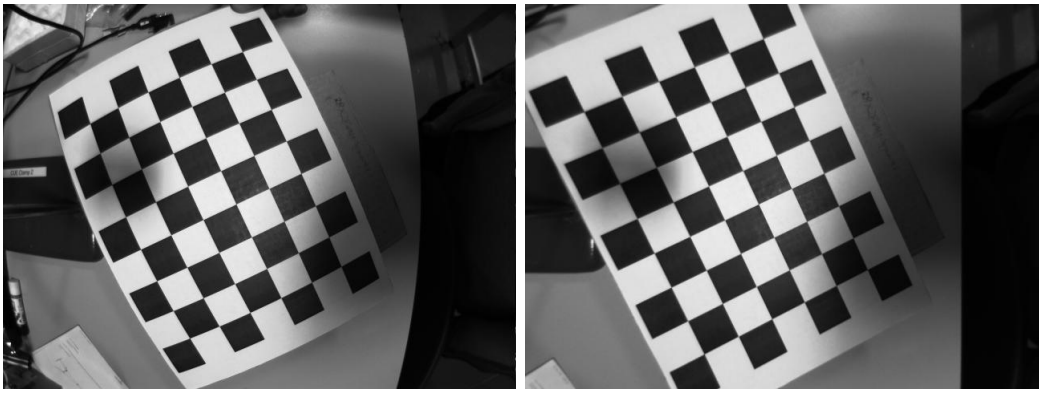
$$\mathbf{x}_d = \mathbf{x}_r + \mathbf{x}_t \tag{4.4}$$

Equation 4.4 is a forward model. Given the distorted coordinates, the objective is to recover the original undistorted coordinates, i.e. the inverse model is required. The inversion of the Equation 4.4 cannot be carried out analytically due to the undistorted point $\mathbf{x}_n$ being a independent of (u_d, v_d). An iterative procedure can be used to minimise an approximate inverse model of the form:

$$\mathbf{x}_n \approx \mathbf{x}'_n = \frac{1}{(1 + k_1 r^2 + k_2 r^4)} \left[\mathbf{x}_d - \begin{pmatrix} 2k_3 x_n y_n + k_4(r^2 + 2x_n^2) \\ k_3(r^2 + 2y_n^2) + 2k_4 x_n y_n \end{pmatrix} \right], \tag{4.5}$$

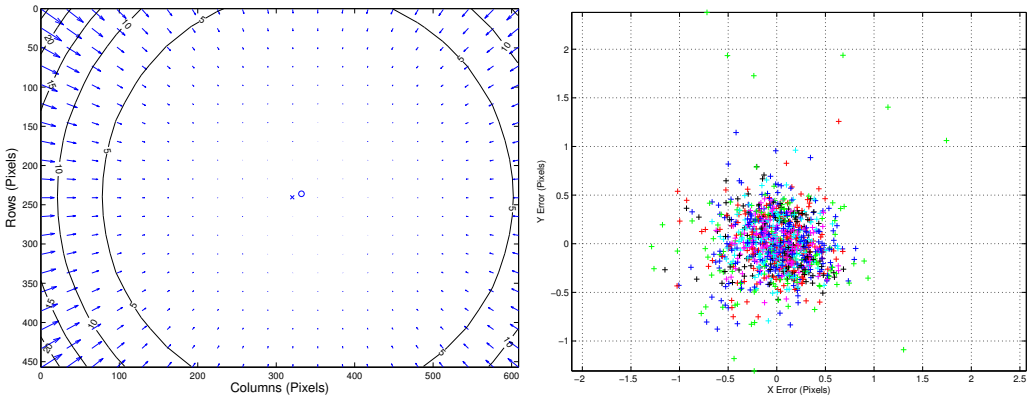
where $\mathbf{x}'_n$ is the approximation of the true undistorted coordinate $\mathbf{x}_n$. Distortion can be handled in two ways: either every pixel of the incoming image is remapped to undo the lens distortion or alternatively if a feature based method is being used then features can be passed through the forward (Equation 4.4) and inverse mappings (Equation 4.5) [80] with the every pixel method being more computationally expensive than the feature based method. In the present implementation the "every pixel" method was used for simplicity.

The Matlab Camera Calibration Toolbox [81] was used to determine the camera intrinsics matrix of Equation 2.7 and the distortion parameters k_1 through k_4. The procedure consisted of capturing 25 images of a calibration artefact with known geometry from varying viewpoints. The grid intersection points from the resultant set of images were manually extracted to subpixel resolution and used by the toolbox to find the mapping from the distorted version of the grid to the ideal rectilinear form. A distorted image of the artefact which consisted of a black and white grid with known cell size is shown in Figure 4.4(a) while the corresponding undistorted version is shown in Figure 4.4(b). As can been seen in Figure 4.4(a), the true straight lines of the grid appear curved with the severity of the distortion being greater at the edges of the image compared to its centre. The undistorted grid corrects for the distortion and in doing so some of the original image data is lost. Figure 4.5(a) visualises the effect of the lens distortion wherein the length and direction of each arrow represents the displacement of the corresponding pixel introduced by the lens.



(a) (b)

Figure 4.4: (a) Distorted image of calibration grid. The straight lines of the grid appear curved. (b) Undistorted image of calibration grid. The straight lines of the grid appear straight following calibration



(a) (b)

Figure 4.5: (a) Visualisation of the tangential and radial lens distortion of pixels in a VGA resolution image (b) XY residuals between true corner points and model predicated points using parameters from Tables 4.1 and 4.2

The estimated parameters of the intrinsics matrix and distortion coefficients are shown in Tables 4.1 and 4.2 respectively with their associated uncertainties. The tangential parameters are an order of magnitude smaller than those for the radial distortion as suggested by the image of the artefact. The error resulting from the

difference between the manually extracted corner points for an image and those predicted through application of the distortion model is plotted in terms of XY error for each image in Figure 4.5(b). The graph shows that the mean error in each of the components is around 0.5 pixels which was considered sufficiently low to proceed with the values calculated from the procedure. Given these parameters, the distortion correction was implemented in C++ using the associated functions in OpenCV [82].

Intrinsic Parameter (Pixels)	**Optimised Value** $\pm 3\sigma$
Focal Length (f_c)	$\begin{pmatrix} 840.45 \\ 838.91 \end{pmatrix} \pm \begin{pmatrix} 9.30 \\ 9.39 \end{pmatrix}$
Principal Point (cc)	$\begin{pmatrix} 331.13 \\ 235.01 \end{pmatrix} \pm \begin{pmatrix} 7.41 \\ 7.44 \end{pmatrix}$
Skew (α_c)	0

Table 4.1: Entries of the intrinsics matrix resulting from the calibration procedure

Distortion Parameter	**Optimised Value** $\pm 3\sigma$
k_1	-0.14 ± 0.044
k_2	-0.52 ± 0.48
k_3	0.0012 ± 0.0015
k_4	0.0024 ± 0.0016

Table 4.2: Distortion Parameters calculated from calibration

The extent to which successive images overlap is a function of vehicle speed and the frame rate of the camera of which the latter is less flexible. Vehicle speed should therefore be chosen to maximise overlap within the time constraints of the inspection. The developed system operates offline therefore for ease of development the entire image was undistorted before proceeding with the feature extraction stage.

4.3 Feature Extraction

An image is a information-rich signal which can enable several functions to be carried out within a robotic system. For visual odometry, only a subset of this data is required. Feature extraction is a commonly used dimensionality reduction technique which constitutes a fundamental task in computer vision. It comprises two sub-problems namely feature detection and description.

A feature may be defined as a region of the image which is dissimilar to its immediate neighbourhood in terms of properties such as intensity, colour and texture [83]. Each feature has an associated *descriptor* consisting of a set of measurements representing the image structure in its neighbourhood. Once defined, the descriptor can be used to identify the same feature in another image which potentially has a different viewpoint of the scene. The extracted feature may have relevance in the context of the image for example if the detector extracts edges these could correspond to lines on the object in the image. Features can also be regions of the image which may not have particular meaning with respect to the image but the composition of the pixels is such that in tracking applications it can serve as an area that can be tracked in an accurate and stable manner. Within this area of research there exists the concept of local and global features. Global features reduce an image to a single set of parameters and are often used in object recognition tasks [84]. This type of feature tends to be sensitive to clutter in the scene and occlusions of the object in question. Local features first investigated by Schmid and Mohr [85] on the other hand produce many features and associated descriptors from a single image. There exist a large range of feature extraction algorithms in the literature that focus on extracting different types of feature including, corner detectors, blob detectors and region detectors.

Local features have gained widespread use in computer vision tasks including texture recognition, robot localisation, building panoramas primarily for their invariance to operations such as scaling, rotation and illumination changes. The paper by Tuytelaars and Mikolajczyk [83] lists several desirable properties that extracted local features should exhibit. The most important is that of *repeatabil-*

ity meaning the ability to re-detect the same point across images with differing viewing conditions. The property of repeatability is a function of the property of *distinctiveness* which refers to the uniqueness of the vector of parameters used as a fingerprint for the feature. In addition to the preceding a good feature should be *local* in order to minimise occlusion effects; there should exist a reasonable *quantity* of features in the image; the features should be *accurately* localised in the image and finally it should be *efficient* to calculate the descriptor.

The descriptor can take the form of an image patch which is effectively an area cut out of the image whose centre lies on the feature. To determine if a particular feature resides in a different image, the correlation of the two patches from each image can be used to determine patch similarity. This method can be made to be illumination and rotation invariant, however, matching failure can occur due to the relative warping of the patches resulting from camera viewpoint changes. Another technique often used is based upon histograms which capture colour information in the image. Evaluating the histogram of patches in the neighbourhood of the feature can been used in combination with shape information to identify a point.

The Scale Invariant Feature Transform (SIFT) developed by Lowe [86] proposes both a detector and invariant descriptor. SIFT features are robust to changes in image scale, rotation and partially invariant to illumination changes. The algorithm comprises of several involved steps, the reader is referred to [86] for a comprehensive description; the following provides a brief overview of its operation. A feature is obtained from scale-space extrema of Difference of Gaussians (DOG) within a DOG pyramid. A Gaussian pyramid is created from the input image through repeated smoothing and subsampling. The DOG pyramid is then computed from the differences between adjacent levels in the pyramid. The coordinates of features are points in the DOG that are extrema. A descriptor comprising of a 128 element vector is created for each SIFT feature calculated from a 16x16 grid of pixels centred on the feature. This window is then divided into a 4x4 pixels block. In each block the gradient and orientation for each pixel is calculated and discretized into eight orientation bins. The SIFT descriptor is composed of the gradient and orientation values from each block.

SIFT is used pervasively in the computer vision based robotics literature. Mikolajczyk and Schmid [87] carry out an exhaustive evaluation of ten local descriptors comprising of histogram, spatial frequency and differential based descriptors through consideration of matching and recognition of the same scene/object under different viewing conditions. In particular, the authors consider geometric and photometric image transformations in scale, rotation, blurring, viewpoint and illumination, the effects of JPEG compression are also considered. It is found that the Scale Invariant Feature Transform (SIFT) performs very well in comparison to the other types of descriptors for most of the tests only being out performed by the gradient location and orientation histogram descriptor proposed by the authors.

Bauer et al [88] carry out a comparison of SIFT and the Speeded Up Robust Features (SURF) [89] algorithm on a set of natural scene outdoor images. Similarly to SIFT, SURF provides both a detector and invariant descriptor. Instead of calculating the DOG, SURF computes the Hessian matrix for identifying feature points within the image. A 64 element descriptor describes the distribution of Haar-wavelet responses in a window surrounding the feature coordinate. The smaller feature length reduces the time required for feature computation and matching compared to SIFT. It was found that the SIFT produced more features per image than SURF and that the associated quality of matching was slightly greater for SIFT according to the metric defined as the ratio of correct to incorrect matches. SURF, however, was found to produce more interest points per unit time in comparison to SIFT which may be of importance for real time systems. The authors conclude that for real time systems the better run-time performance of SURF outweighs the slightly better feature quality of SIFT. The greater accuracy of SIFT was favoured over SURF in the current implementation. With the rise of GPU implementations [90], the reduction in realtime performance was not considered to be an issue. The widely used MATLAB/C++ based SIFT implementation due to Vedaldi [91] was used.

The power of the feature extraction concept lies in *matching* features between images using the descriptors. This is implemented in terms of the nearest neighbour (subject to a threshold constraint) in terms of Euclidean distance in the descriptor

space. Carrying out the matching procedure yields a set of *point correspondences* between the images which can be used to infer motion of the camera from changes in feature image locations. Invalid matches can occur when the descriptor vectors assume similar values as a results of, for example, the image containing self similar patches or sensor noise. Failure to remove such false positives or *outliers* will produce incorrect estimates of motion and thus have a significant impact on accuracy of the motion estimates. The following section presents a technique for performing outlier rejection.

4.4 Random Sample Consensus

In the set of correspondences there generally exist erroneous matches which can introduce error into the calculation of the image to image transform. The Random Sample Consensus (RANSAC) algorithm introduced by Fischler and Bolles [92] was used to perform outlier rejection. RANSAC is an iterative algorithm which uses a hypothesise-verify framework to estimate model parameters. This is achieved through generating a model from putative inliers and subsequently checking the deviations of the model with respect to the remaining data. Specifically a random subset of the data is selected where the cardinality of this subset is the minimum necessary to compute the free parameters of the model. This fundamentally differs from least squares techniques that make use of all available data to estimate the model parameters. The point wise errors of the model with respect to the data is then evaluated. If the sum of such errors is less than a user defined threshold, ϵ, which is a function of the noise contaminating the points then the current parameters are stored in a *consensus* set. If this error is greater than ϵ, the model is discarded and the preceding steps are iterated until the threshold constraint is satisfied. As a result of multiple iterations, several consensus sets will be generated and can be ranked according to their cardinality. When the total number of iterations has been completed the model parameters associated with the largest consensus set is selected as the output of the algorithm.

The number of iterations, N, of the algorithm may be set through the following analysis [93]. Let p be the probability that at least one of the random sets does not contain an outlier. If the probability of selecting an inlier is assigned, u, then the probability of a random set of size, m, elements containing inlier's exclusively is u^m assuming the points are independent. The probability that a least one point in the random set is an outlier is, $1 - u^m$, which is greater than the probability that they are all outliers $(1 - u)^m$. Given the preceding the following equation may be established assuming the iterations are independent:

$$1 - p = (1 - u^m)^N \tag{4.6}$$

Through rearrangement of Equation 4.6 a lower bound for N can be derived as follows:

$$N = \frac{\log(1 - p)}{\log(1 - u^m)} \tag{4.7}$$

In the current implementation N was set to 1000. Given the set that has produced the best fitting model parameters, the inlier's making up this set can be used to recompute the model parameters using a least squares technique. In this system, RANSAC was used to estimate the homography, H, relating two images. A homography is an invertible image transform that can describe general perspective changes between two images. H was estimated using the Direct Linear Transform (DLT) [28] with a random subset size of four. The inliers corresponding to the best fitting homography were used to compute the rigid body transform relating two images described in Section 4.7.

4.4.1 Outlier Rejection Example

The image processing pipeline in this system is demonstrated by a simple example. Figure 4.6 displays consecutive images of a grid acquired by the robot during a traversal of approximately 20 mm, this is illustrated by the downward

translation of the handwritten "25" in going from one image to the next. A subset of the extracted SIFT features are shown in the left of Figure 4.7. The corresponding features matched to this set through computing the nearest neighbour in descriptor space are plotted in the right and connected by a line. Notice that the majority of lines look to have a gradients of similar value across the two images while the two outliers are clearly visible with a markedly different gradient: the purpose of RANSAC is eliminate such outliers from the set of point correspondences.

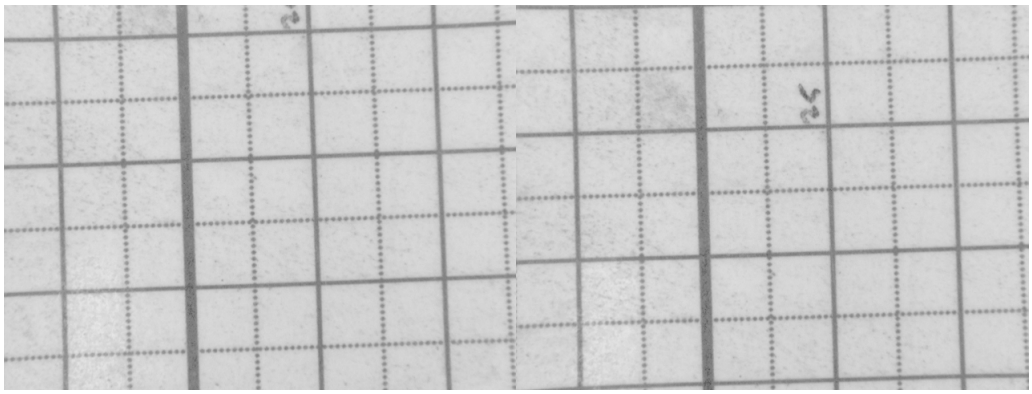


Figure 4.6: Left: image_{t-1}, Right: image_t. During interval Δt the vehicle has translated approximately 20 mm

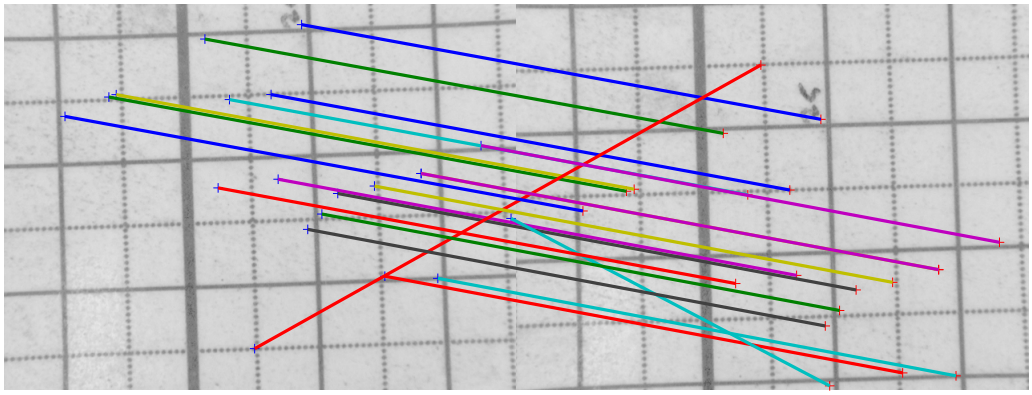
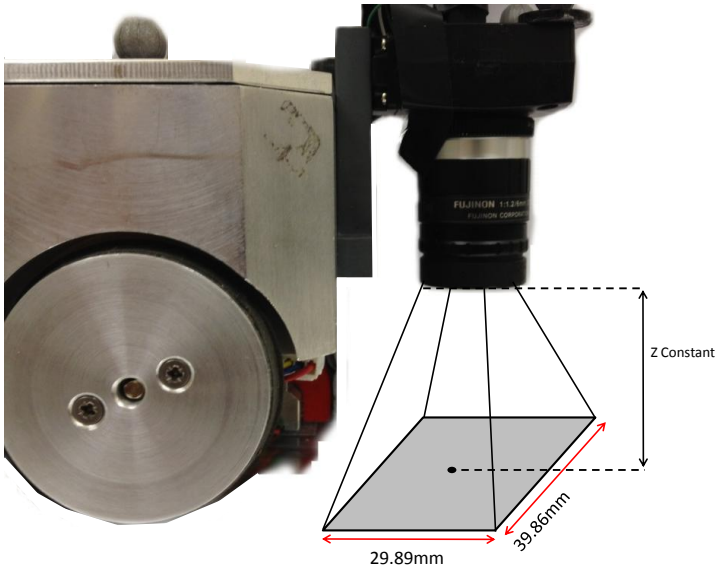


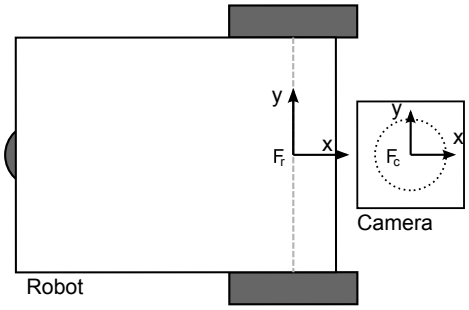
Figure 4.7: Subset of SIFT matches prior to outlier removal. The lines connect matched features between the two images.

4.5 Conversion to Metric

The use of a monocular camera as metric device in this system depends upon the assumption that the robot is travelling on a flat surface which gives rise to the constraint of constant depth, z. The configuration is shown in Figure 4.8(a) where the camera captures a fixed area, in this case, 29.89 mm x 39.86 mm, of the surface over which it is moving. The focusing distance and field of view of the lens determine the area that can be imaged. This constraint means that it is possible to determine a conversion factor to transform from pixels to mm. This is similar to the constant radius assumption used in the pipe crawler in [74]. The planarity assumption is not very restrictive in the context of NDE as many scenarios consist of inspecting large panel type structures.



(a)



(b)

Figure 4.8: (a) Point Grey Research Chameleon camera mounted perpendicularly to the direction of travel. The Z offset was $\approx$ 40 mm. (b) Robot and camera coordinate frames

4.6 Feature Extraction on Industrial Surfaces

The applicability of the technique relies critically on the texture of the surface over which the vehicle is travelling. Highly textured surfaces meaning those that exhibit large variations in pixel intensity generally yield a high count of distinctive features and thus provide many points from which to compute the inter-image

transform. Low texture surfaces conversely provide a low feature count and also a higher probability of the extracted descriptors being similar due to the underlying homogeneity of the surface. Such surfaces can therefore prove to be difficult both in terms of an insufficient number of features to compute the transform (minimum of three points are required) coupled with erroneous matching amongst these features due to the similarity of the descriptors: these factors can cause failure of the method.

This section assesses the effectiveness of feature extraction on a range of NDE materials through the computation of mean feature densities. SIFT features and descriptors are extracted on samples of aluminium, steel, bricks and substantially rusted steel. Images of the laboratory floor are also considered - such images are used in the results section to demonstrate application of the technique on a large dataset. The different materials are shown in Figure 4.9 where it is apparent that the surfaces display varying levels of reflectivity and texture under the ambient lighting conditions.

4.6.1 Illumination

Illumination of the scene plays an important role in any image processing application: in this study experiments were conducted under controlled laboratory lighting. In practice the available lighting is a function of the inspection scenario. For example, limited access could mean that external lighting needs to be deployed into the area before the inspection can take place. Independence of the ambient conditions in-situ could be gained through mounting the camera in an enclosure that internally projects light uniformly onto the surface. The experiments approximated the use of uniform lighting via external flood lights arranged so as to minimise shadows.

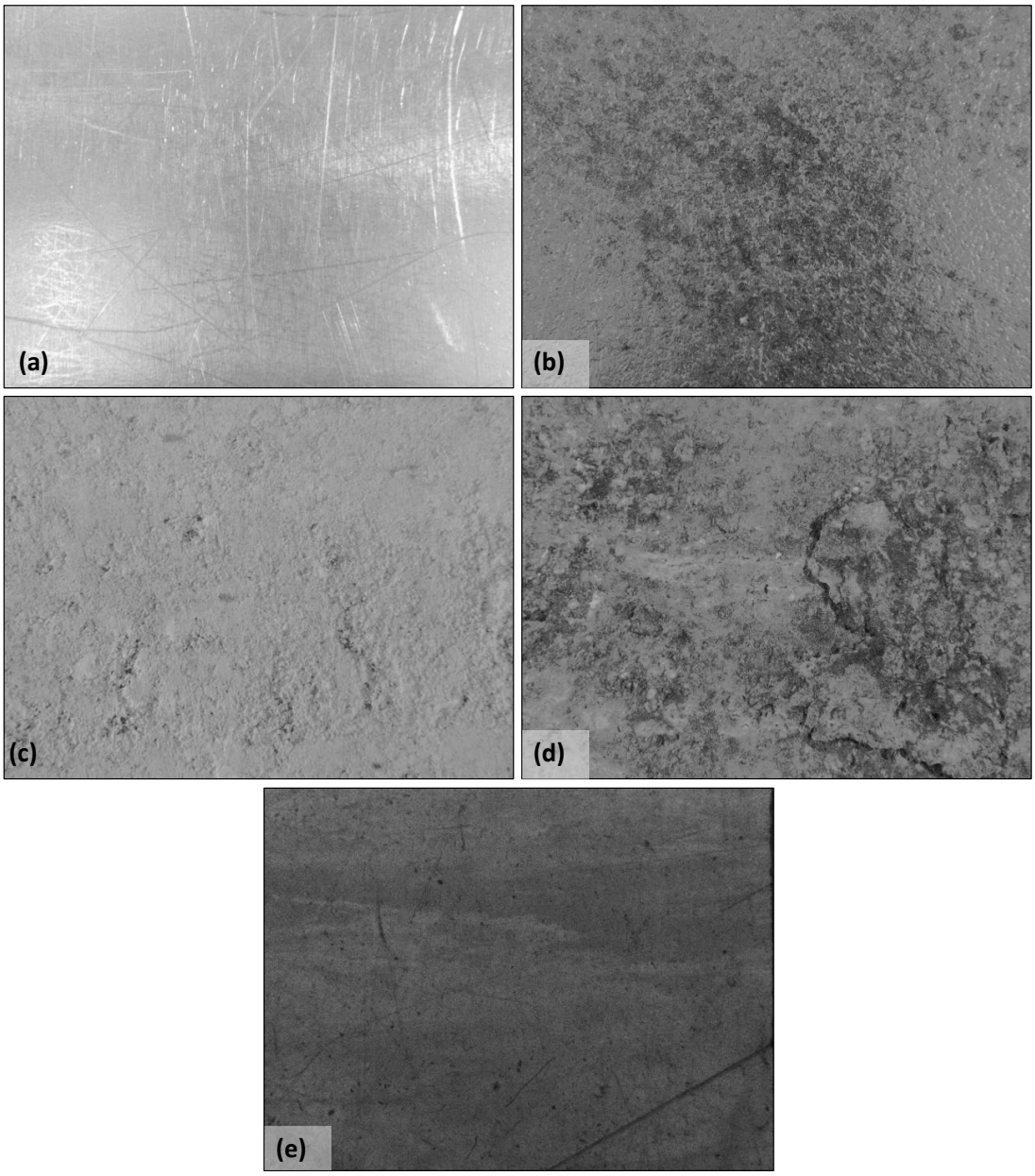


Figure 4.9: Typical image of the samples used for evaluation (a) Reflective aluminium plate section (b) Smooth steel plate section (c) Construction bricks (d) Weathered section of steel containing flaking patches of rust and areas pitting (e) Matte laboratory floor

4.6.2 Methodology

A test sequence of 60 images was recorded for each type of material. The metric of feature density measured at sequential stages of the feature extraction pipeline was used as the evaluation criteria for the materials. Specifically the densities associated with the following were considered:

1. Extracted SIFT features
2. Successfully matched features detected in 1.
3. Inliers found by RANSAC resulting from the set of matched points in 2.

The vehicle was instructed to drive a linear path at an average speed of 20 mm/s resulting in a mean overlap of 75% between consecutive images. The SIFT matching and RANSAC thresholds were set to values found to be empirically optimal from observation of path errors evaluated from many runs of the experiment. Feature densities were measured simply by dividing the number of features in steps 2. and 3. by the area of image intersection calculated from VCMS ground truth. The total feature density in 1. was calculated with respect to the total fixed area captured by each image as shown in Figure 4.8(a).

It was considered from visual inspection of the samples in Figure 4.9 that the aluminium sample would present a particular challenge due its reflectivity and self similar appearance. The remaining surfaces conversely displayed very little or no reflections and had clear textural patches suitable for satisfying the constraints specified by the SIFT detector. These observations are borne out in Table 4.3 which summarises the feature densities associated with the total number of extracted features, the number of successfully matched features and the number of inliers determined by RANSAC.

Material	Mean Feature Density(pts/mm^2)		
	Total	Matched	RANSAC
Aluminium	0.92	0.0041	0.0026
Steel Plate	1.06	0.19	0.18
Bricks	1.07	0.58	0.58
Rusted Steel	1.65	0.85	0.84
Lab Floor	0.96	0.16	0.13

Table 4.3: Feature densities measured for a set of NDE materials. SIFT match threshold was set to 3 and the RANSAC threshold was set to 0.0055.

4.6.3 Results

The aluminium sample had a total extracted feature density of 0.92 pts/mm^2 which is of similar magnitude to the density for the bricks and steel samples. There is, however, a dramatic reduction in the matched and RANSAC densities for the aluminium sample in comparison to the steel and bricks, the implication being that although there are comparable feature densities the descriptors are indistinct. This is further evidenced by the relatively large reduction in going from the matched to RANSAC densities - this reduction indicates that there is a high level of incorrect matches. In addition to this, in applying SIFT to this test sequence there were several occurrences of no detected inliers. It is considered that the lighting conditions and reflectivity of the material may have been too testing for the partial illumination invariance property of the SIFT descriptors. This test demonstrated that aluminium was not suitable for visual odometry due to high levels of incorrect matches.

The rusted steel sample produced the highest total feature density, a result suggested by observation of Figure 4.9(d) which shows a high degree of surface texture. The quality of the SIFT matches is indicated by the small difference between the matched and RANSAC densities; this shows that the majority of SIFT matches were valid. The steel plate, bricks and lab floor image sets yield

similar levels of total feature densities with the bricks and lab floor giving rise to less matched points. There exists small difference between the matched and RANSAC densities again indicating a high quality of SIFT matching. These measurements correlate with the visible texture shown by the surfaces in Figure 4.9.

Real NDE inspection scenarios generally contain texture rich surfaces as a result of the environmental conditions. The rusted steel sample is particularly representative of the condition of materials likely to be encountered in the field. The high quality of SIFT matching on this material demonstrates the suitability of the feature based technique for real inspections.

4.7 Image Based Odometry

The problem of estimating the transform between images is illustrated in the schematic of Figure 4.10. In the interval between time $t-1$ and t the robot has both translated and rotated giving rise to the arrangement shown. Given common features (circles in this case) identified in both images, the objective is to recover the transform, T, relating the poses of the robot when the images were captured. Note that the images are captured at an offset $\mathbf{t}_{offset} = (\Delta x_{CR}^{CAM}, \Delta y_{CR}^{CAM})$ with respect to the centre of rotation CR. The offsets must be estimated in order to transform from the camera frame to the robot's frame of reference.

The pipeline for the image processing used to determine T is shown in Figure 4.11. It operates in an incremental fashion whereby incoming images are immediately processed and integrated into the positional estimate. The image arriving at time, $t-1$, is firstly undistorted followed by the extraction of a large number of SIFT features, $\mathbf{f}_{t-1}$, whose associated descriptors, $\mathbf{d}_{t-1}$, are then stored. The same process is carried out for the image arriving at time t yielding $\mathbf{f}_t$ and $\mathbf{d}_t$. A search for the descriptors obtained for image_{t-1} is carried out in image_t. The

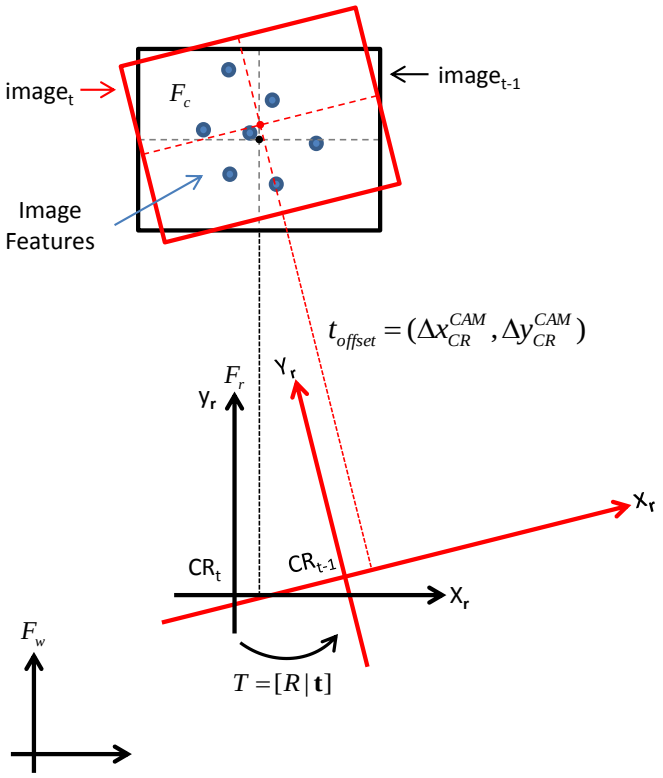


Figure 4.10: Schematic showing the geometry of the problem through overlapping image$_t$ (red) and image$_{t-1}$ (black).

result is a set of correspondences, $\mathbf{f}_{matched}$. Incorrect matches within this set are removed using RANSAC to yield the set of inliers $\mathbf{f}_{inliers}$ which are subsequently used to compute the rigid body transform relating the two point sets.

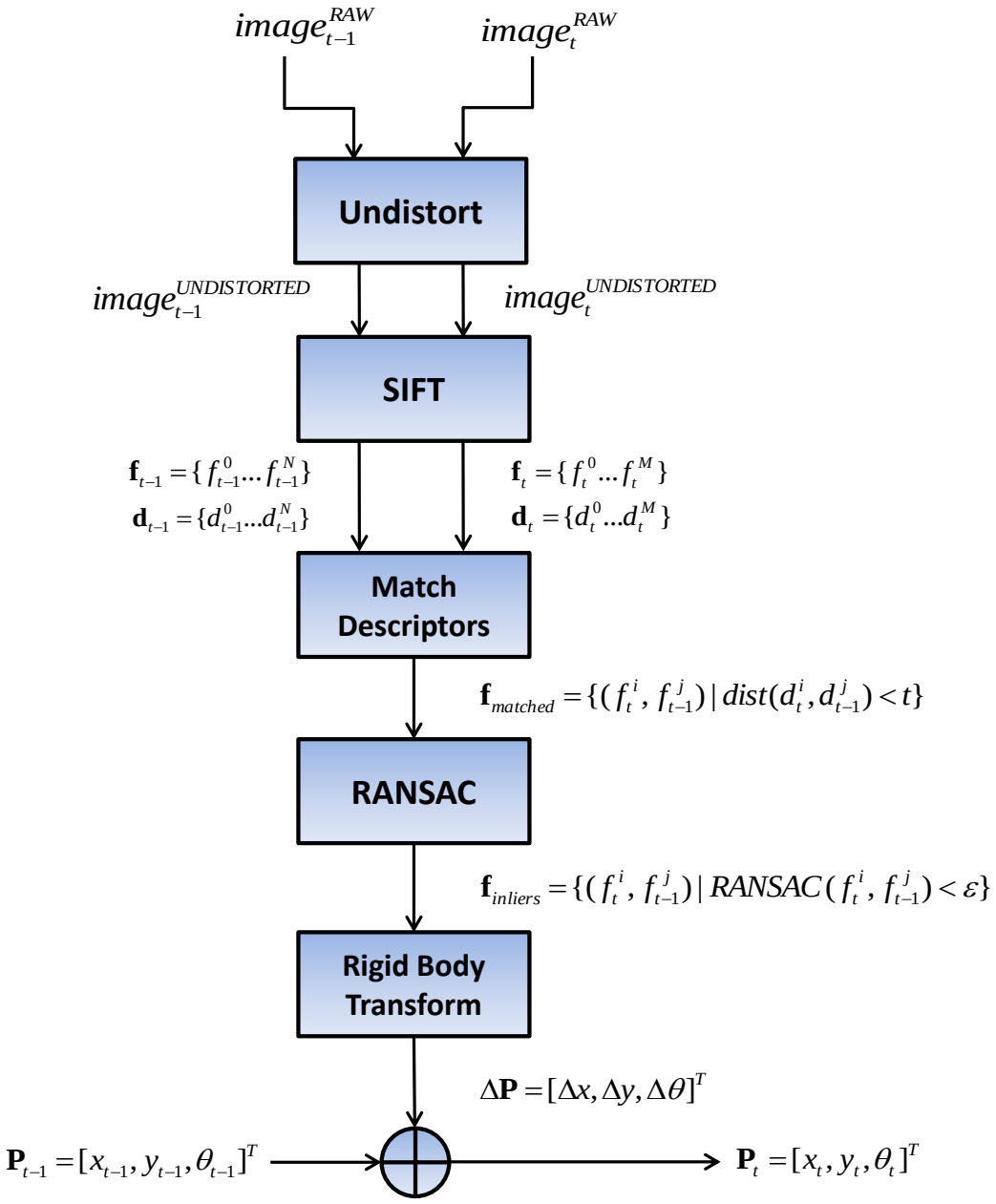


Figure 4.11: Diagram of image processing pipeline

4.7.1 Estimating the rigid body transform

Given two N-point data sets, $M = \{\mathbf{m}_i | i = 1 \dots N\}$ and $D = \{\mathbf{d}_i | i = 1 \dots N\}$, containing point correspondences from image $t-1$ and image t respectively the objective is to determine the parameters of the rigid body transform, R and $\mathbf{t}$, that minimises the cost function:

$$E(R, \mathbf{t}) = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{m}_i - (R\mathbf{d}_i + \mathbf{t})\|^2, \tag{4.8}$$

where in the 2D case R is the 2×2 matrix:

$$R(\Delta\theta) = \begin{bmatrix} cos(\Delta\theta) & -sin(\Delta\theta) \\ sin(\Delta\theta) & cos(\Delta\theta) \end{bmatrix}, \tag{4.9}$$

and $\mathbf{t}$ is a 2-vector:

$$\mathbf{t} = [\Delta x, \Delta y]^T, \tag{4.10}$$

note that the notation $\|\|$ denotes Euclidean distance. Minimisation of such a function is not limited to the work described in this thesis - this a general problem that occurs when trying to register data sets arising from any sensor that produces point cloud data e.g. LIDAR. The paper by Nüchter et al [94] provides an in depth review of the known closed form solutions and also presents some linearised techniques. This problem may be cast as a constrained optimisation where the constraint pertains the particular form that the matrix R may assume. A rotation matrix is one in which the columns are orthonormal, $\det(R) = 1$ and $R^{-1} = R^T$. The closed form method developed in [95] based upon the Singular Value Decomposition (SVD) of the correlation matrix generated from the point correspondences was selected for its simplicity and ease of implementation. This selection is supported by the paper by Lorusso et al [96] in which it is demonstrated that the SVD approach performs well in terms of accuracy and

stability with respect to degenerate datasets in a simulation based comparison with three other well known algorithms. The centroid of each dataset M and D is firstly calculated:

$$\mathbf{c}_m = \frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i \tag{4.11}$$

$$\mathbf{c}_d = \frac{1}{N}\sum_{i=1}^{N}\mathbf{d}_i \tag{4.12}$$

and removed from the points comprising D and M respectively such that each point cloud has been translated to the origin. The correlation matrix C is then calculated as according to Equation 4.13:

$$C = \sum_{i=1}^{N}\mathbf{m}_i'^T\mathbf{d}_i'^T = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}, \tag{4.13}$$

where $\mathbf{m}_i' = \mathbf{m}_i - \mathbf{c}_m$ and $\mathbf{d}_i' = \mathbf{d}_i - \mathbf{c}_d$. The SVD of C is evaluated to yield the following decomposition:

$$C = U\Lambda V^T, \tag{4.14}$$

where U and V are unitary matrices and Λ is a diagonal matrix with non-negative real numbers on the diagonal. Given these matrices rotation is calculated as follows:

$$R_{t-1}^t = VU^T, \tag{4.15}$$

where the change in angle can be recovered by:

$$\Delta\theta = \arctan(\frac{R^t_{t-1,12}}{R^t_{t-1,22}}). \tag{4.16}$$

Note that computation of the rotation matrix is independent of the units of the coordinates used to generate the correlation matrix, C. Given, R^t_{t-1}, the translation between the point clouds is given by:

$$\mathbf{t} = \mathbf{c}_d - R^t_{t-1}\mathbf{c}_m \tag{4.17}$$

Equation 4.17 computes the translation in the frame of reference of the camera F_c, see Figure 4.8(b). The trajectory of the robot is recursively estimated via the following set of update equations:

$$\mathbf{p}_{t,\theta} = \mathbf{p}_{t-1,\theta} + \Delta\theta \tag{4.18}$$

$$\mathbf{p}_{t,(x,y)} = \mathbf{p}_{t-1,(x,y)} + R^W_r(\mathbf{p}_{t,\theta})\mathbf{t}', \tag{4.19}$$

where the term $R^W_r(\mathbf{p}_{i,\theta})$ is:

$$R^W_r(\mathbf{p}_{i,\theta}) = \begin{bmatrix} cos(\mathbf{p}_{i,\theta}) & -sin(\mathbf{p}_{i,\theta}) \\ sin(\mathbf{p}_{i,\theta}) & cos(\mathbf{p}_{i,\theta}) \end{bmatrix}, \tag{4.20}$$

is the rotation matrix mapping the translation calculated in the F_r to the global world frame, F_w, and $\mathbf{t}' = [-\Delta x, \Delta y]^T$. The update equations are independent of an underlying kinematic model describing vehicle motion across successive time steps and thus the applicability to vehicles such as the 4 wheel drive robot in Figure 4.1 is clear.

4.7.1.1 Transform Mean Square Error

The transform Mean Square Error (MSE) is given by instantiating Equation 4.8 with the model parameters calculated in the preceding steps.

$$MSE = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{m}_i - (R_{est}\mathbf{d}_i + \mathbf{t}_{est})\|^2 \tag{4.21}$$

Pose estimation in the robot frame, F_r, is achieved through application of the transform relating the camera and robot coordinate frames in accordance with Equation 4.22. Note that this equation is applied before mapping the pose estimate into the global frame.

$$\mathbf{p}_t^r = \mathbf{p}_t^c - R(\mathbf{p}_{t-1,\theta})\mathbf{t}_{offset} \tag{4.22}$$

4.8 Estimate of Transform Uncertainty

As a result of the recursive nature of the estimation the uncertainty in the visual odometry estimate can grow without bound - this is the case for any odometry system. A useful property of the visual technique is that there is a clear method for characterisation of this uncertainty through analysis of the inter image transform. This uncertainty arises from:

1. Detector Noise
2. The random behaviour of RANSAC

As a result of 1) the relationship between the correspondences is in general non-rigid necessitating the optimisation procedure to determine the best fitting rigid parameters. A sequence of static images of the materials considered in Section 5 is used to evaluate detector noise. In 2) the features selected to form the set

of inliers is a function of the RANSAC threshold, ϵ, and thus the rigid body parameters can be viewed to be a function of ϵ; the influence of this parameter on the transform is investigated.

4.8.1 Detector Noise

A sequence of 50 images for the materials (b)-(d) shown in Figure 4.9 was acquired from a stationary robot. In order to consider the noise associated with the detector, a particular point, $\mathbf{p}$, was extracted in image$_1$ and a set of correspondences, $\mathbf{p'}_i$, were detected in the remaining images using the descriptor obtained in image$_1$. The variances in the x and y components for each material are shown in Table 4.4. The steel and aluminium samples display differing variances for the x and y components while those for the bricks and rusted steel are of similar value. These measured variances will be used as an approximation of detector noise when the robot is in motion. These values do not take account of image blurring, however, this was minimised as far as experimentally possible through prudent selection of the camera shutter time, therefore, blur made negligible contribution to the computed variances.

Material	σ_x^2 (Pixels^2)	σ_y^2 (Pixels^2)
Steel Plate	0.17	0.023
Bricks	0.012	0.012
Rusted Steel	0.010	0.023
Lab Floor	0.025	0.058

Table 4.4: Detector variance evaluated for different materials

4.8.2 Effect of Threshold on Outlier Rejection

SIFT and the matching procedures are deterministic functions, RANSAC injects a source of random behaviour. One of the key parameters controlling RANSAC is the distance threshold, ϵ, used to determine if a data point is an inlier or oulier from its distance from the corresponding model point. As ϵ is increased the mean number of inliers increases, this is shown in Figure 4.12. Using the same input image pair ten equally spaced ϵ settings lying in the range $[0.001, 0.01]$ are plotted against mean number of inliers where ten trials have been used to compute the mean. The error bars denote the maximum and minimum values obtain during these trials. Note that the data points are normalised such that their average distance from the origin is $\sqrt{2}$ and thus a fixed valid range of $[0.001, 0.01]$ was used for ϵ.

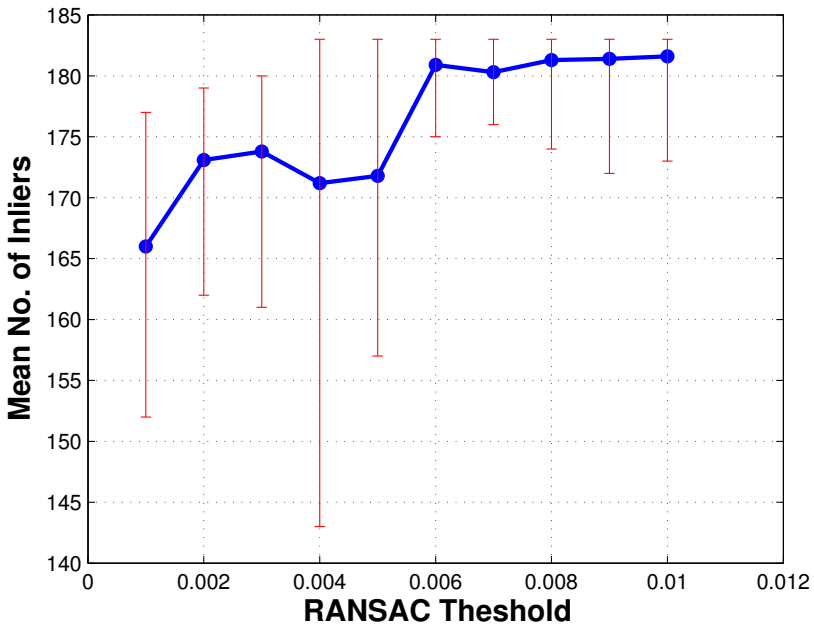


Figure 4.12: Plot of the mean number of inliers over a range of RANSAC threshold settings

The distance threshold controls the number of points available to compute the parameters of the rigid body transform relating two images. The purpose of this section is to consider the effect of this threshold on the variability of the resultant transform. The distance threshold is swept from $[0.001, 0.01]$ to determine the variance of the transform parameters - this procedure is carried out for each type of type of material in Figure 4.9 apart from aluminium. This test was performed on several image pairs from each of the preceding datasets. For a representative image pair from each dataset the variance of the computed transform parameters are shown in Table 4.5. The conclusion that can be drawn here is that for the particular type of images considered in this scenario, the distance threshold has little influence on variability of the parameters.

Material	σ_x^2 (Pixels^2)	σ_y^2 (Pixels^2)	σ_θ^2 ($^{\circ 2}$)
Steel Plate	5.23×10^{-4}	7.99×10^{-4}	3.26×10^{-5}
Bricks	3.67×10^{-5}	7.39×10^{-5}	2.02×10^{-6}
Rusted Steel	1.76×10^{-4}	1.82×10^{-4}	1.44×10^{-6}

Table 4.5: Variance of transform parameters resulting from ϵ sweep

4.9 System Evaluation

The system was evaluated on four different surface types using the differential drive robot shown in Figure 4.8(a). This enabled a comparison with encoder odometry to be made. Given the recursive nature of the estimation, it was expected that the system would display accurate inter image alignment but would exhibit large error over long distances due to drift - the results show this to be the case. The error with respect to Vicon ground truth is evaluated for each image set. Visualisation software (developed by Dr Gordon Dobie in CUE) enabling images to be displayed inside a virtual world with a specific (x, y) location and orientation θ was used to create large image composites to demonstrate the stitching function of the system.

4.9.1 Experimental Setup

Vicon markers were affixed both to the centre of rotation of the robot defining the robot frame, F_r, and directly onto the camera to define the camera frame, F_c, the configuration is shown in Figure 4.8(b). The visual trajectory is estimated in F_c while encoder and ground truth are estimated in F_r. In the experiments presented in the Results section 4.9.2 the transform mapping F_c to F_r as determined from Vicon data was applied to the image based estimate to yield robot centred coordinates allowing a comparison with the encoder odometry and ground truth.

Prior to conducting the experiments the UMBmark [97] procedure was carried out to fine tune the robot kinematic model parameters in the form of the inter wheel distance and wheel diameters ensuring optimal encoder based pose estimation.

4.9.2 Results

This section presents the results of applying the algorithm on four datasets comprising images of:

1. Grid paper (open loop)
2. Steel plate (open loop)
3. Laminate surface (open loop)
4. Laminate surface (closed loop)

where *open loop* indicates that the robot updates its position from encoder odometry while *closed loop* denotes Vicon positional updates. The technique is firstly applied on a calibration dataset consisting of images of grid paper. Importantly, this dataset demonstrates the optimal operation of the ground truth system and was used to set the scaling factor mapping image pixels to millimetres. An NDE dataset comprising of images of a steel plate (Figure 4.9(b)) is used to demonstrate the technique in a realistic inspection scenario. The laminate floor surface

was selected on the basis of providing a large area to test the method and the easily identifiable features within the images to make apparent alignment issues. The type of path used in the steel plate dataset is repeated for application on the laminate floor and comparisons are made. Finally a large dataset consisting of laminate floor images is used to compare the short term and long term accuracy of the system; closed loop control was used in order to ensure sufficient overlap for illustrating image alignment.

4.9.2.1 Grid Paper Dataset

An image composite of grid paper, composed of square cells of side $10mm$, constructed from ground truth data is shown in Figure 4.13 (a). The straight lines composing the grid are approximately aligned with lateral errors being $\approx \pm 2$ pixels. The mapping from pixels to millimetres was determined through averaging the scaling factor over several grid cells in the virtual environment.

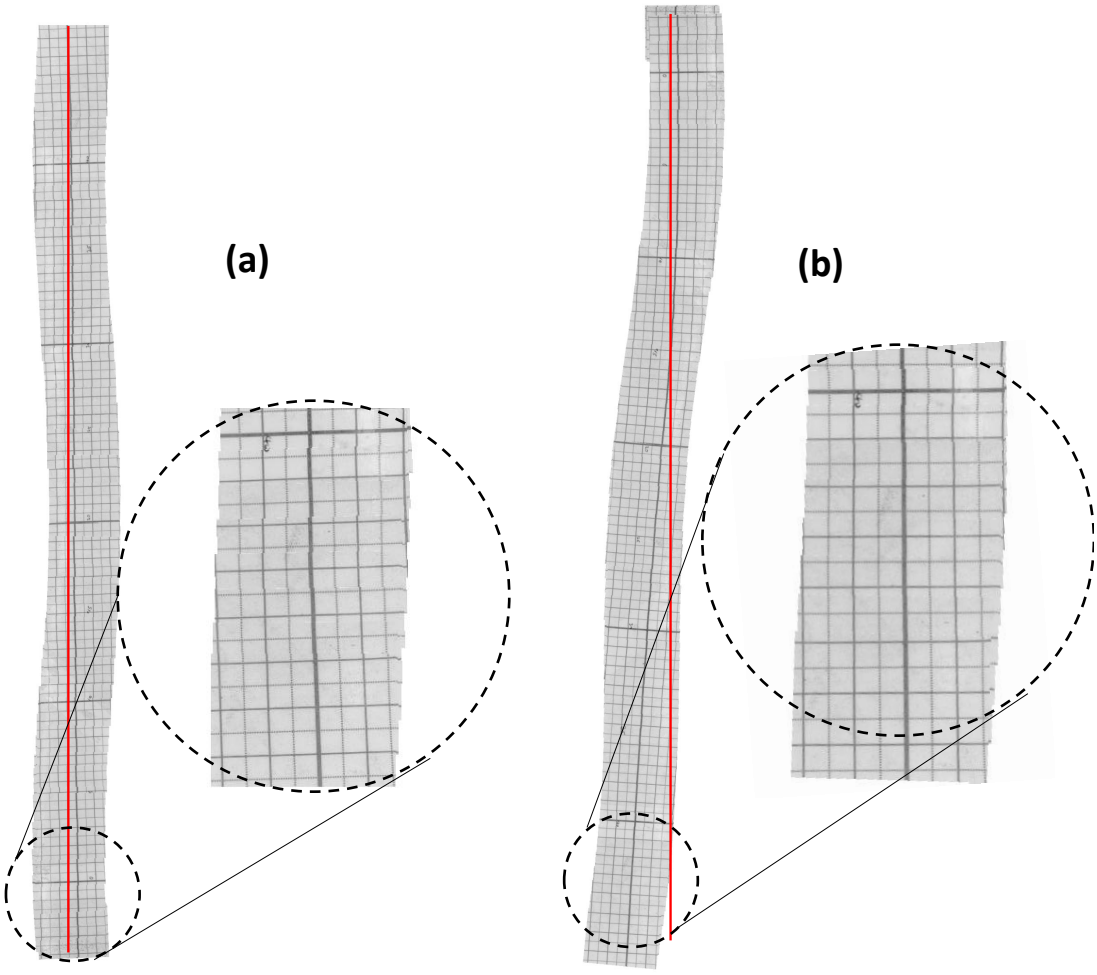


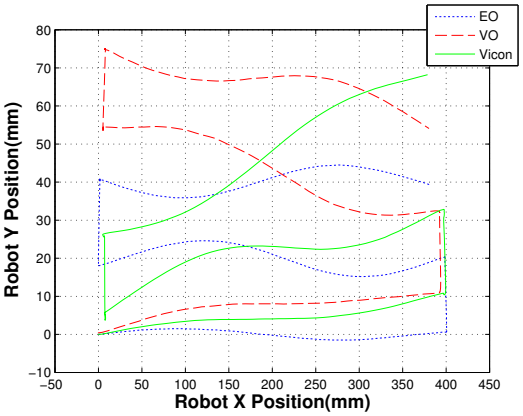
Figure 4.13: Grid paper dataset (a) Vicon placement. (b) Visual odometry placement. The enlarged subsections of each path displays improvement in image alignment resulting from the visual odometry. The complete mosaic exhibits drift due to the build up of error.

The regular structure of the grid can be used to show some properties of the algorithm, namely that of accurate image to image alignment and overall drift. Using image data only the visual odometry was run to produce the composite displayed in Figure 4.13 (b). The image to image alignment has improved as shown by the enlarged subsections, however, over the length of the path driven

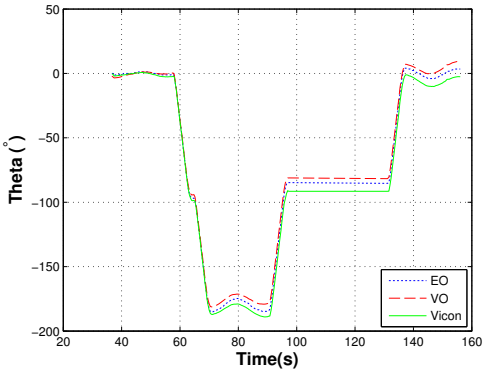
by the robot, the departure from the true rectilinear pattern is significant. The error in the linearity of the grid appears large as indicated by the overlaid red line, however, the inter image error is small; this occurrence may be explained by there being a large number of small of errors distributed across a long distance.

4.9.2.2 Steel Plate Dataset

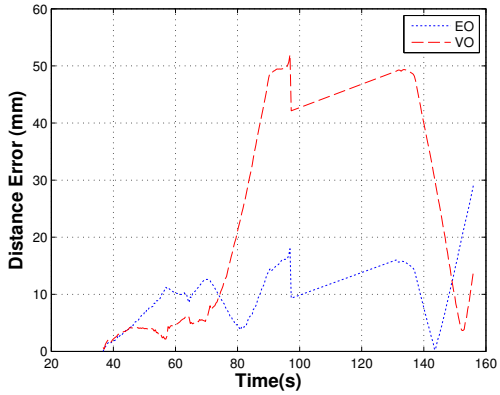
This dataset demonstrates the technique on a realistic NDE sample in which the robot (under open loop control) completes a partial raster scan consisting of 400 mm and 100 mm horizontal and vertical sweeps respectively. Figures 4.14(a) and (b) show the XY and θ estimates of the path while the errors with respect to ground truth for both visual and encoder odometry are shown in Figure 4.14 (c) and (d). The major error in the path occurs between sections (2) and (3) in Figure 4.15(b) where the angle is incorrectly estimated. The first point of intersection at $\approx$ 70 seconds of the curves in Figure 4.14 (c) coincides with a large bias in the angular estimate around the same time. Note that the error curves in Figure 4.14 (c) and (d) display a strong correlation indicating that the two systems have similar behaviour. Neither system accurately estimates the true path of the robot with visual odometry giving rise to greater error (integral of visual odometry error curve) than the encoders. The raw matches and RANSAC filtered matches in Figure 4.14(e) show a strong correlation indicating that the feature descriptors are distinctive. The transform MSE in Figure 4.14(f) attained a maximum value of 0.01 mm^2 with a mean of about 0.005 mm^2. The mosaics resulting from ground truth and visual odometry are shown in 4.15 (a) and (b) respectively. Note that the arcs result from F_c not being coincident with F_r.



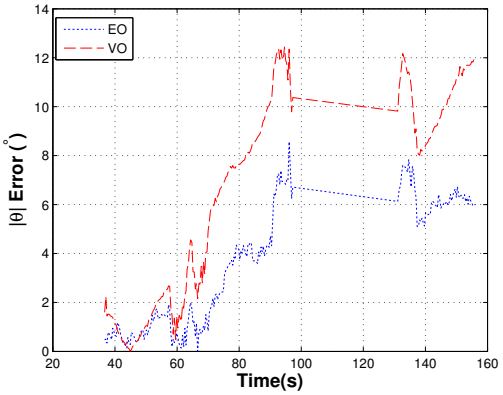
(a) X-Y Estimate



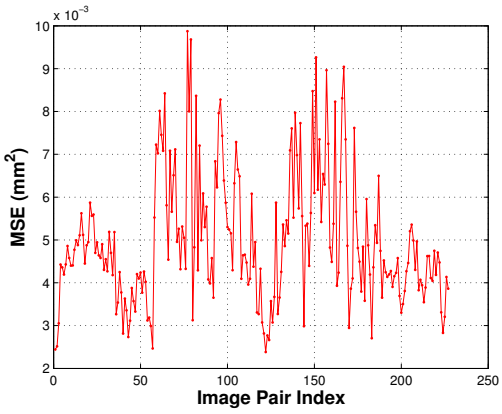
(b) θ Estimate



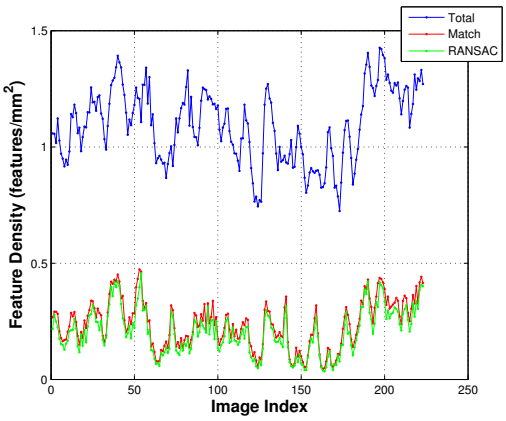
(c) X-Y Error



(d) $|\theta|$ Error



(e) Inter image transform MSE



(f) Feature Densities

Figure 4.14: Steel Plate Dataset. The mean feature densities for total, matched and RANSAC were 1.092, 0.30 and 0.26

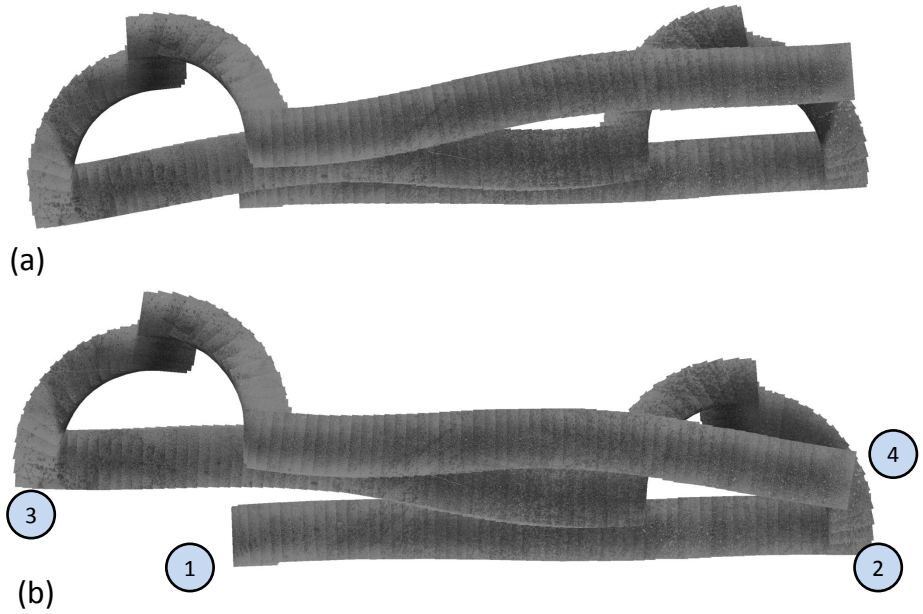
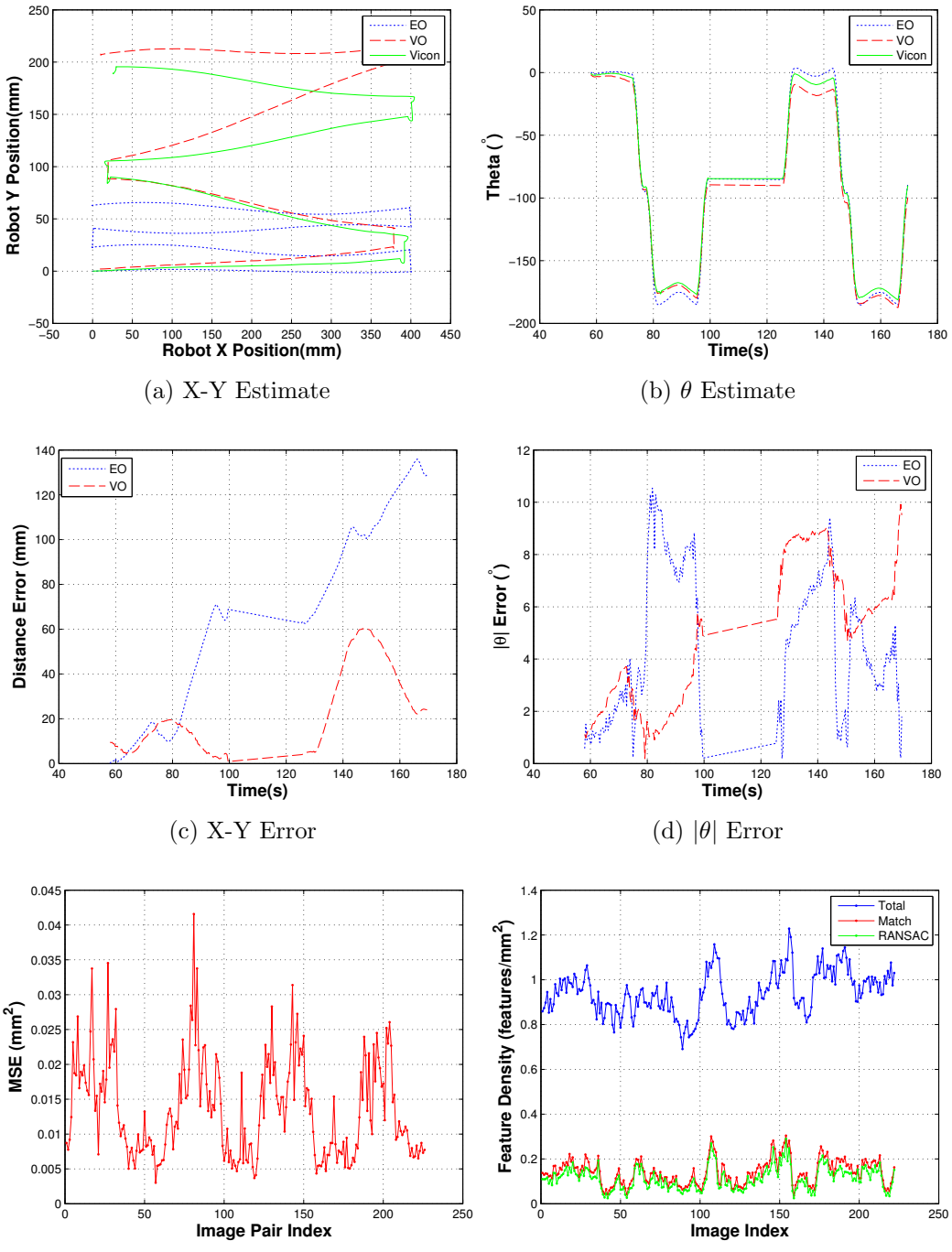


Figure 4.15: Steel plate mosaic (a) Vicon placement (b) Visual odometry placement

4.9.2.3 Laminate Floor Dataset (Open Loop)

This section applies the method to the laminate floor of the laboratory where the robot was instructed to complete a similar motion to that conducted in the steel plate dataset. The horizontal and vertical sweeps were 400 mm and 25 mm during this test. In this and the following dataset, a grid of 1 m x 1 m cells composed of masking tape was applied to the surface such that the grid lines spanned multiple images:this is used to highlight misalignment. The XY and θ estimates are shown in Figures 4.16(a) and (b). As shown by Figure 4.16(a), visual odometry in this cases produces less error than the encoders in terms of the distance error. The angular error is also smaller for visual technique compared to encoder odometry as shown in Figure 4.16(d). The mosaics resulting from Vicon and visual odometry are show in Figure 4.17 in which the path has been divided

into sections (1) - (4) to aid the analysis. As can be seen from Figure 4.17(b), the greatest error occurs between sections (3)-(4) which correlates with a large bias in the estimate of θ in Figure 4.16(b). The areas showing the masking tape have been highlighted. In Figure 4.17(a) a continuous grid line is indicated by the red dashed line, however, the same line in Figure 4.17(b) is reconstructed as two lines: the red and blue which should be the same. The raw matches and RANSAC filtered matches in Figure 4.16(e) show a strong correlation indicating that the feature descriptors are distinctive. The transform MSE in Figure 4.16(f) attained a maximum value of 0.04 mm^2 and a mean of about 0.015 mm^2.



(a) X-Y Estimate

(b) θ Estimate

(c) X-Y Error

(d) $|\theta|$ Error

(e) Inter image transform MSE

(f) Feature Densities

Figure 4.16: Laminate floor dataset. The mean feature densities for total, matched and RANSAC were 0.94, 0.20 and 0.15 respectively.

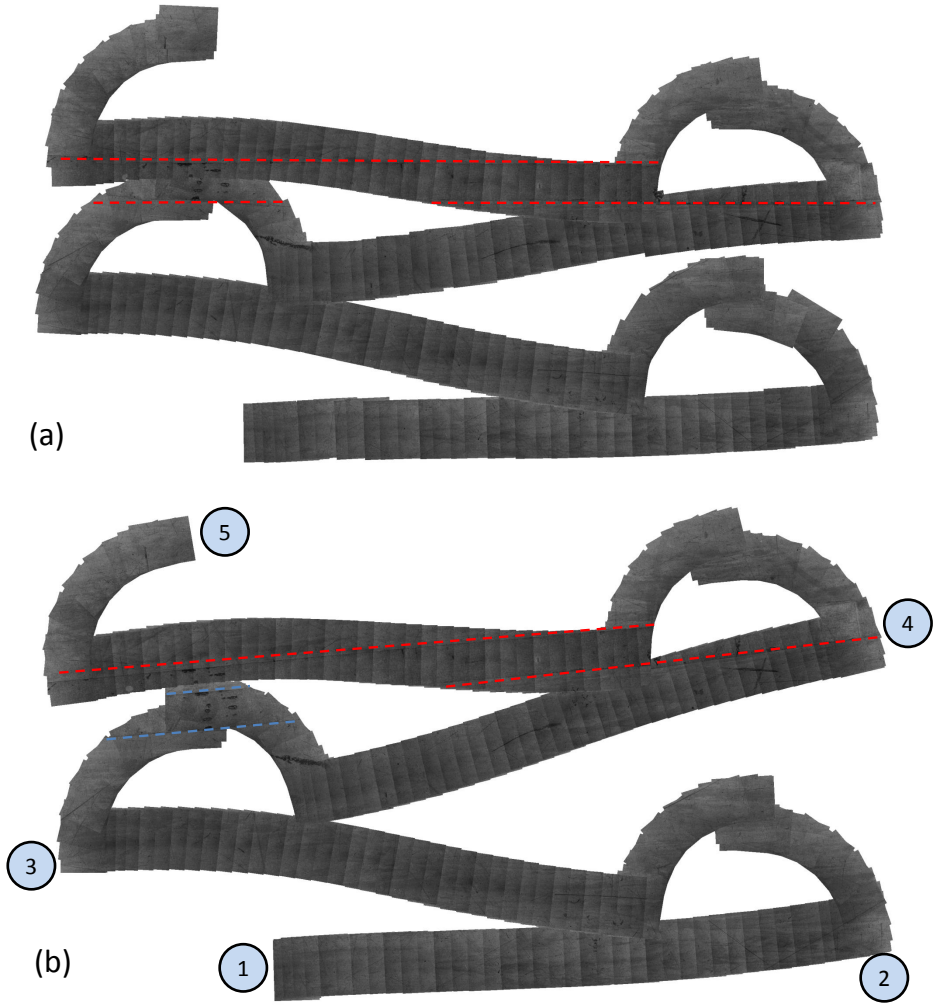


Figure 4.17: Laminate floor mosaic (a) Vicon placement (b) Visual odometry placement.

4.9.2.4 Comparison of steel and laminate floor datasets

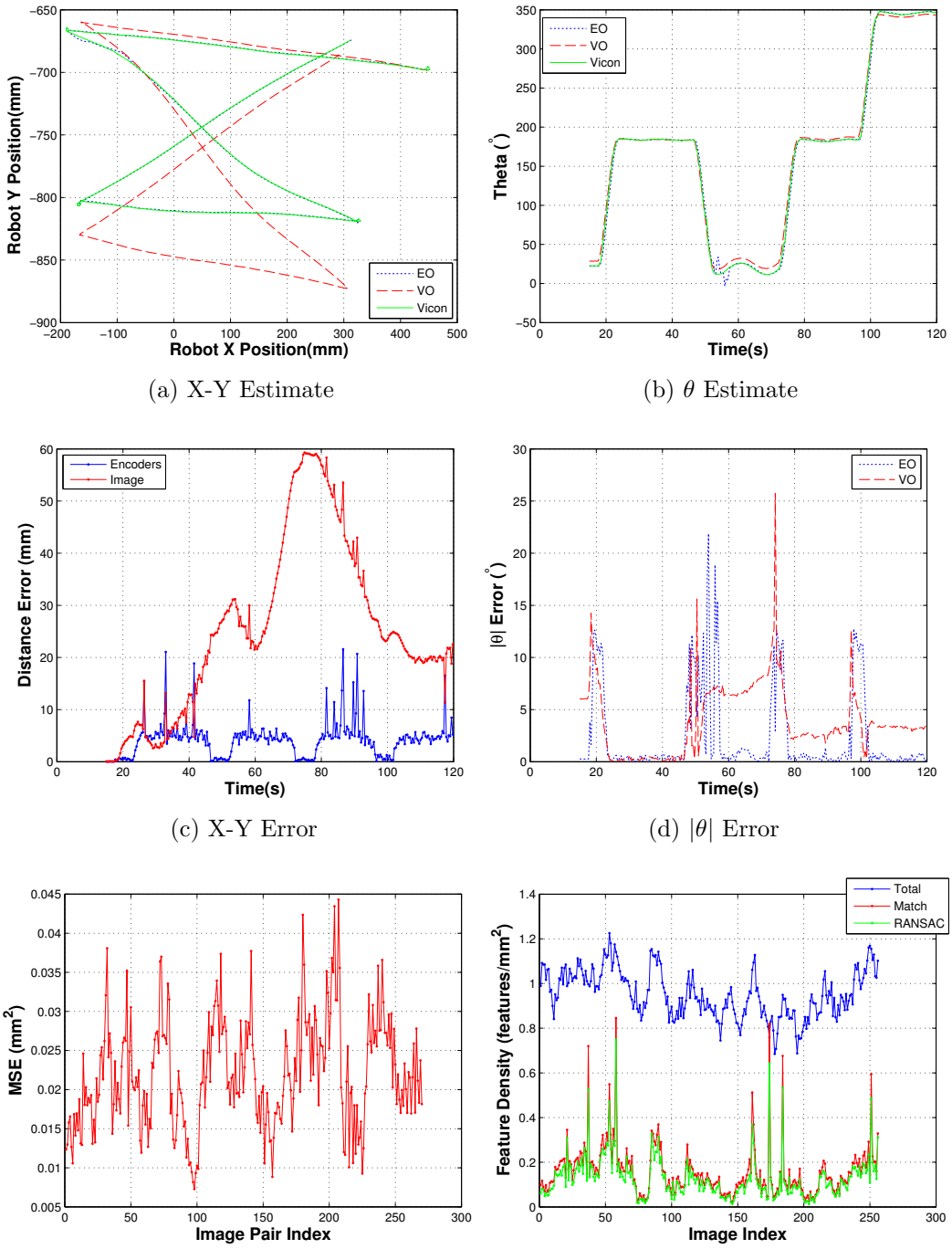
In Equation 4.17, translation is expressed as a function of rotation and thus error in the former is compounded by any error in the estimate of the latter. The accuracy with which the rotation is estimated is therefore of key importance for reducing drift. The magnitude of the θ error in Figure 4.16(d) for the laminate surface is upper bounded by 4° until $\approx 100s$ after which point the path estimate begins to drift substantially. For steel, in Figure 4.14(b) the angular error increases more quickly than in the laminate surface case causing an earlier onset of drift in the visual odometry path.

4.9.2.5 Laminate Floor (Closed Loop)

A dataset acquired from executing a figure of eight path over the laminate surface is used to demonstrate the technique on a large area. The figure of eight path was chosen such that the dataset contained images of the same area distributed with large temporal separation across the dataset - this is used to highlight alignment problems. The XY and θ estimates for the encoders and visual odometry are shown in Figures 4.18(a) and (b). As a result of closed loop control the encoder estimate is very closely aligned to Vicon ground truth in Figure 4.16(a). However, due to time delays associated with applying path corrections calculated from Vicon they are not equivalent as shown in Figure 4.18(c) and (d). In these error plots the encoder and visual odometry curves display a strong correlation, up to around 45 s, thereafter the visual odometry XY error increases substantially as a result of angular error. The raw matches and RANSAC filtered matches in Figure 4.18(e) are strongly correlated indicating that the feature descriptors are distinctive. The transform MSE in Figure 4.18(f) had a maximum value of 0.045 mm^2 and a mean of 0.02 mm^2.

The composites resulting from the ground truth and visual odometry pose estimates are shown in Figures 4.19 (a) and (b) respectively. The areas showing the masking tape have been highlighted where the red, green and blue dashed lines indicate the same grid line on the surface. Note that the path has been divided

into sections (1) - (5) to aid the analysis. In Figure 4.19 (b) it is evident by inspection that the images between locations (1) and (2) have been assembled locally in an accurate fashion, this is shown by the alignment of the long piece of tape as highlighted by the green lines. Figure 4.19 (a) shows a similar result. In section (2) - (3), there exists an area of overlap with (1) - (2) located after the turn. The tape within this region should ideally be aligned with the tape imaged in section (1) - (2) which is the case for the Vicon composite, however, the build up in error in the visual odometry case has caused drift resulting in misalignment: this shows the global error. Surprisingly, in the upper region of the path between locations (2) and (4) (highlighted in red).



(a) X-Y Estimate

(b) θ Estimate

(c) X-Y Error

(d) $|\theta|$ Error

(e) Inter image transform MSE

(f) Inliers

Figure 4.18: Errors for laminate floor figure of 8 dataset. The mean feature densities for total, matched and RANSAC were 0.93, 0.19 and 0.17 respectively.

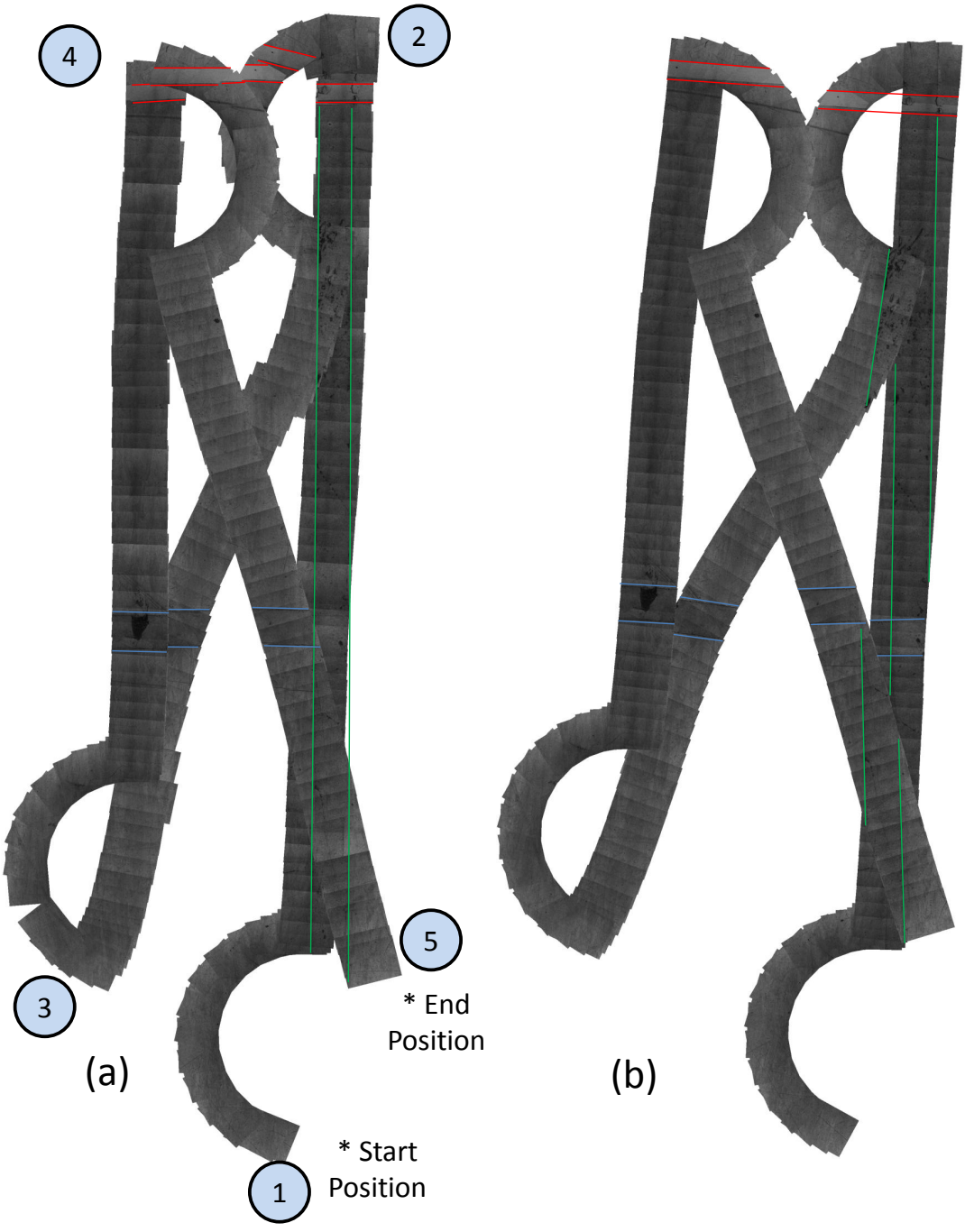


Figure 4.19: Figure of 8 mosaic on laminate floor (a) Vicon based composite (b) Visual odometry based composite.

4.9.3 Discussion

This section discusses the advantages and disadvantages of the visual method for odometry when compared to the encoder based technique. Together the two methods offer the capability to estimate motion in a range of environments. The major benefit of visual odometry is that it can be utilised in slippery environments and can be used for vehicles where wheel slippage plays a major part in the method of locomotion. It is, however, reliant upon a constant height above the surface, uniformity of illumination and availability of surface texture. Visual odometry also makes use of computationally intensive image processing and optimisation in order to produce the pose estimates. Unlike the visual method, encoder odometry requires a kinematic model of the robot but is otherwise comparatively simple enabling pose estimates to be made on-robot with minimal computational resource. The accuracy of the encoder odometry technique is a function of the kinematic model parameters and the amount of wheel slip experienced. The accuracy of visual odometry is primarily a function of the quality of feature matching which in turn is a function of the illumination and distinctiveness of surface features. Given a scenario in which the conditions are such that either system could be applied, encoder odometry would be the preferable choice due to its lower computational requirement. There also exists the possibility of using both encoder and visual odometry in a hybrid approach where, for example, the visual estimates could potentially compensate for wheel slippage. The correlation between feature matches and RANSAC filtered matches in combination with the transform MSE could be used to weight the contribution of the visual estimate to the the overall estimate of motion. Fusing data in this manner could be implemented using a Bayesian filtering approach, this technique is demonstrated for fusing different positional data in Chapter 6.

4.10 Sources of Error

This section discusses a number of error contributing sources in the estimation process. Aligning the virtual VMCS object coordinate frames, F_c and F_r with the physical axes of the robot is not a straightforward problem. With reference to Figure 4.8(b) it can be seen that the x-axis should ideally be normal to the axis connecting the wheels of the robot, any discrepancy in this location could cause a systematic error in the estimate of angle. The corresponding axes of F_c and F_r would ideally be parallel, angular error in the configuration of these axes could again induce systematic errors into the visual odometry estimate. In order to reduce coordinate frame alignment errors a tool was manufactured to systematically align the frames; this procedure aligned the frames as best as experimentally possible. Using this technique the translation error was less than a millimetre while the angular error was less than half of a degree.

4.11 Conclusion/Future Work

This chapter has described an image based motion estimation system that is applicable in planar operating environments. The primary objective of the system is to provide a means of pose estimation for vehicles for which encoder based estimation is not possible. Such a robot is shown in Figure 4.1, where due to the dependence upon wheel slip for turning, encoder based odometry would drift very quickly. Features extracted from images are used to infer camera motion. This inference involves finding the parameters that optimally align successive images. Alignment of the images is equivalent to stitching and thus the system has a desirable NDE based output of creating mosaics of the surface upon which the robot is travelling. Importantly the composite image could then be used in performing visual NDE. The applicability of the feature extraction technique was investigated for materials of the type encountered in NDE and shown to be usable

for matte texture rich surfaces and problematic on shiny surfaces. The system was demonstrated on several datasets and compared to Vicon ground truth and encoder odometry.

The visual odometry technique was tested on steel and laminate surfaces under controlled laboratory conditions. Both surfaces yielded approximately 1 features/mm^2 while the RANSAC filtered matches were 0.26 features/mm^2 for steel and $0.15 - 0.17$ features/mm^2 for laminate. Although the feature densities in the steel dataset are greater, the angular error increases more quickly than for the laminate surface resulting in drift earlier in the visual odometry path. In the steel dataset the greater error for encoder odometry was 28 mm while the visual odometry was 52 mm. However, for the open loop laminate case the maximum error was 60 mm while the error for the encoders was around 138 mm. In the closed loop laminate dataset, visual odometry again attained a maximum error of 59 mm.

The presented system operates in an incremental fashion whereby incoming images are immediately processed (potentially in real time) and integrated into the estimate of vehicle pose. Drift is a problem that stems from the recursive nature of the estimation where *only* successively overlapping images are considered. In the path driven by the robot, however, the same point may be passed at non-successive times as was shown in the dataset of Section 4.9.2.5. Taking account of such non-sequential overlap through forcing *all* common features sets to align optimally could substantially reduce drift. The optimisation would take as input the image positions as specified by the visual odometry and then look to refine them through the non-successive overlap. Such an optimisation could be implemented in the framework of Structure from Motion [98] or SLAM and would be the logical extension to the work presented in this chapter. The work presented here has made use of algorithms to extract features from images of the operating environment. However, the techniques are not limited to visual data, features extracted from other types of NDE sensors could be used for example sub-surface defects detected by Eddy currents or defects detected by ultrasond fired into the structure.

Chapter 5

Air Ranging Methods

5.1 Introduction

The estimation of range by a positioning device from itself to an object is fundamental to determine the objects location. A review of the conventional methods to compute range was presented in Chapter 2 where, in summary, the techniques consisted of triangulation, time of flight and phase based methods. Indeed, each 3D positioning system employed in this thesis has used one of these techniques: the VMCS uses stereo vision based triangulation and the LAT derives distance from a phase measurement. This chapter introduces the Cricket Indoor Location System, a low cost, modular, ultrasonic based positioning system that uses the time of flight of two pulses to compute range. The accuracy of this system is characterised in the 1D and 2D cases and calibration procedures are developed to improve this accuracy. The results of this analysis are used in the following Chapter 6 in which Bayesian filtering is used to fuse ultrasonic measurements with wheel encoder data.

Of the systems considered in this thesis, the ultrasonic positioning system is particularly attractive in terms of its relative low cost and small footprint; this property could be of interest in access restricted inspections. Consider an inspection by a multi-robot system where access limitations preclude use of the laser

tracker. In this scenario, the modularity of the ultrasonic system gives rise to the possibility of utilising, as the nodes of a positioning system, a subset of vehicles equipped with the associated modules. The locations of such vehicles would become fixed upon entry into the inspection site and following the determination of inter module distances, these robots would act as transmitters to provide positional updates to vehicles conducting the inspection. With a view to creating a more *accurate* version of the ultrasonic system in the future through improvement of the specific ranging technique employed, this chapter also provides a brief comparison of several methods for through-air ultrasonic ranging.

5.2 Ultrasound

Acoustic waves with a frequency greater than 20 kHz (upper limit of human hearing) are said to be ultrasonic. Ultrasonic waves are used in variety of applications including ranging, chemical processing and medical/NDE imaging. Ranging techniques using pulsed waveforms such as ultrasound were discussed in Section 2.3.4. A fundamental limit in ranging applications using sound waves is knowledge of the speed of sound in the propagation medium which is greatly effected by temperature. The speed of sound in air is a function of temperature and therefore error in measuring temperature is propagated to error in range. The dependency is expressed as follows:

$$c = \sqrt{\frac{\gamma p}{\rho}} = \sqrt{\frac{\gamma RT}{M}}, \tag{5.1}$$

where γ is the adiabatic index of air, p is the pressure, and ρ is the density. Note that Equation 5.1 can be approximated by

$$c = 300 + 0.6T, \tag{5.2}$$

where T is in Celsius. Given Equation 5.1, the fractional error in a range measurement $\Delta d/d$ (where Δd is the standard deviation, or error, of the range measurement d) may be computed as a function of the fractional error in both time of flight $\Delta t_F/t_F$ and temperature $\Delta T/T$ using the following equation:

$$\left(\frac{\Delta d}{d}\right)^2 = \left(\frac{\Delta t_F}{t_F}\right)^2 + \frac{1}{4}\left(\frac{\Delta T}{T}\right)^2. \tag{5.3}$$

Equation 5.3 shows the difficulty in obtaining accuracy in many practical scenarios. In Equation 5.3, the fractional error derived from an uncertain temperature can dominate the fractional error in the time of flight measurement and thus the fractional error in distance would be dominated by the temperature inaccuracy.

Methods for temperature-compensated ranging include direct temperature measurement through the use of thermistors and resistance temperature detectors. If there are temperature gradients in the medium, however, then this method could result in poor accuracy.

In practice, the uncertainty in the speed of sound may be reduced through measuring it directly. Chande et al [99] describes a method that avoids the need to accurately measure temperature. The technique measures the time-of-flight of an acoustic pulse over a known distance while simultaneously ensonifying the environment. The speed of sound is measured independently of the medium, assuming that the fixed distance is known accurately.

5.3 Acoustic Beacon Location System

A commercially available indoor acoustic positioning system was used to provide global position measurements. The fundamental operating principal is described in Section 2.3.4. Developed by Priyantha [100], the ultrasonic system provided an update rate of 3 Hz. The system comprises of a collection of modules each configurable to be a transmitter or receiver, a module is shown in Figure 5.1. The

system performs multi-lateration through measurement of at least three inter-module distances. The distances are estimated through measurement at the receiver of the time difference of arrival (TDoF) between the transmitter emissions of an ultrasonic pulse and radio frequency signal. The radio signal encodes the module identifier while the ultrasonic pulse serves to enable the TDoF calculation. The piezoelectric transducers, in Figure 5.1, have a resonant frequency 40 kHz and are used for transmission/reception of the ultrasonic pulse. The 1 cm aperture and operating frequency produces a beamwidth at the -3 dB point of $\pm 26^{\circ}$ with respect to the line perpendicular to the transducer face. The system compensates for changes of sound speed due to temperature through using the average of the temperatures measured (using thermistors) by each module.

Cricket may operate in two modes: the transmitters are fixed and the receiver is mobile such that the transmitters must use a time multiplexing approach (in order to avoid signal interference amongst different modules) or the alternate mode where the transmitter is mobile and the receivers are fixed resulting in non-simultaneous updating of the robots. The former is preferred when multiple robots are in use allowing the platforms to simultaneously update their locations.

Figure 5.1: A Cricket transceiver module

Assuming three receiver to transmitter distances have been acquired and given that the receivers lie on a ring with radius r the location of the transmitter is calculated by trilateration [100] as follows:

$$\begin{bmatrix} x_{tx} \\ y_{tx} \\ z_{tx} \end{bmatrix} = \begin{bmatrix} \frac{1}{2r}(d_1^2 - d_2^2 + r^2) \\ \frac{1}{2r}(d_1^2 - d_3^2 + r^2) \\ \pm\sqrt{(d_1^2 - x^2 - y^2)} \end{bmatrix}, \tag{5.4}$$

where d_1, d_2 and d_3 are three receiver - transmitter distances output from Cricket. The receiver corresponding to d_1 is the origin of the coordinate frame, the positive x-axis is defined by receivers 1 and 2 and the positive y-axis by receivers 1 and 3. The arrangement is shown in Figure 5.2 in which the modules were rigidly affixed to a frame such that the receiver transducers of each module were positioned to be 1 m apart and 1.1 m above the plane containing the sample. The z-axis of the system coordinate frame was defined by the line perpendicular to the sample surface passing through RX_1. The x and y axes were defined to lie along the edges of the sample.

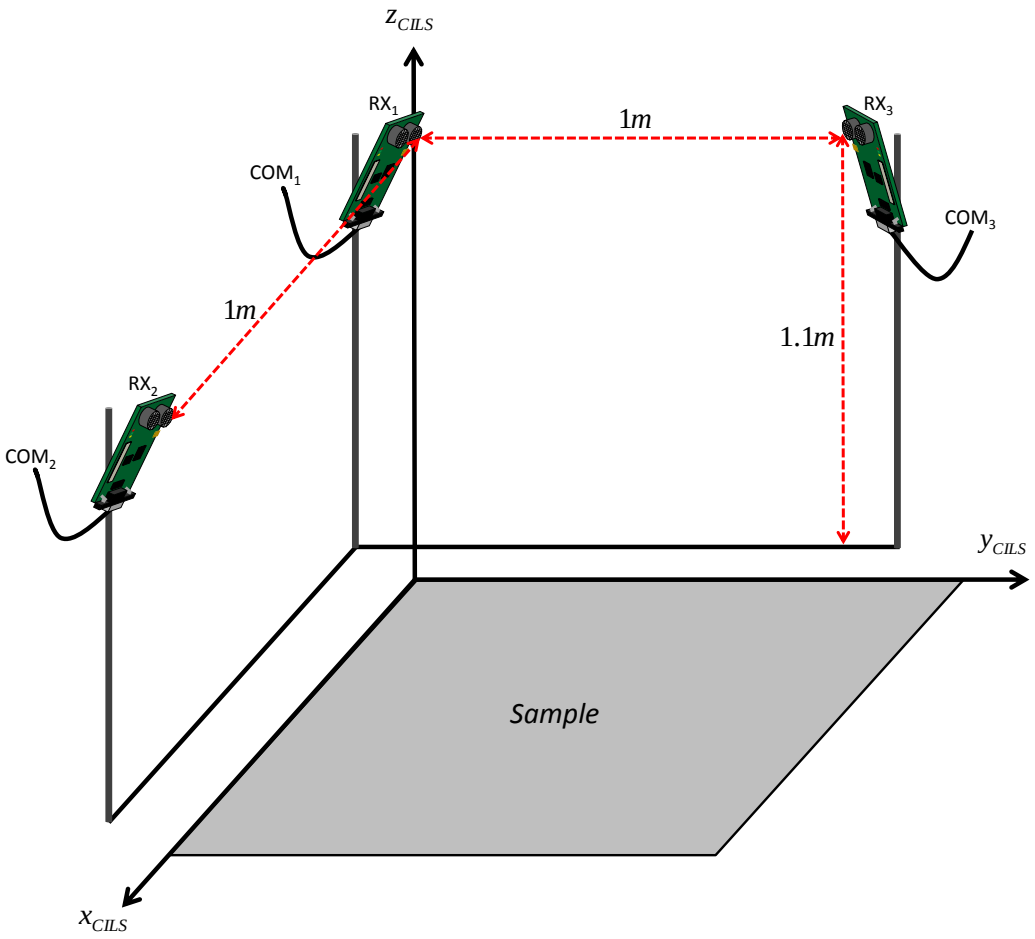


Figure 5.2: Schematic showing the experimental setup of the acoustic positioning system.

Given a robot (x, y) position and the location of the beacons, the distances can be easily recovered as follows:

$$d_b = \sqrt{(x - x_b)^2 + (y - y_b)^2 + z_b^2}, \tag{5.5}$$

where (x_b, y_b, z_b) are the beacon locations for $b \in \{1, 2, 3\}$. The following section quantifies the uncertainty in the Cricket estimated position/distance for use in the filtering algorithms.

5.3.1 Error Characterisation of Acoustic System

The accuracy and standard deviation of measurements for the ultrasonic system was considered in the simplest case of 1D followed by consideration of the 2D case assessed over the working area used in the experiment.

5.3.1.1 One Dimensional Test

The 1D experiment consisted of holding a transmitter static while a receiver unit was moved in increments of 20 mm along a measurement rail with 1 mm resolution from $0 - 2000$ mm. At each step 50 distance measurements were taken. The range of 2000 mm was chosen because it matched the distance covered in the experimental evaluation section. Plotting the mean measured ultrasonic distance against the true distance (the line $y = x$) yields the graph shown in Figure 5.3 which clearly shows an error in gradient and offset with respect to the ideal.It was considered, following inspection of the software, that this error was due to the assumed clock frequency of 8 MHz differing from the actual oscillator frequency running at 7.37 MHz. A calibration procedure was carried out to compute the gradient and y-offset terms that minimise the error between the ultrasonic measurements and true range: the residual errors are shown in Figure 5.4 with mean, maximum and minimum errors of -1.80×10^{-5} mm, 4.18 mm and -4.79 mm respectively. The standard deviation evaluated from the 50 distance measurements at each location as a function of range is plotted in Figure 5.5 with mean, maximum and minimum values of 1.45 mm, 3.37 mm and 0.53 mm respectively. The uncertainty of the distance readings returned from Cricket was evaluated after the spatial realignment and it was found that the histogram of

distances pertaining to each measurement location was a function of the grid position and in the worst case followed a zero mean Gaussian density with a variance of 23 mm^2.

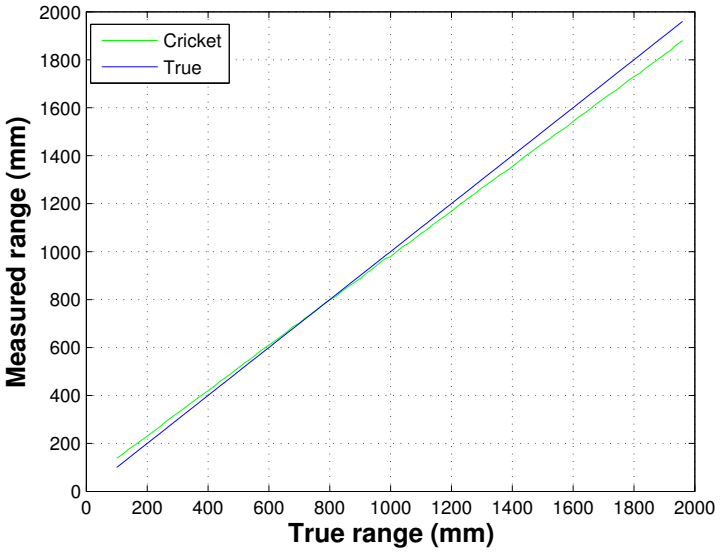


Figure 5.3: Average ultrasonic measured distance vs True distance over a distance of 0 to 2000 mm

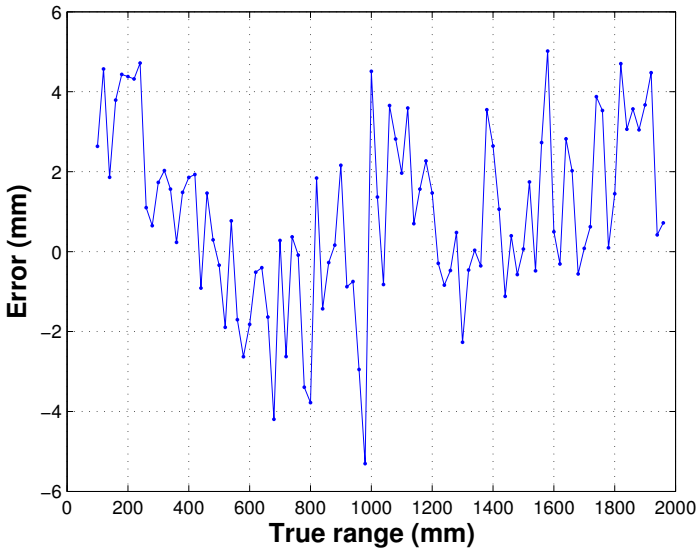


Figure 5.4: Error in ultrasonic measurements following 1D calibration procedure

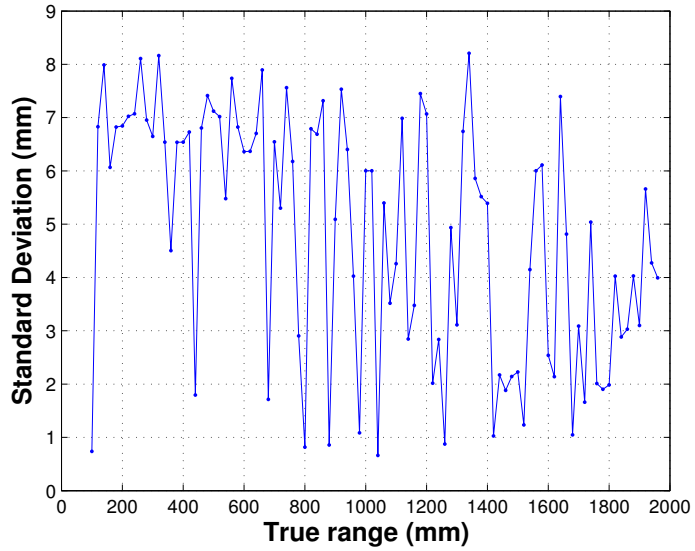


Figure 5.5: Standard Deviation in ultrasonic distance measurements vs True distance over a distance of 0 to 2000 mm.

5.3.1.2 Two Dimensional Test

The positional accuracy of the ultrasonic system in two dimensions was measured through the acquisition of 300 measurements at each intersection point of a 7 x 7 grid with divisions of size 100 mm x 100mm with and a resolution of 0.5 mm; finer dimensions than considered in [100].

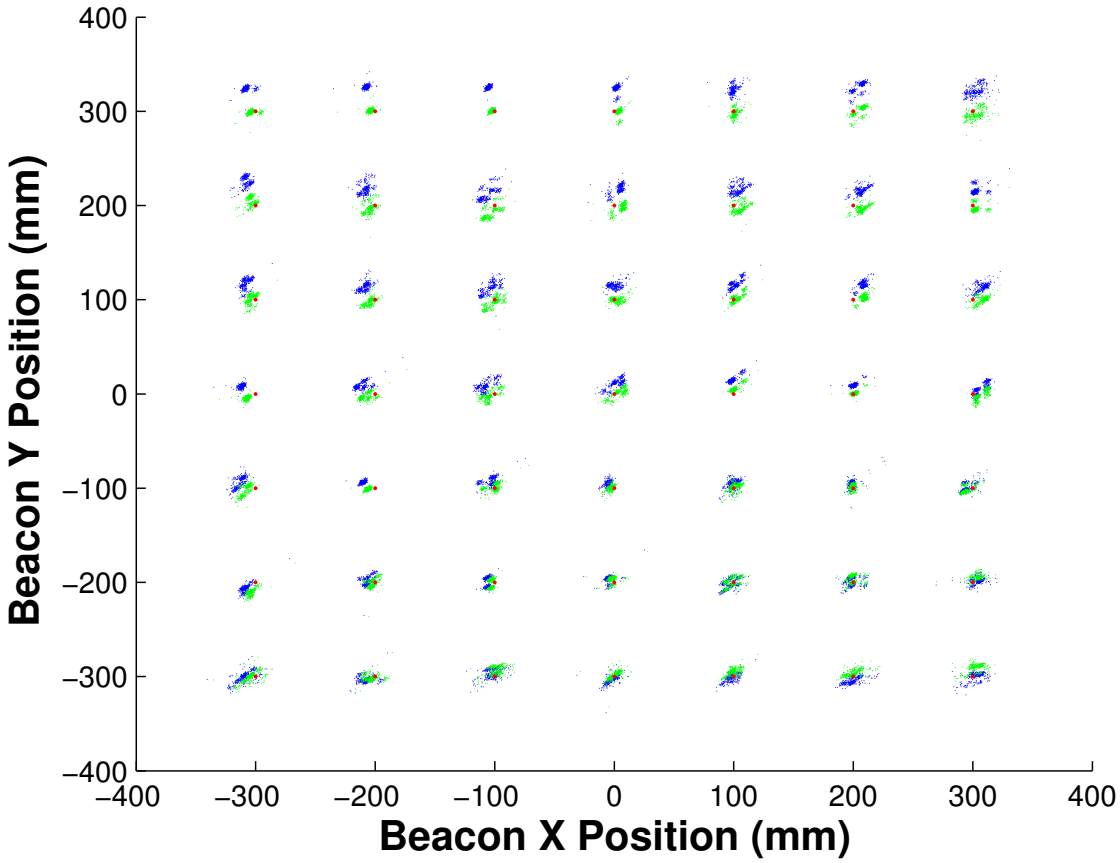


Figure 5.6: Uncertainty in ultrasonic measurements in the XY plane. The blue clusters represent raw measurements with the 1D correction applied, the green clusters result from the spatial correction and the red dot indicates the true location of the ultrasonic system module.

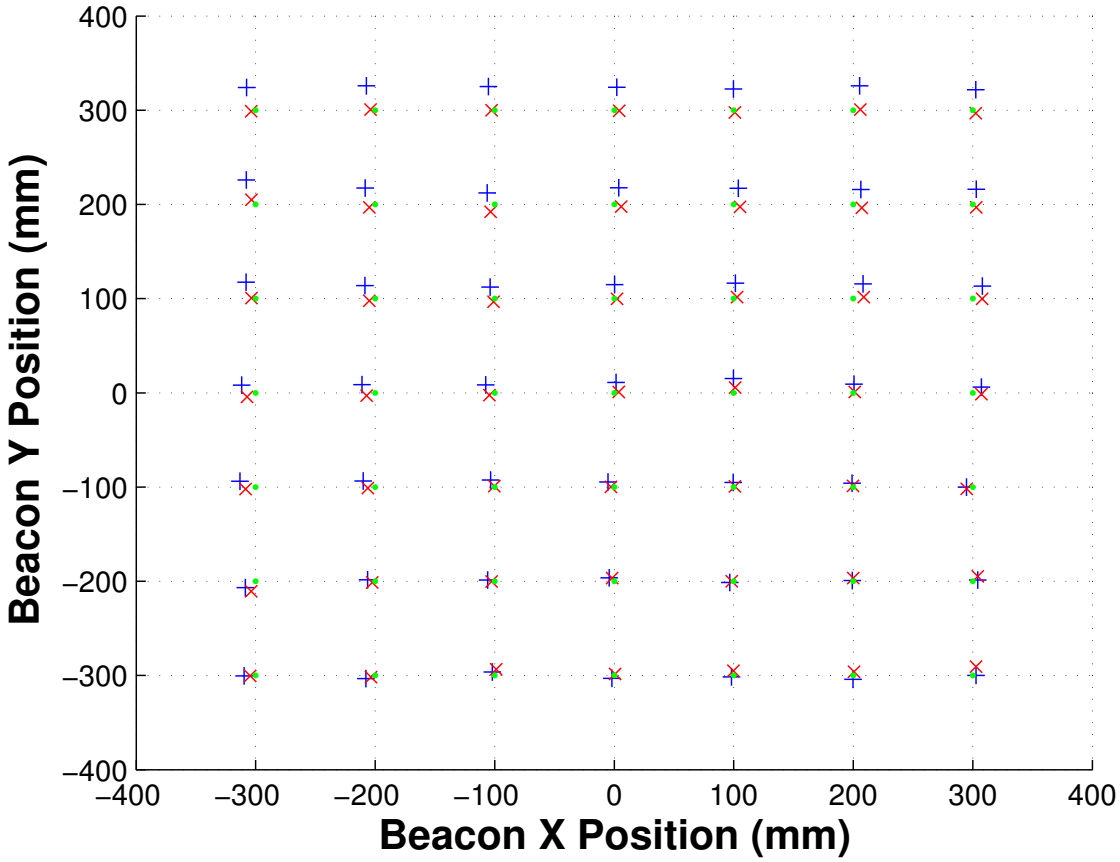


Figure 5.7: Plot of mean positions of the measurement clusters. The blue cross denotes raw ultrasonic data, the red cross indicates the calibrated location, while the green dot shows the true transmitter location.

The spatial distribution of raw ultrasonic measurements in the plane assumed the form shown in Figure 5.6; it can be seen that the uncertainty could be well approximated by a Gaussian distribution. Evaluating the mean position of each measurement cluster results in Figure 5.7 where it is apparent that the (x, y) estimates display a degree of distortion. A calibration procedure was applied to correct this distortion in order to simplify the measurement equations of the

filters described in the following sections. The calibration process consisted of measuring with Cricket known corner points of the rectangular 600 mm x 500 mm inspection area used in the experiments to form the following matrix:

$$A = \begin{pmatrix} BR_x & TL_x \\ BR_y & TL_y \end{pmatrix}^{-1}, \tag{5.6}$$

where (BR_x, BR_y), (TL_x, TL_y) were the bottom right and top left corners of the inspection area respectively. An ultrasonic measurement $\boldsymbol{p}_m = (x, y)$ then undergoes a transform by Equation 5.6:

$$P' = AP_m \tag{5.7}$$

yielding the point $P' = (x', y')$. The calibrated point P_{calib} is then given as follows:

$$P_{calib} = (\frac{x}{1 - (\frac{y}{y'})(1 - x')})(\frac{y}{1 - (\frac{x}{x'})(1 - y')}). \tag{5.8}$$

This calibration procedure was chosen for its simplicity and speed as only two points need to be recorded; it could subsequently be applied in online operation very easily. Using the mean point of each raw measurement cluster as input to the calibration process yielded the cross points in Figure 5.7. The displacement on a per grid point basis is illustrated in figure 5.8 for both the uncalibrated and calibrated cases. It is evident from these graphs that the spatial calibration has reduced the offset error to approximately zero.

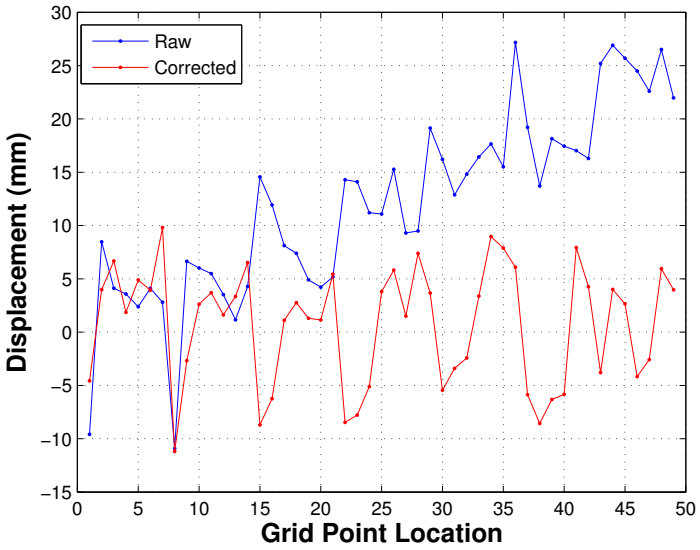


Figure 5.8: Error in the calibrated and uncalibrated cases

5.4 Ultrasonic Ranging Methods

This section reviews several ranging techniques comprising of threshold detection, cross correlation, phase measurement and the Biologically Inspired Ranging Algorithm.

5.4.1 Threshold detection

Threshold detection simply involves detecting the arrival of an ultrasonic signal by triggering the event when the receiver voltage exceeds a predefined threshold, ϵ. The equation is given by Equation 5.9. The signal can be a single-frequency amplitude-modulated signal, or take the form of an impulse function. In general, signals are tone-bursts - several cycles of a single frequency, often windowed to reduce the bandwidth of the signal. Threshold based systems are typically used in combination with narrow-bandwidth transducers to maximise the intensity

of the sound output and therefore range. The ultrasonic system uses threshold detection in order to detect the ultrasonic pulse used for the time difference of arrival calculation.

$$d(t) = \begin{cases} 1, & \text{if } s_T(t) \geq \epsilon \\ 0, & \text{if } s_T(t) < \epsilon \end{cases} \tag{5.9}$$

In general, the threshold technique does not require complex calculations or data storage, therefore, the required computational effort is low. It can be implemented with inexpensive single frequency ultrasonic transducers such as the 40 kHz transducers employed by the ultrasonic system. Regarding error the main sources consist of, low sampling frequency, low signal to noise ratio, and inherent bias due to the requirement of a threshold parameter. A low sample rate could introduce error due to the possibility of the sample points missing the true arrival time - using a high frequency increases the likelihood of capturing the true start point of the signal. In a low signal to noise ratio case, the noise could trigger the threshold causing error in the time estimate.

5.4.2 Cross-correlation

In ultrasonic time of flight systems, knowledge that the received signal will be temporally similar to the emitted signal means that a cross correlation between a transmitted signal and a received signal will result in substantial noise reduction. Cross correlation is also improved through signal design; a frequency modulated signal produces a sharper cross correlated peak than that obtained from a single frequency tone burst. For these reasons, cross-correlation has been called *matched filtering*, and when used in conjunction with frequency modulated signals, *pulse compression*. At its most simple, the method of cross correlation takes a trans-

mitted and received signal ,$s_T(t)$, and $s_R(t) \sim s_T(t - \tau_0)$) and produces a time domain signal whose maximum occurs at the delay time τ_0. The cross correlation, $c(t)$, is calculated for these real signals by:

$$c(t) = \int_{-\infty}^{\infty} s_T(\tau) s_R(t + \tau) d\tau \tag{5.10}$$

It is often simpler in practice to perform the calculation in the frequency domain, where $C(f) = S_T^*(f) S_R(f) = \mathcal{F}(s_T(t))^* \mathcal{F}(s_R(t))$, and $c(t)$ is the inverse Fourier transform of the cross spectrum $C(f)$. Cross correlation has noise-reduction properties, because the cross-correlation of random noise is theoretically zero. Application of cross correlation then has the potential to reduce the noise associated with the onset time of the received signal.

Cross correlation can be very accurate when the waveform is a frequency-modulated signal (FM), such as a linear chirp. The width of the pulse (ΔT) created when performing cross correlation is inversely proportional to the bandwidth (Δf) of the transmitted signal. FM chirps are widely used in conjunction with cross correlation for RADAR and SONAR applications. They require wideband transducers, and if appropriate wideband technology is available, then cross correlation using FM chirps is an optimal ranging method.

5.4.3 Phase Measurement

Measurement of the phase of the received signal, $s_R(t)$, can be used to determine range. This scheme can operate in a single frequency or multi-frequency configuration. The technique requires the entire time record of the signal, otherwise, offsets will be introduced causing the result to be incorrect.

5.4.3.1 Single frequency signals

Time of flight can be determined from measuring the phase of a single-frequency signal. A wave, $s(x,t)$, at a spatial position, x, and time, t, in a non-dispersive medium may be expressed as follows:

$$s(x,t) = Ae^{i(kx-2\pi ft)} = Ae^{i\left(2\pi f\left(\frac{x}{c}-t\right)\right)}, \tag{5.11}$$

where, A, denotes amplitude, k, is the wavenumber, f, is the frequency of the signal and c is the speed of sound in the medium. According to Equation 5.11 at any point x, the Fourier transform of $s(x,t)$ yields a spectrum where the phase ϕ of the frequency f is given by $\phi = 2\pi fx/c$. Since $x/c = t_F$, the phase is a direct measure of the time of flight. Phase is 2π periodic, therefore for any time of flight $t_F > 1/f$, there exists ambiguity. In the context of through air ranging, this ambiguity occurs even at very small distances, i.e. if $f = 40$ kHz and $c = 345\text{ms}^{-1}$, then ambiguity occurs for any distance between transmitter and receiver greater than ~ 8.6 mm. Practically, therefore, a single-frequency phase method cannot work for large-scale time of flight measurements in air, unless coupled with other methods. In the literature, this technique is often referred to as the phase shift method [99].

5.4.3.2 Multi frequency signals

A larger unambiguous range may be obtained through the use of two or more frequencies in the transmitted signal [99, 101]. In the single-frequency case, unambiguous operation over longer distances requires signal frequency to be reduced. For example, to measure over a range of 0-345 mm in air, a signal with a frequency ~ 1 kHz must be used. In the multi-frequency case, rather than using a 1 kHz single frequency signal, two signals separated in frequency by ~ 1 kHz, for example 40 kHz and 41 kHz, can be utilised. It can be shown that the *difference* in the phases of each signal is unambiguous for ~ 345 mm.

Given two frequencies f_1 and f_2, the time of flight t_F may be computed as a function of the two frequencies phases φ_2 and φ_1 as follows:

$$\begin{aligned} t_F &= \frac{1}{\Delta f}\left(\Delta n + \frac{\Delta\varphi}{2\pi}\right) \\ &= \frac{1}{f_2 - f_1}\left(n_2 - n_1 + \frac{\varphi_2 - \varphi_1}{2\pi}\right), \end{aligned} \tag{5.12}$$

where $\Delta n = n_2 - n_1$ is the difference in the number of cycles n_i accumulated by frequency f_i. Ambiguity occurs when $\Delta n \neq 0$; in this case when $t_F > \Delta f$.

5.4.4 Frequency modulated chirps and the cross spectrum

Wideband chirps provide an extension of the multi-frequency phase technique: a change from discrete frequency operation to continuous frequency representation. This is apparent by noting that Eq. 5.12 is a simple case of a more general equation that describes the behaviour of phase as a function of frequency and its relation to the time-of-flight of an acoustic pulse. In the limit of $\Delta f \to 0$, with $\Delta n = 0$, Equation. 5.12 becomes

$$t_F = \lim_{\Delta f \to 0}\left(\frac{1}{\Delta f}\frac{\Delta\varphi}{2\pi}\right) = \frac{1}{2\pi}\frac{\partial\varphi}{\partial f}, \tag{5.13}$$

which is a general result relating the time-of-flight $t_F = x/c$ to the gradient of relative phase plotted against frequency. Therefore, with a wideband signal, one only needs to know the relative phase between the two signals over a range of frequencies to obtain a value for the time of flight.

The simplest method to calculate the relative phase between two chirped signals is by calculating the cross spectrum. For a transmitted and received signals, $s_T(t)$ and $s_R(t)$ respectively, the cross spectrum is $C(f) = \mathcal{F}(c(t)) = S_T^*(f)S_R(f)$ where $c(t)$ is the cross correlation of $s_T(t)$ and $s_R(t)$, with $S_T(f)$ and $S_R(f)$ their Fourier

transforms. The phase of the cross spectrum is $Arg(C(f)) = Arg(S_T^*(f)S_R(f)) = Arg(S_R(f)) - Arg(S_T(f))$; that is, the phase of the cross spectrum is the difference in the phases between received and transmitted signal.

An FM chirp phase based measurement can therefore be performed as follows: The cross spectrum $C(f)$ between the transmitted and received signals is calculated. The relative phase $\Delta\varphi_C$ is obtained through $\Delta\varphi_C = \tan^{-1}(\Im(C(f))/\Re(C(f)))$. This relative phase must be unwrapped to reveal a linear relative phase as a function of frequency. At this point, a simple linear fit will give the value of $d\varphi/df$ to be used in Equation.5.13 to calculate the time of flight t_F of the acoustic signal.

5.4.5 Biologically Inspired Ranging Algorithm

Bats employ complex airborne ultrasonic chirp sequences enabling detection of sub-millimetre targets. They display an impressive ability to discriminate closely spaced overlapping echoes in the received signal - it has been shown experimentally [102] that bats are able to resolve separation delays smaller than the inverse bandwidth limit imposed by man-made systems based on coherent matched filtering receivers [103]. The *Eptesicus fuscus* bat can discriminate echoes in air with a resolution of 2 μs vs 10 μs of conventional systems assuming a signal bandwidth of 100 kHz. It is believed that the sonic emissions potentially composed of multiple spectral components and the associated signal processing of such signals enable this high resolution discrimination. The underlying assumption is that bats possess this ability to identify the acoustic signature of their prey [102].

The Bio-Inspired Range Finding Algorithm (BIRA) developed by Devaud *et al* [104], based upon the previous computational models by Matsuo *et al* [105] and Saillant *et al* [106], attempts to replicate, through time/frequency processing, the resolution with which bats are able to echo-locate. BIRA processes a received signal $x_{rec}(t)$ that is composed of a sum of overlapping Linear Frequency Modulated (LFM) chirps inspired by bat signals. This signal is processed temporally to extract time-of-flight while the *fine* delays arising from reflections, or glints, of

the complex reflecting object are determined through frequency-domain processing. It is the temporal processing that is the focus of this section, see [104] for details of the fine delay extraction algorithm.

5.4.5.1 Temporal Processing

The temporal processing component of BIRA seeks to mimic the operation of the basler membrane, a component of the bat inner ear. This membrane transforms received acoustic vibrations into electrical signals where this conversion is a function of varying frequency sensitivity along the length of the membrane. The temporal block makes use of a filterbank as a first order approximation of the basler membrane [106]. The nature of the sweep function in real bat chirps varies among species but in general assumes a hyperbolic shape in time. LFM chirps were chosen as an engineering simplification, however, the temporal subsystem can be reconfigured to process more realistic sweep functions. The received signal $x_{rec}(t)$ is convolved with a filterbank composed of 101 Gaussian chirplet filters characterised by the following equation [105]:

$$F(f_j, t) = \exp(\frac{-t^2}{\alpha}) \exp\left(2\pi i(f_j t + \frac{1}{2}st^2)\right) \tag{5.14}$$

where f_j is the starting frequency for the j^{th} chirplet filter, s is the sweep rate of the emitted wave and α is a parameter for the width of the Gaussian window function. The filterbank serves to decompose the received signal into a set of subbands each appearing in time as function of the linear frequency modulation, an effect similar to the action of the mammalian cochlea. The computational complexity of the algorithm is quadratic in the number of filters, N, and therefore the selection of 101 filters represents a compromise between an efficient discretization of the bandwidth and a manageable computational complexity, following the model of Matsuo [105].

The action of the temporal block can be viewed as an approximate parallel form of cross correlation wherein the signal is simultaneously convolved with a number of filters each of which processes a particular spectral band of the input. It is essentially carrying out a cross correlation of the received signal with a discretized version of the transmitted chirp - the action of the temporal block is illustrated in Figure 5.9.

The time of flight information is extracted through aligning the outputs of the filterbank with respect to the frequency which occurs first in time. The sum of the filter outputs across the time axis is then evaluated resulting in a single waveform the maximum peak of which is taken to be the time-of-flight of $x_{rec}(t)$. This process is illustrated in Figure 5.9(c) where a subset of filterbank outputs is shown and subsequently aligned and summed to produce the single waveform used to estimate time-of-flight.

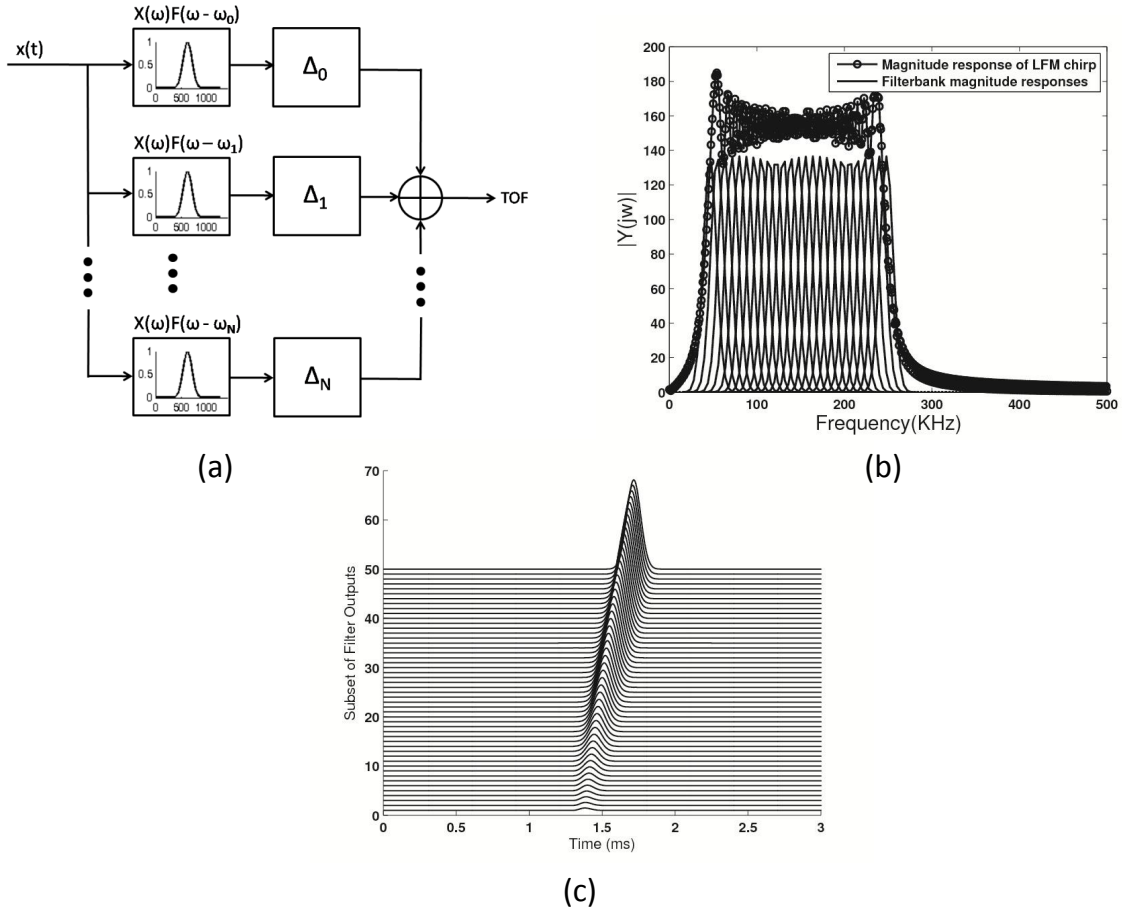


Figure 5.9: Operation of BIRA (a) The input signal $x_{rec}(t)$ is convolved with a filterbank composed of 101 Gaussian chirplet filters (b) The filters composing the filterbank process different frequency bands across the spectral occupancy of the input chirp (c) The chirp is converted into a series of pulses (subset shown). Note that the linear trend corresponds to the linear sweep rate of the chirp

5.5 Experimental Setup and Results

An experimental comparison of cross correlation, the cross spectrum technique and the BIRA was carried out using linearly swept chirps with frequency rising from 20 kHz to 120 kHz. The signal were transmitted from a wideband capacitive ultrasonic transducer to a calibrated microphone (1/8" Bruel and Kjaer, Naerum,

Denmark) placed at various distances on the acoustic axis of the transducer. The microphone was placed on a linear encoder with a positioning accuracy of ±250 μm, over a range of 10 to 90 cm (the limits of the linear stage). For each distance, 10 single-shot chirps were recorded (at 10 MS/s), and the distance calculated using the different algorithms.

A plot of real distance against accuracy for the three techniques is shown in Figure 5.10 while the standard deviation as a function of distance is shown in Figure 5.11. The methods have similar accuracies with the cross spectrum method being the most accurate. The standard deviation curves show the cross correlation and cross spectrum techniques to be very similar in repeatability while the BIRA display a comparatively lower level of repeatability.

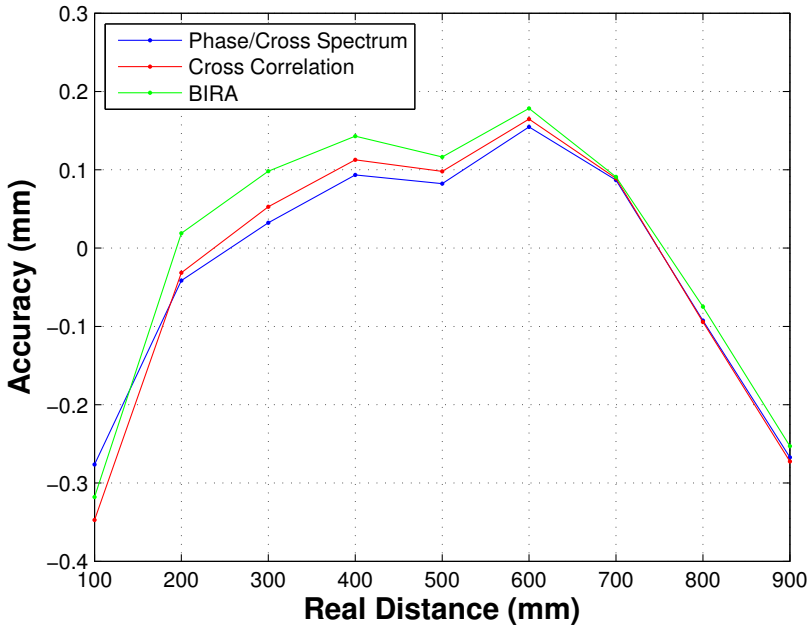


Figure 5.10: Accuracy plot for the three ranging techniques

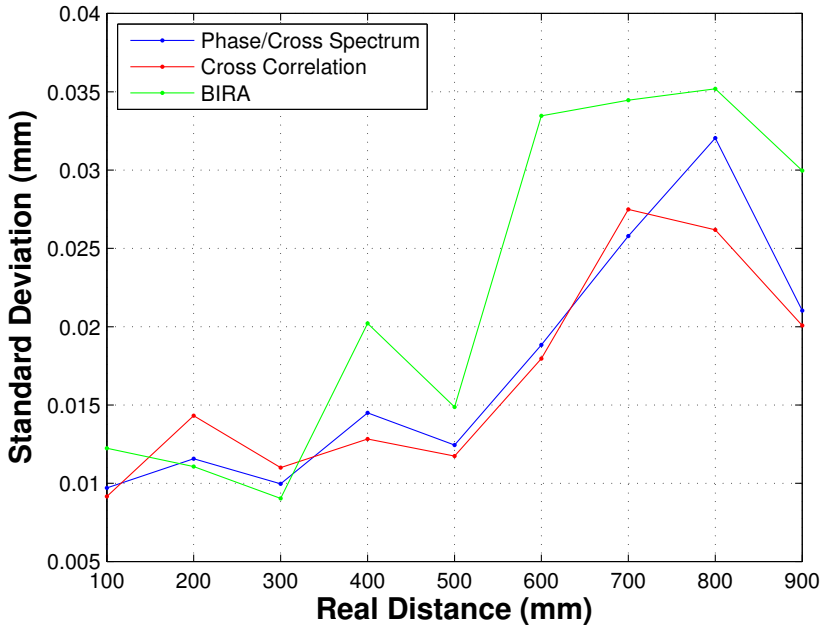


Figure 5.11: Standard deviation for the three ranging techniques

5.6 Conclusion

This chapter introduced the Cricket Indoor Location System, a low cost acoustic based positioning system. The accuracy of this system was characterised and calibration procedures were developed to correct range measurements and remove a radial type distortion observed in the 2D positional measurements. An overview and experimental comparison of a set of through air ranging techniques was also presented. This set comprised of the conventional threshold detection, cross correlation and phase based methods and the novel BIRA. The accuracy of these ranging techniques were tested upon signals acquired over a range of 10 to 90 cm. If was found that the methods produced very similar accuracy curves as shown in Figure 5.10. The minimum magnitude errors for the BIRA, cross correlation and phase based methods were 0.02 mm, 0.03 mm and 0.04 mm while

the largest magnitude errors were 0.31 mm, 0.35 mm and 0.29 mm respectively. However, the BIRA requires reater computational expense due to the filterbank in comparison to the other techniques.

In terms of repeatability as measured by standard deviation, the cross spectrum and cross correlation methods display similar results. The minimum minimum standard deviations were 0.008 mm and 0.01 mm for the cross spectrum and cross correlation respectively while the maximum deviations were 0.0275 mm, 0.0325 mm. However, the BIRA yields greater standard deviation for ranges greater than 300 mm, rising to a maximum of 0.035 mm.

The motivation for comparing different ranging methods stems from the desire to create a more accurate version of the ultrasonic system which as mentioned previously has the advantage of modularity and a small footprint in comparison to the other positioning systems considered in this thesis. An important factor in through air ultrasound is the dependency of sound speed on temperature, the effect of which can be substantial. Approaches for mitigating the temperature dependency were also discussed namely measurement of the ambient temperature and a calibration procedure make use of a known distance.

Chapter 6

Comparison of Bayesian Filters for Data Fusion

6.1 Introduction

All sensors have limited perception capabilities imposed by the physical mechanism employed to perform measurements. This means that they have different accuracies, uncertainties and failure modes. In mobile robot tracking, therefore, multiple positional sensors are utilised to continuously estimate vehicle pose. In such systems, the data resulting from each sensor is combined to compute a single estimate with the objective of having lower error than using each source in isolation. Since all measurements have an associated uncertainty their contribution to the overall estimate should be weighted accordingly. A probabilistic approach for sensor measurement fusion in which uncertainty is explicitly modelled has proven to be very effective in the robotics literature as exemplified by Thrun et al [107]. This approach casts the problem of merging multiple streams of noisy sensor data into the framework of *Bayesian filtering* which enables the joint estimate to be recursively updated by incoming measurements.

This chapter performs an experimental evaluation of two Bayesian filtering algorithms using multiple positional sensing data to track the planar motion of a vehicle. A real time, on-robot implementation of an Extended Kalman and Particle filter was used to control a robot performing representative raster scanning of a sample. Both absolute and relative positioning were employed - the absolute method being an indoor acoustic positioning system that required careful calibration. The performance of the tracking algorithms are compared in terms of computational cost and the accuracy of trajectory estimates.

6.2 Bayesian Filtering

Probabilistic estimation of a robot's position through fusion of multiple sensor outputs is a strongly researched area in robotics. It is a long-standing problem in the field and is considered a fundamental requisite of autonomous systems [17]. A typical component of a wheeled robotic system is odometry in the form of rotary encoders attached to the drive mechanism of the robot. These devices return pulses resulting from discrete increments of rotation thus providing a low-level source of positional information - the principle of operation was as illustrated in Chapter 1 section 2.3.2.2. Such sensors although providing accurate short-term tracking are subject to long term accumulation of errors introduced, in particular, by wheel slippage (driving on uneven terrain or slippery surfaces) and interaction with a priori unknown objects in the environment that may perturb the course of the robot [24]. These accumulated errors eventually lead to gross error between the true location and the encoder reported location. In relative positioning systems the uncertainty associated with the positional estimate can grow without bound as described in Chapter 2 2.3.2. The effect is illustrated in Figure 6.1 in which the ellipses which represent the contours (at the 3σ bounds) of Gaussian distrbutions over pose expand with distance. A simulation of a robot executing a raster scan is shown in Figure 6.2 in which its position is estimated by encoders. The blue line corresponds to the path reported by encoder odometry

while the green line corresponds to the actual motion of the robot that resulting from wheel slippage on the corners of the path. The error between the odometry reported path and actual path increases with trajectory length.

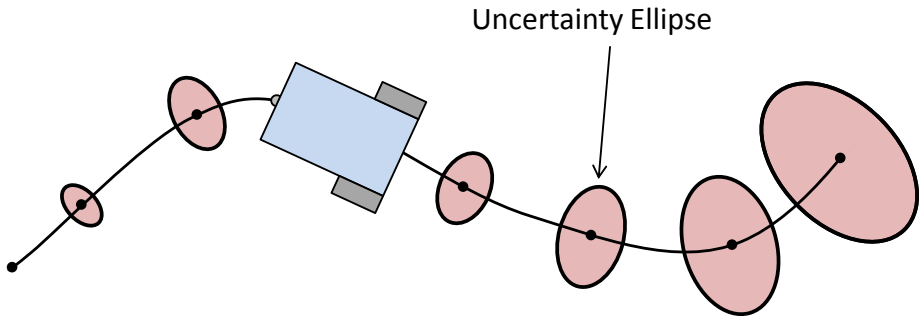


Figure 6.1: The ellipses representing positional uncertainty grow with distance travelled in encoder based positioning

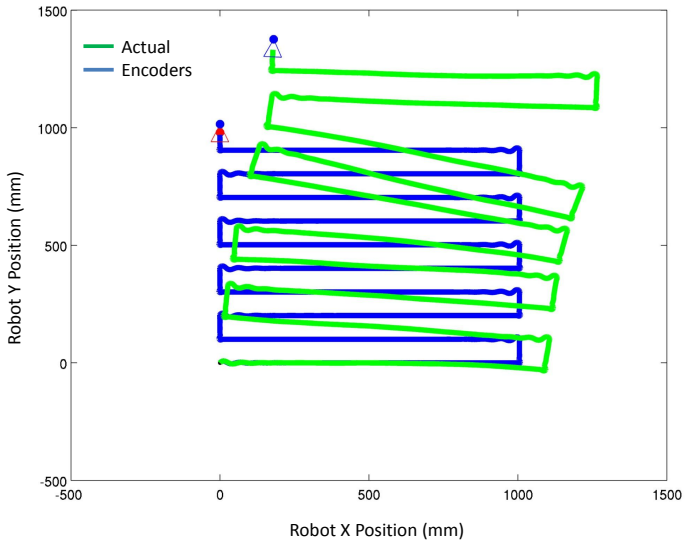


Figure 6.2: Typical raster path for NDE. Uncertainty in the robot's position grows with path length.

In the case of relative pose estimation, the build up in error of the estimate cannot be removed without the aid of some other form of sensing. This extra data can be provided by fixed beacons - forming an absolute positioning system - embedded in the operating environment to correct the path. It is also possible to use purely onboard sensors through application of SLAM as described in Chapter 2 2.3.5 wherein beacons are in effect automatically created from the environment itself. Both techniques although applicable in different cases make use of the same underlying filtering theory in order to combine sensor outputs.

6.2.1 Recursive Bayesian Filtering

The recursive Bayesian filter provides a probabilistic framework to fuse multiple noisy measurements to estimate a quantity, $\boldsymbol{x}_k$ (where k is the time increment), modelled as a Random Variable (RV). The quantity, $\boldsymbol{x}_k$, is termed the *state* and, in general, it assumes the form of a random vector whose elements comprise all variables of interest in the system: the specific problem under consideration is then known as *state estimation*. The objective of the filter is to combine the measurements to generate an estimated output, $\hat{\boldsymbol{x}}_k$, such that the estimation error $\boldsymbol{x}_k - \hat{\boldsymbol{x}}_k$ is minimised.

At each time instant a Probability Distribution Function (PDF) $p(\boldsymbol{x}_k)$ expresses the probability of the system being in each state comprising the state space which may be continuous or discrete. The PDF is in general multi-modal. Over time the system transitions from state to state and thus the shape of $p(\boldsymbol{x}_k)$ should be warped accordingly with time. This is the function of the Bayesian filter; it manipulates $p(\boldsymbol{x}_k)$ given the previous state and information provided by the set of incoming measurements, $\boldsymbol{Z}^k = \{\boldsymbol{z}_0, \ldots, \boldsymbol{z}_k\}$, to form the conditional density $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1}, \boldsymbol{Z}^k)$.

The operation of the Bayesian filter may be derived as follows (following the exposition in [108]). Let the PDF associated with the state $\boldsymbol{x}_{k-1}$ at time $k-1$ be denoted by $p(\boldsymbol{x}_{k-1}|\boldsymbol{Z}^{k-1})$. In the absence of any new measurements, $\boldsymbol{x}_{k-1}$ may be projected into the next time step using Equation 6.1 (using the theorem of total probability [109]):

$$p(\boldsymbol{x}_k|\boldsymbol{Z}^{k-1}) = \int p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})p(\boldsymbol{x}_{k-1}|\boldsymbol{Z}^{k-1})d\boldsymbol{x}_{k-1}, \tag{6.1}$$

where the term, $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$, corresponds to the PDF over the state following state transition between $k-1$ and k. The term $p(\boldsymbol{x}_k|\boldsymbol{Z}^{k-1})$ is referred as the prediction PDF and application of Equation 6.1 is known as filter *prediction*. This prediction is subsequently updated in light of a measurement $\boldsymbol{z}_k$ arriving at time k using Equation 6.2:

$$p(\boldsymbol{x}_k|\boldsymbol{Z}^{k}) = \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k)p(\boldsymbol{x}_k|\boldsymbol{Z}^{k-1})}{p(\boldsymbol{z}_k|\boldsymbol{Z}^{k-1})}, \tag{6.2}$$

where $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$ is termed the *likelihood* function, $p(\boldsymbol{x}_k|\boldsymbol{Z}_{k-1})$ from Equation 6.1 is now termed the *prior* PDF and $p(\boldsymbol{x}_k|\boldsymbol{Z}_k)$ is referred to as the *posterior* PDF: Equation 6.2 is know as the filter *update*. The denominator is a normalising constant that ensures that the posterior is a valid PDF. Note that any function that lies on or above the x-axis and integrates to 1 (sums to 1 in the discrete case) is a valid PDF [110]. In Equation 6.2 the prior can be considered to express the knowledge about the state before the arrival of a measurement. Upon arrival of $\boldsymbol{z}_k$, the likelihood function evaluates the $\boldsymbol{x}_k$ most likely to have given rise to the observed measurements. The posterior is then the PDF over the state integrating using both knowledge of how the state evolves in time *and* the observations which effectively refine the prediction through measurements.

Figure 6.3 illustrates the predict-update steps of the Bayesian filter. The prior, due to the uncertainty in state transition, has the effect of broadening the distribution over $p(\boldsymbol{x}_{k-1}|\boldsymbol{Z}^{k-1})$ and sliding it along the x-axis to the predicted state . Upon arrival of $\boldsymbol{z}_k$, the $\boldsymbol{x}_k$ most likely to have given rise to $\boldsymbol{z}_k$ is evaluated:

this is typically a narrower PDF that depends upon the sensor. The product (normalised) is taken to form the belief in $\boldsymbol{x}_k$ in light of $\boldsymbol{z}_k$; this is the posterior distribution.

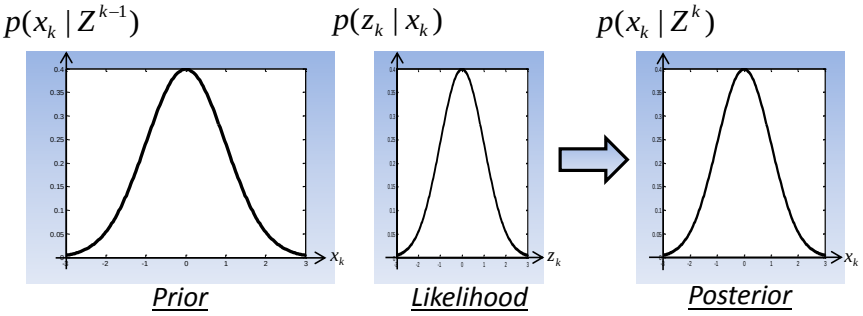


Figure 6.3: Depiction of predict-update cycle of the Bayesian filter assuming Gaussian PDF's

The practical use of Equations 6.1 and 6.2 is, in general, computationally intractable because the PDF, in real world systems, needs to be maintained over a potentially very large number of states. Consider, for example, a discrete case in which a 3-vector state may only take on values drawn from a fixed set where each variable in the vector can take on 100 different values. The Bayesian filter in this example would have to manipulate probabilities for 100^3 possible states; the values in this example are not unrealistic yet propagating a PDF comprising 10^6 entries would be very difficult to compute especially on an RSA. In the continuous case the filter would require to maintain the PDF over an infinite number of states. Practical implementation requires the introduction of constraints to make the manipulation of the PDF feasible.

6.2.2 Filter Implementation

Realisation of the predict-update cycle of the Bayesian filter requires *process* and *measurement* models; where the former pertains to calculation of the prior and the latter is involved in the evaluation of the likelihood. The process model

describes the state transition calculated in the predict step which consists of a deterministic function, $f_k(\boldsymbol{x}_{k-1})$ which is mostly likely to be nonlinear in real systems. This transition step is however uncertain and thus has an additive random noise component as follows:

$$\hat{\boldsymbol{x}}_k = f(\hat{\boldsymbol{x}}_{k-1}) + \boldsymbol{\epsilon}_k, \tag{6.3}$$

where $\boldsymbol{\epsilon}_k$ is random vector drawn from the process noise PDF; the expansion of the state PDF (prior) in Figure 6.3 is due to this noise. There is a also a requirement for a model describing how the system state maps to the measurement space in which the different sensors are measuring; this model is encoded in the (non-linear) function $h_k(\boldsymbol{x}_{k-1})$ while the noise perturbing incoming measurements is represented by the random vector $\boldsymbol{\delta}_k$.

$$\hat{\boldsymbol{z}}_k = h(\hat{\boldsymbol{x}}_{k-1}) + \boldsymbol{\delta}_k, \tag{6.4}$$

where $\boldsymbol{\delta}_k$ is drawn from the measurement noise PDF. Two quantities need to be taken care of in the filter namely the estimate of the state and probability distribution over all states. A diagram showing the sequence of steps that take place in the filter as a function of time is shown in Figure 6.4 (adapted from [111]).

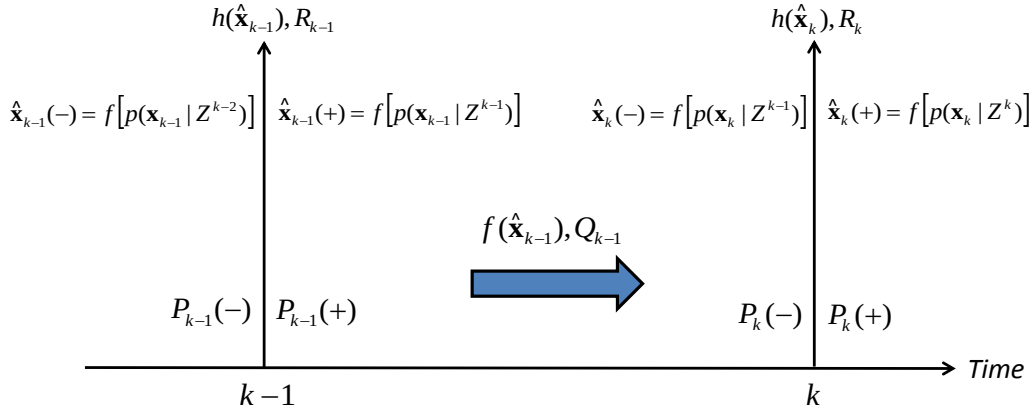


Figure 6.4: Filter Timing Diagram

Note that the $(-)$ and $(+)$ denote before and after integration of the measurement $\boldsymbol{z}_k$. At time step $k-1$ the estimate of the true state vector is denoted $\hat{\boldsymbol{x}}_{k-1}(-)$. Through application of the function, $h_k(\hat{\boldsymbol{x}}_{k-1})$, in combination with the measurement covariance R_k, the update step is executed to yield the updated state estimate $\hat{\boldsymbol{x}}_{k-1}(+)$. The function f defines the property of the PDF used as the estimate output by the filter. This may simply be the state coinciding with the maximum point of the PDF or it may be defined as the expected value, E of the distribution. The expectation operator is defined in the discrete case as follows:

$$E(\boldsymbol{x}) = \sum_{i}^{N} \boldsymbol{x}_i p(\boldsymbol{x}_i) \tag{6.5}$$

which is in effect a weighted average of the all possible states (of which there are N) that the estimated state can assume. The function, $f_k(\hat{\boldsymbol{x}}_{k-1})$, in combination with the process covariance Q_{k-1}, drives the state transition from $k-1$ to k at which point the process repeats. As well as updating the state estimate, the filters require to update the estimate of the uncertainty associated with the state. In Figure 6.4 the PDF is represented by the quantity P_k which is also subject to

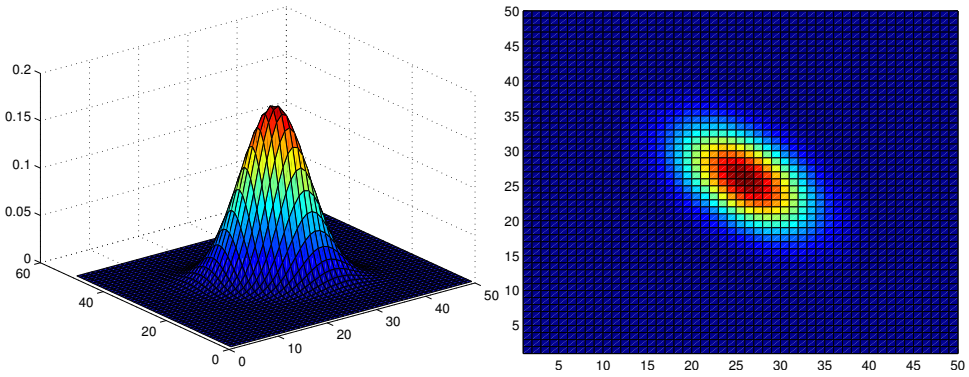
the predict-update cycle. The following sections describe the Extended Kalman Filter and Particle Filter which introduce constraints to make the implementation possible.

6.2.3 Extended Kalman Filter

The Kalman Filter (KF) [112] is an analytic approximation of the Bayesian filter that constrains the PDF to have a fixed form. The derivation of the KF makes the assumption that the distribution over the state space at any time instant is *Gaussian*. The unimodal property of this distribution implies that the system state is mostly likely to reside in one location of the state space. A multivariate Gaussian PDF over the RV, $\boldsymbol{x} \in \Re^N$, is defined as follows:

$$\boldsymbol{g}(\boldsymbol{x}) = \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})), \tag{6.6}$$

where $\boldsymbol{\mu} \in \Re^N$, is the mean and, $\Sigma \in \Re^{N\times N}$, is the covariance of the distribution, note that, $||$, denotes matrix determinant; a plot for the $N = 2$ case is shown in Figure 6.5(a).



(a) (b)

Figure 6.5: Gaussian assumption of EKF (a) Multivariate Gaussian PDF over $\boldsymbol{x} \in \Re^2$ (b) Plot of contour at 3σ bounds

The variable $\boldsymbol{\mu}$ (which coincides with the mode and median of the distribution) is defined to be the output of the filter and may be extracted by the expectation operator as follows:

$$\hat{\boldsymbol{x}} = E(\boldsymbol{x}). \tag{6.7}$$

The covariance term is computed from the expectation of the estimation error $\tilde{\boldsymbol{x}} = \boldsymbol{x} - \hat{\boldsymbol{x}}$ as follows:

$$\Sigma = E((\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}})^T). \tag{6.8}$$

The covariance matrix is a symmetric, positive semi-definite matrix that encodes the shape of the contour of the PDF as shown in 6.5(b). It is a multi-dimensional measure of the spread of the distribution, in the 1D case covariance becomes variance. The diagonal entries of Equation 6.8 are the variances in the corresponding elements of $\boldsymbol{x}$. The off-diagonal terms are a measure of the extent to which the elements of the state vary together. The covariance matrix can also be viewed to be the MSE of the estimate with respect to the true state. The task of the KF then, is to propagate the mean and covariance of the state PDF over time which is equivalent to propagating the state estimate and associated estimate uncertainty over time.

6.2.3.1 Gaussian Assumption

The underlying assumption of the KF forces the random vectors in Equations 6.3 and 6.4 to be drawn from zero mean Gaussian PDFs. This condition is expressed as follows:

$$E(\boldsymbol{\epsilon}_k) = \mathbf{0}, \qquad E(\boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^T) = Q_k, \tag{6.9}$$

$$E(\boldsymbol{\delta}_k) = \mathbf{0}, \qquad E(\boldsymbol{\delta}_k \boldsymbol{\delta}_k^T) = R_k, \tag{6.10}$$

where Q_k and R_k are covariances of the process and measurement noise respectively.

Because a linear transform of a Gaussian RV results in another Gaussian RV, the derivation of the KF relies upon the functions in Equations 6.3 and 6.4 being linear. If these functions are non-linear the Gaussian assumption no longer holds and the estimate will diverge. There exist several extensions to the KF for non-linear systems of which the Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) [107] are the mostly commonly used. The EKF simply linearises Equations 6.3 and 6.4 around the current state estimate and then proceeds to make use of the standard KF equations. The UKF selects several samples of the state space surrounding the current estimate, passes these points through the non-linear transform and then constructs a Gaussian PDF from the resultant set of points. In practical applications the performance difference between the EKF and UKF has been found to be negligible [107] with the latter also having greater computational overhead. In this study the EKF was selected for the experimental comparison with the Particle Filter.

6.2.3.2 Prediction

The EKF linearises the process and measurement models through truncating terms with order greater than or equal to 2 in the multidimensional Taylor expansion around the predicted state estimate. The coefficients of the complete expansion progressively reduce in magnitude and thus the lower order terms dominate in contribution [113]. It is possible then, to maintain an accurate approximation through retaining only the terms defining the linear approximation. The accuracy of the approximation is ultimately a function of the local nonlinearity

surrounding the region in which the function linearisation is being evaluated. In the case of highly nonlinear functions, the EKF could be a poor state estimator. In the EKF, filter prediction is implemented by Equation 6.11:

$$\hat{\boldsymbol{x}}_k = f(\boldsymbol{x}_{k-1}), \tag{6.11}$$

which corresponds to the expected value of Equation 6.3. The predicted state uncertainty is computed through first order propagation of the uncertainty from the previous time step. Consider the 1D case, where a function, g, depends upon an RV x. The variance in $g(x)$ resulting from the variance in x is, to first order:

$$(\frac{dg}{dx})\sigma_x^2(\frac{dg}{dx}). \tag{6.12}$$

Equation 6.13 shares this structure in expressing the predicted covariance as follows:

$$\hat{\Sigma}_k = F_k \Sigma_{k-1} F_k^T + R_k, \tag{6.13}$$

where $F_k = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_{k-1}}$ is the Jacobian of the function $\boldsymbol{f}$ with respect to the state from the previous time step $\boldsymbol{x}_{k-1}$. The covariance of the process noise R_k is also added at this stage which signifies that the prediction induces uncertainty into the estimate of the state - this is shown by the prior in Figure 6.3.

6.2.3.3 Update

The Kalman gain matrix, K_k, can be viewed as minimising he sum of the diagonal entries of $\hat{\Sigma}_k$ [111] and is given as follows:

$$K_k = \hat{\Sigma}_k H_k^T (H_k \hat{\Sigma}_k H_k^T + Q)^{-1}, \tag{6.14}$$

where $H_k = \frac{\partial h}{\partial \boldsymbol{x}_{k-1}}$ is the measurement Jacobian. The gain matrix is used to weight the contribution of the information provided by the sensor measurement to the state estimate as follows:

$$\boldsymbol{x}_k = \hat{\boldsymbol{x}}_k + K_k(z_k - h(\hat{\boldsymbol{x}}_k)). \tag{6.15}$$

The covariance matrix in light of the measurement is computed by:

$$\Sigma_k = (I - K_k H_k)\hat{\Sigma}_k, \tag{6.16}$$

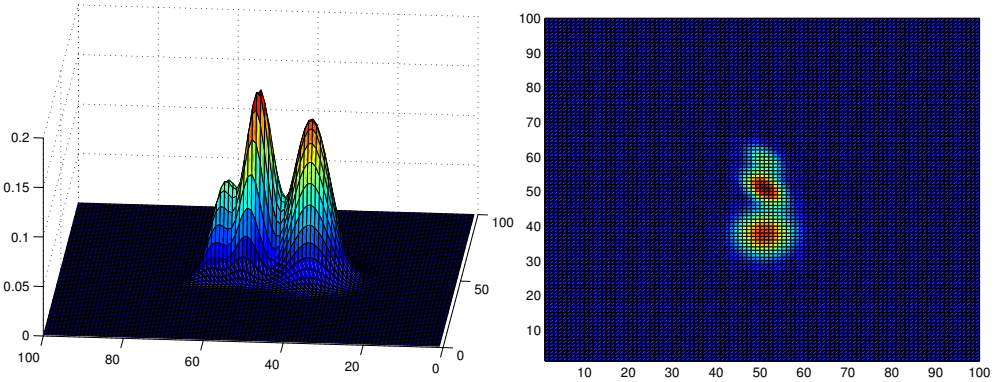
where I is the identity matrix with dimension N x N. The addition of the measurement information has the effect of reducing the covariance of the distribution. Equations 6.11 - 6.16 implement the EKF. For fixed size of state vector, these equations can be computed within a fixed time period.

6.2.4 Particle Filter

A Particle Filter (PF) is a recursive Monte Carlo technique that uses a sample based representation of the probability distribution associated with the system state. The constraint introduced by the PF is one of predicting and updating only a *finite* set of samples in the state space rather than manipulating the set of *all* possible states. The PF does not enforce the Gaussian assumption made by the EKF. This means that both the Gaussian noise *and* function linearity constraints to be relaxed [108]. A simple example demonstrates the differences between two the PF and EKF.

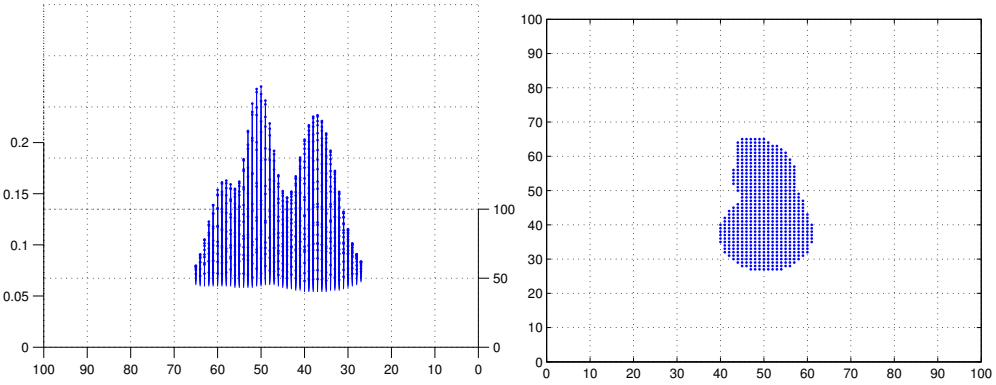
The multi-modal PDF shown in Figure 6.6(a) displays a distribution that could arise from the state passing through a non-linear transform. With reference to Figure 6.6(b), it is apparent that there exists two states with relatively high probability: an accurate representation of such a PDF in the EKF would not be

possible. A PF representation of this distribution in terms of a Probability Mass Function (PMF) (a discrete distribution) is shown in Figures 6.7(a) and (b) in which a finite set of samples are used to capture the shape of the distribution.



(a) (b)

Figure 6.6: (a) Example of a multi-modal PDF that could result from model non-linearity or non-Gaussian noise sources (b) Corresponding non-elliptical contour plot of PDF showing 2 areas of large probability



(a) (b)

Figure 6.7: (a) PF sample based representation (b) Sample in the XY plane of the distribution

The distribution associated with the state is represented as a discrete set of *state*, *weight* pairs as follows:

$$\{\boldsymbol{x}_k^i, w_k^i\}_{i=1}^N, \tag{6.17}$$

where N is the number of samples taken to approximate the true distribution, as $N \to \infty$ the better the approximation becomes and greater the computation is required. The term, $\boldsymbol{x}_k^i$ corresponds to the *ith* sample while w_k^i is its weight which is proportional to the probability of being the true state of the system. Since the weights represent probability they sum to unity as follows:

$$\sum_i w_k^i = 1. \tag{6.18}$$

The pair, $\{\boldsymbol{x}_k^i, w_k^i\}$, is referred to as a *particle*. The distribution over the state is then represented as a weighted shifted sum of impulse functions [114] as follows:

$$p(x_k|\boldsymbol{Z}^k) \approx \sum_{i=1}^N w_k^i \delta(\boldsymbol{x}_k - \boldsymbol{x}_k^i), \tag{6.19}$$

where $\boldsymbol{Z}^k$ is the set of all measurements received up until time k. The output of the PF can be extracted in a number of ways from the underlying PMF. This is usually taken to be the state coinciding with maximum probability or the expected value of distribution calculated by:

$$\hat{\boldsymbol{x}}_k = \sum_{i=1}^N \boldsymbol{x}_k^i w_k^i. \tag{6.20}$$

The task of the PF is then to predict and update the samples and weights representing the state PDF over time.

6.2.4.1 Importance Sampling

A central problem in the PF concerns the method used to draw samples from the posterior distribution, $p(\boldsymbol{x}_k|\boldsymbol{Z}^k)$, to propagate through the filter over time. In general an analytic expression for the posterior is not readily available and it has been shown that for high-dimensional state vectors the associated distribution is concentrated in a small region of the state space [115]. Uniformly sampling from this space, therefore, is very likely to result in many samples representing improbable states.

To mitigate this problem the concept of *importance sampling* is used [108]. Instead of sampling from the posterior distribution, importance sampling samples from a different distribution that shares the *same* domain as the posterior; this distribution is referred to as the *importance* or *proposal* density. The key to the method is that is easier to sample from the importance density than it is from the posterior. The conventional choice of proposal is the transitional distribution $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$ [108].

6.2.4.2 Prediction

The prediction step of the PF is implemented through drawing samples from $p(\boldsymbol{x}_k|\boldsymbol{x}_{k-1})$. This corresponds to sampling N times from Equation 6.3 where in the framework of the PF, $\boldsymbol{\epsilon}_k$ can be a non-Gaussian PDF and the (non-linear) function f is used directly.

6.2.4.3 Update

Given the choice of the transitional density as the proposal, the resultant update step of the PF in which a measurement $\boldsymbol{z}_k$ is integrated can be shown to have the following form [114]:

$$w_k^i = w_{k-1}^i \frac{p(\boldsymbol{z}_k|\boldsymbol{x}_k^i)}{\sum_{i=1}^N w_k^i} \tag{6.21}$$

where the likelihood, $p(\boldsymbol{z}_k|\boldsymbol{x}_k^i)$, is implemented through sampling from the (non-linear) measurement model in Equation 6.4 where the noise vector, $\boldsymbol{\delta}_k$, can be drawn from a potentially non-Gaussian PDF.

6.2.4.4 Resampling

After a small number of cycles of the filter a problem known as *degeneracy* can arise - this is due to the variance of particle weights increasing with time [114]. Degeneracy is the condition in which all but one particle has negligible weight. Practically, this means that the filter is updating particles which have negligible contribution to the filter output. Resampling is a technique used to combat degeneracy which involves discarding particles with low weight and replicating those particles with higher weights due to their higher probability of being the true state. The number of effective particles, N_{EFF}, can be used as a measure of the degeneracy and may be computed as follows:

$$N_{EFF} = \frac{1}{\sum_{i=1}^N (w_k^i)^2}. \tag{6.22}$$

A resampling procedure can then be invoked every time N_{EFF} falls below a threshold, N_{THR}.

The objective of the resampling procedure is to generate a new set of particles, $\{\boldsymbol{x}_k^{i*}, w_k^{i*}\}_{i=1}^N$, from the old set, $\{\boldsymbol{x}_k^i, w_k^i\}_{i=1}^N$, such that $Pr(\boldsymbol{x}_k^{i*} = \boldsymbol{x}_k^j) = w_k^j$ where the new weights are set to $\frac{1}{N}$. Resampling with replacement is used, therefore, and particles with high weight can be selected multiple times during the procedure. There exist numerous algorithms for resampling including systematic, stratified and residual sampling.

6.2.4.5 Sample Impoverishment

The possibility for multiple copies of the same particle in the resampled set can lead to a loss of diversity in the set of samples. This condition is mitigated in part through the process noise which adds a random offset to the samples such that they are spread in the state space; in practice this may not be enough. A technique known as *roughening* can be used which artificially increases the process noise in order to further spread the samples in the state space [116].

6.2.5 Filter Tuning

6.2.5.1 Filter Consistency

A filter is said to be *consistent* if it produces state estimates which converge to the true system state as the number of measurements increases. Consistency implies that the estimate is *unbiased* meaning that the filter output is as close as possible to the true system mean [113]:

$$\hat{\boldsymbol{x}}_k \approx E[\boldsymbol{x}_k]. \tag{6.23}$$

It also requires that the difference between the filter estimated covariance and the true covariance is minimised, this is known as *efficiency* and may be expressed as follows:

$$\Sigma_k - E\big[(\boldsymbol{x} - \hat{\boldsymbol{x}})(\boldsymbol{x} - \hat{\boldsymbol{x}}))^T\big] \geq \mathbf{0}. \tag{6.24}$$

The approximations employed by the EKF and PF can result in estimates which are *inconsistent* thus violating Equations 6.23 and 6.24. Inconsistency can lead to the filter diverging from the true state.

Inconsistency in the EKF results from the linearisation of the process and measurement models. Julier and Ulhmann [113] demonstrate EKF inconsistency through a simple example in which a range, bearing measurement with respective standard deviations of $\sigma_r = 2$ cm and $\sigma_\theta = 15^\circ$ is converted to Cartesian form. The covariance resulting from the linearisation is compared to a Monte Carlo estimate of covariance which uses 3.5×10^6 samples. It is found that the mean of the linearised estimate is 96.7 cm while the true value is $1m$ and the ellipse representing the linearised covariance is smaller than the true value. This shows the problem of bias in the mean and an over optimistic MSE associated with the EKF. The authors indicate that the issue of bias cannot be removed although the covariance under-estimation can be mitigated through the introduction of *stabilisation* noise which artificially inflates the acekf covariance estimate such that Equation 6.24 is satisfied.

Heijden [117] outlines the reasons for inconsistency in PFs. The first reason concerns the number of particles employed: the number of samples used may be too small to represent the state PDF with sufficient statistical significance. The second reason concerns the complexity of the process and measurement models. For detailed models containing many variables in the state vector it is difficult to generate a representative set of samples for the reasons mentioned in Section 6.2.4.1 - the problem of an accurate representation is then problematic. An addi-

tional problem of using detailed models is that variation in the model parameters can have a large effect upon the output of the model. In the case of using a model with too few parameters the filter estimate is likely to contain bias.

6.2.5.2 Process and Measurement Covariances

The accuracy of a state estimator is a function of the accuracy with which the process and measurement covariances have been characterised [118]. The matter is complicated by the previously mentioned issues of EKF stabilisation noise and roughening in the PF. Measured values of covariance obtained through experimentation then do not necessarily correspond with those needed to achieve optimal filter performance.

6.2.6 Literature Review

Tong and Barfoot [119] carried out a comparison of an EKF and sigma point Kalman filters for fusing encoder and LIDAR measurements acquired from a 4-wheeled skid steer vehicle. The performance metric comprises root mean square error with respect to ground truth provided by a Vicon motion capture system. A PF using a very large number of particles is used to simulate the theoretical best performance of the Bayes Filter. The PF outperforms both filters while the sigma point variant of the Kalman filter was found to be more robust to noise and provided higher overall accuracy in comparison to the EKF.

Peralta-Cabezas [120] et al provide a comparison of ten filtering techniques comprising of KF based filters, Monte Carlo filters and hybrid filters which are formed from combining elements of the preceding two approaches. Filter performance is compared for the problem of estimating the location of a three-wheel omnidirectional vehicle using data from a video camera. The main performance criteria are the size of the estimation error, the required computational effort and the

robustness of these techniques to non-Gaussian noise. The results obtained indicate that, generally, the EKF and PF yield accurate predictions with the latter requiring significant computational effort.

Bellotto and Hu [121] examined the use of PF and Kalman filter based techniques for tracking people using a camera and laser range finder mounted on a robot. The authors showed that the Unscented Kalman filter performs approximately to the level of the PF at less computational cost.

6.3 Process Model

The location of a robot in 2D space, defined in Figure 6.8, is determined by 3 variables: the (x, y) position of the drive axis midpoint and θ the angle of rotation with respect to the world coordinate frame.

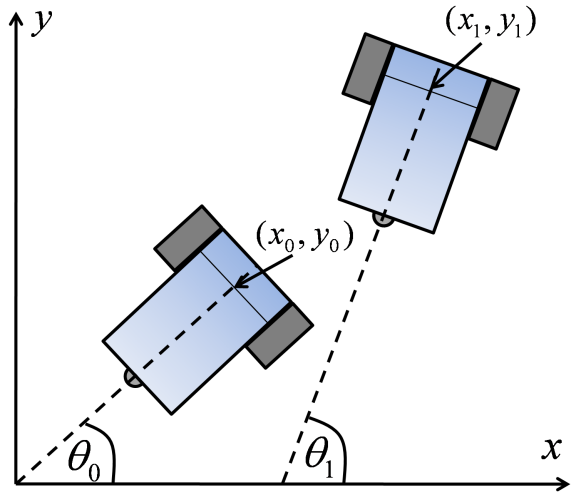


Figure 6.8: Definition of robot model parameters. The (x, y) location of the robot (drive axis midpoint) in the plane of motion and the angle of the centre with respect to the x-axis define the pose of the robot. Note that positive angles result from counterclockwise measurement relative to the x-axis

The state vector, $\boldsymbol{x} \in \Re^3$, to be estimated by the EKF and PF is then defined to be the pose of the robot as follows:

$$\boldsymbol{x} = [x, y, \theta]. \tag{6.25}$$

The error covariance matrix, $\Sigma \in \Re^{3\times3}$, associated with the state estimate $\hat{\boldsymbol{x}}$ has the following form:

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_x\sigma_y & \sigma_x\sigma_\theta \\ \sigma_y\sigma_x & \sigma_y^2 & \sigma_y\sigma_\theta \\ \sigma_\theta\sigma_x & \sigma_\theta\sigma_y & \sigma_\theta^2 \end{bmatrix}, \tag{6.26}$$

where the diagonal entries express the variance in the corresponding state elements and the off-diagonal terms express the extent to which the state elements vary together; note the covariance may be viewed as a correlation matrix as shown in Section 2.3.6. The state transition function, $\boldsymbol{f}$, from Equation 6.3 models the transformation of pose between successive time steps.

State transition is described by the robot kinematic model, $\boldsymbol{f}(\boldsymbol{x}_{k-1}, \boldsymbol{u}_k)$, that is driven between states by the control vector $\boldsymbol{u}_{k-1}$ comprising of encoder pulse counts, Δ_r and Δ_l, accrued by the left and right wheels respectively between sampling times. The differential drive robot shown in Figure 6.9 was used during the comparison experiments; the encoders are directly coupled to the drive wheels to provide a onboard relative source of position data.

Counting the number of pulses accrued between sampling time instants enables prediction pose change which can then be used to update the absolute pose estimate through the following set of equations [24] for a differential drive robot:

$$\boldsymbol{x_k} = \boldsymbol{x_{k-1}} + \begin{bmatrix} r\cos(\theta_{k-1} + \Delta\theta) \\ r\sin(\theta_{k-1} + \Delta\theta) \\ \Delta\theta \end{bmatrix}, \tag{6.27}$$

where $\boldsymbol{x}_{k-1}$ is the pose of the robot at the previous time step:

$$\Delta\theta = \frac{c(\Delta_r - \Delta_l)}{b}, \tag{6.28}$$

and

$$r = \frac{c(\Delta_r + \Delta_l)}{2}, \tag{6.29}$$

are the change in angle and the arc length traversed by the wheels between time steps k and $k-1$, c is the conversion factor between pulses and linear displacement and b is wheelbase of the robot. As with any relative system there exists a cumulative build of error as illustrated in Figure 6.2; the error in the estimate grows with distance and can lead to gross error between the actual pose and the encoder reported pose. The estimate, derived from the encoders, is susceptible to wheel slip and any unmodeled influences that could perturb the course of the vehicle.

Unmodeled effects that could influence the state vector are collected into the random noise vector $\boldsymbol{\epsilon}$ (assumed to be independent of time, hence the subscript k is dropped). It was assumed that the vector was drawn from a Gaussian distribution.

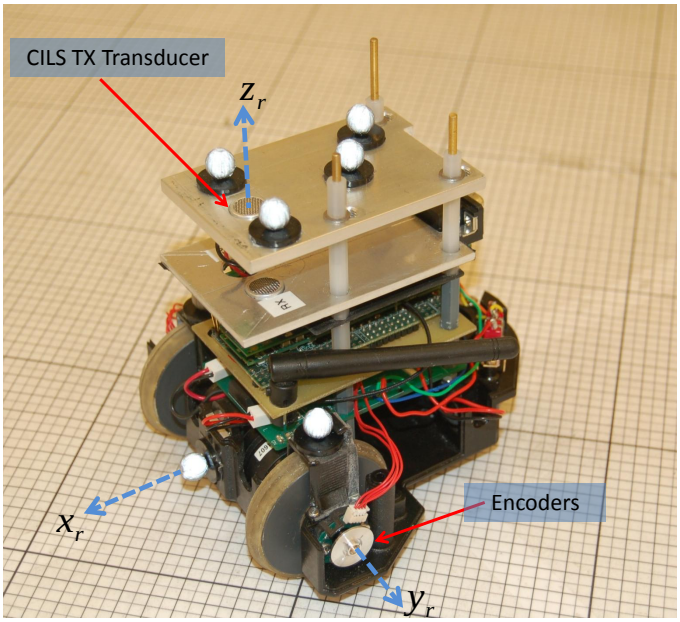


Figure 6.9: Differential drive RSA used in the experiment. The transmitting transducer was precisely aligned with the midpoint of the wheel base to coincide with the odometry tracked point.

6.3.1 Implementation

In the EKF the Jacobian matrix is implemented as follows:

$$F_k = \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}_{k-1}} = \begin{bmatrix} 1 & 0 & r\cos(\theta_{k-1} + \Delta\theta) - r\cos(\theta_{k-1}) \\ 0 & 1 & r\sin(\theta_{k-1} + \Delta\theta) - r\sin(\theta_{k-1}) \\ 0 & 0 & 1 \end{bmatrix}, \tag{6.30}$$

The prediction stage of the PF involves drawing N samples from:

$$\boldsymbol{x}_k = \boldsymbol{f}(\boldsymbol{x}_{k-1}, \boldsymbol{u}_k) + \boldsymbol{\epsilon}. \tag{6.31}$$

6.4 Measurement Model - Acoustic Beacon Location System

The measurement prediction of Equation 6.4 is implemented by Equations 5.4 repeated here for convenience:

$$\begin{bmatrix} x_{tx} \\ y_{tx} \\ z_{tx} \end{bmatrix} = \begin{bmatrix} \frac{1}{2r}(d_1^2 - d_2^2 + r^2) \\ \frac{1}{2r}(d_1^2 - d_3^2 + r^2) \\ \pm\sqrt{(d_1^2 - x^2 - y^2)} \end{bmatrix} \tag{6.32}$$

6.4.1 Implementation

In the EKF, the measurement Jacobian $H_k \in \Re^{3\times 3}$ is given as follows:

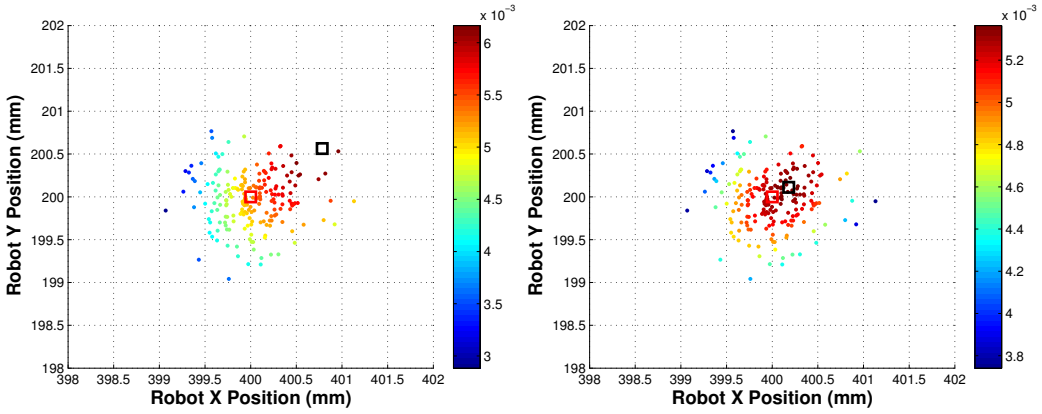
$$H_k = \frac{\partial h}{\partial \boldsymbol{x}_{k-1}} = \begin{bmatrix} (x_1 - \bar{x}_{k-1})/d_1 & (y_1 - \bar{y}_{k-1})/d_1 & 0 \\ (x_2 - \bar{x}_{k-1})/d_2 & (y_2 - \bar{y}_{k-1})/d_2 & 0 \\ (x_3 - \bar{x}_{k-1})/d_3 & (y_3 - \bar{y}_{k-1})/d_3 & 0 \end{bmatrix}, \tag{6.33}$$

where $d_b = \sqrt{(\bar{x} - x_b)^2 + (\bar{y} - y_b)^2}$ for beacons $b \in \{1, 2, 3\}$.

In the PF the incorporation of measurement information is implemented through the likelihood function, $p(\boldsymbol{z}_k|\boldsymbol{x}_k)$. In section 5.3.1.1, it was shown through experimentation that the noise corrupting distance measurements could be modelled as Gaussian with a variance of $\sigma_{beacon} = 23$ mm^2. Assuming that the distance measurements are Independent and Identically Distributed (IID), the joint likelihood for each particle, $\boldsymbol{x}_k^i$, becomes a product of the marginals distributions for each component of, $\boldsymbol{z}^k$, as follows:

$$p(\boldsymbol{z}_k|\boldsymbol{x}_k^i) = \prod_{b=1}^{3} \frac{1}{\sqrt{2\pi\sigma_{beacon}^2}} exp\left(\frac{(\sqrt{(x_i - x_b)^2 + (y_i - y_b)^2 + z_b^2} - d_b)^2}{-2\sigma_{beacon}^2}\right), \tag{6.34}$$

where (x_i, y_i) is the predicted location of the robot by the ith particle. The likelihood function in effect assigns more weight to those particles closer to an incoming beacon measurement, where the size of the weight is a function of the variance of the distance measurements. The weight assignment function of the Equation 6.34 is demonstrated in Figure 6.10. In Figure 6.10 (a) the subset of particles in the upper right area of the arrangement receive greater weighting in accordance with their shorter distance to the ultrasonic reading. In Figure 6.10 (b) the ultrasonic measurement lies close to the predicted location of the robot and thus the centre of the particle cloud is assigned relatively high weighting.



(a) (b)

Figure 6.10: Example of particle weight assignment by the likelihood function. The black square indicate the location of the ultrasound reported robot location, while the red square shows the mean of the particle positions. The colour of the particles indicate weight according to the colourbar. (a) Ultrasonic reading far from the predicted location (b) Ultrasonic reading close to the predicted location

6.5 Experimental Evaluation of Bayesian Filters

Efficient C++ implementations of the EKF and PF were written for execution on-board the RSA. To comply with best practice, the UMBMark [97] procedure was carried out to fine tune the wheel base and wheel diameters to ensure optimal encoder odometry estimation. Ground truth was provided by the VMCS. The

following sections describe the method used to align the coordinate frames of the different tracking systems used in the experiment followed by the evaluation of the implemented algorithms on real world data representative of NDE scanning. The experiment was conducted in the centre of the volume where previously it was shown that the ground truth system had the best accuracy (Section 3.3.11.1).

6.5.1 Coordinate Frame Alignment

Three coordinate frames had to be aligned during the experiment to ensure all systems involved were tracking the same point in space; the ultrasonic transmitter transducer and VMCS markers were rigidly aligned to track the midpoint of the axis defined by the drive wheels as shown in Figure 6.9. In order to ensure the constant radius, r, of the ring of three receiving transducers in Equation 5.4, markers were also placed on the ultrasonic receiver modules where marker placement was carefully chosen to avoid perturbing the received ultrasonic pulse. The modules were then aligned using the positions reported by the VMCS.

The systems were configured to track the same physical point on the robot through acquisition of 300 measurements at five grid locations yielding the point sets $\boldsymbol{p}_v$ and $\boldsymbol{p}_c$ for the Vicon and Cricket systems respectively. The locations used to compute the transform spanned the working area as shown in Figure 6.11, the system coordinate frames are shown. The procedure described in Section 4.7.1 was used to obtain a least squares estimate of the translation and rotation parameters relating the two frames; the transform is applied in results analysis to enable comparison between systems.. The average residual error between the transformed VMCS points and the ultrasonic measurements was -3.7 mm: this was the best error that could be achieved in practice and manifests as an offset in the errors calculated in the following section.

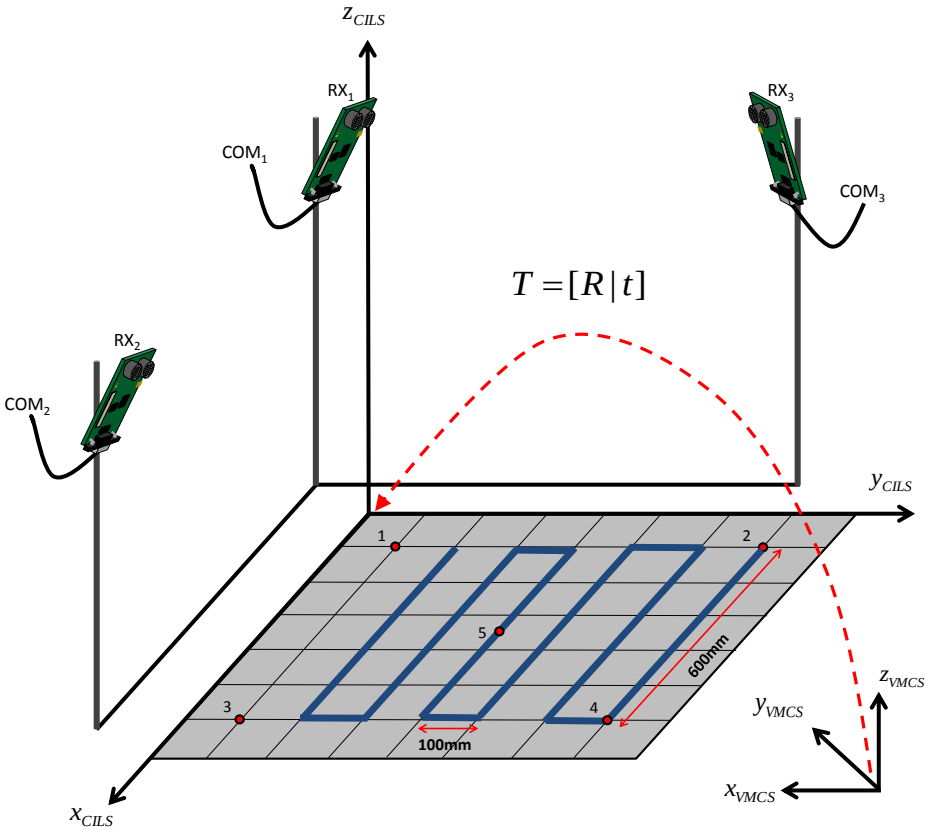


Figure 6.11: Diagram showing the five points used to align the coordinate frames of the VMCS and the ultrasonic system. The raster scan path used during the experiment is plotted on the grid.

6.5.2 Raster Scan Experiment

To simulate a typical course employed in a real NDE scan a robot was instructed to execute a raster scan consisting of 600 mm horizontal sweeps and 100 mm vertical sweeps contained in the rectangle of dimension 600 mm x 500 mm starting with the pose $\boldsymbol{p}_{start} = [300 \text{ mm}, -300 \text{ mm}, 180°]^T$ in the plane - the path is illustrated in Figure 6.11. This scan was repeated five times to generate multiple datasets for analysis.

6.5.3 Results Analysis

The trajectory estimates from all estimation sources for dataset 1 are shown in the graph of Figures 6.12 and 6.15.

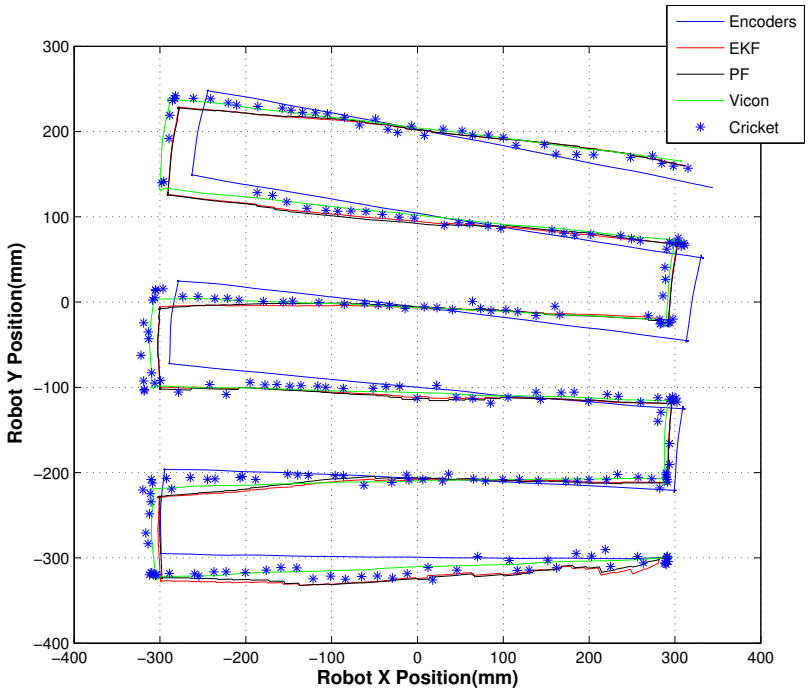


Figure 6.12: The trajectory estimates from all positional sources

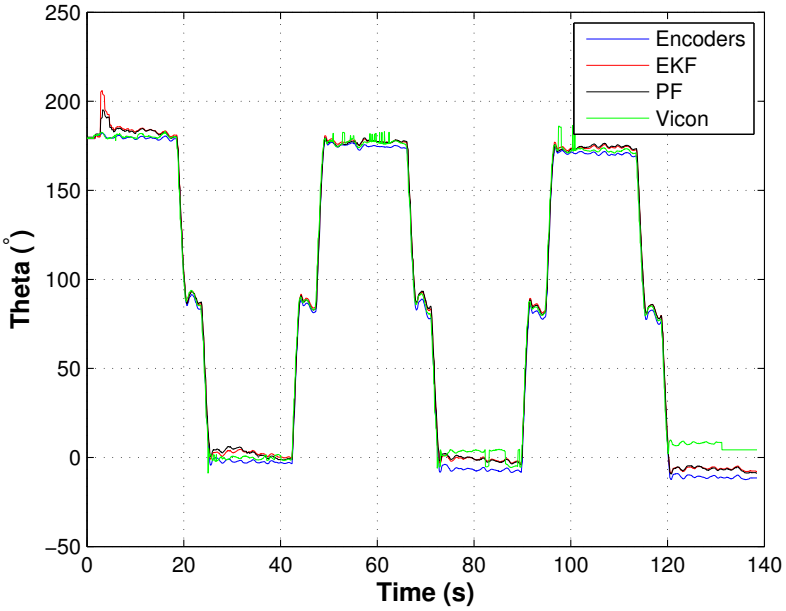


Figure 6.13: Estimate of robot orientation

The EKF and PF (using 200 particles) curves have a stronger correlation with the VMCS curve than odometry which becomes increasingly erroneous. The estimation errors of each source with respect to the VMCS is shown in Figures 6.14 and 6.15 while the numerical MSE errors are tabulated in Table 6.1.

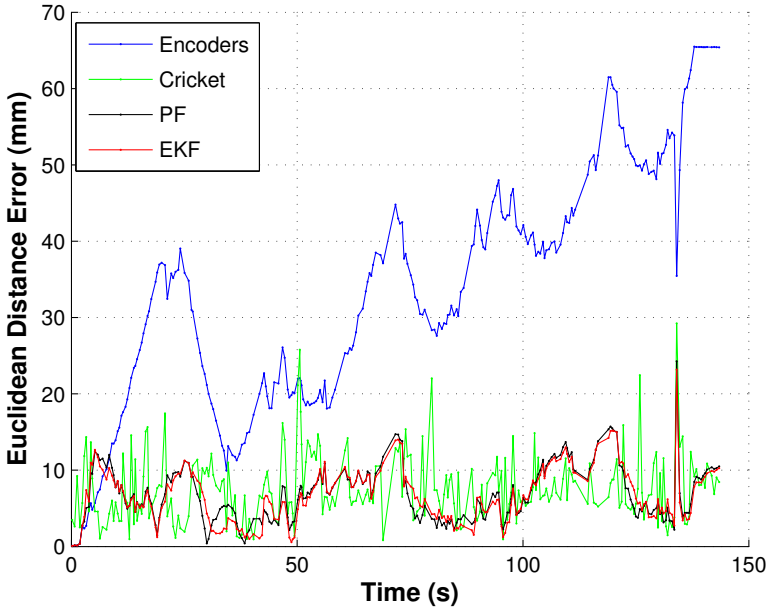


Figure 6.14: Error in (x, y) position of trajectory estimates with respect to Vicon

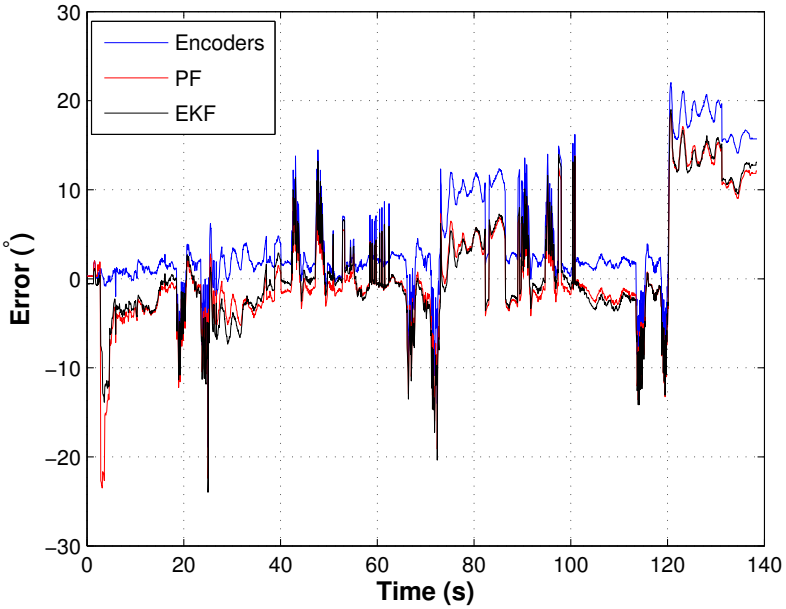


Figure 6.15: Error of robot orientation

Dataset	Estimation	XY MSE (mm^2)	θ MSE ($^{\circ 2}$)
1	EKF	66.53	36.06
	PF	67.86	33.17
	Cricket	146.58	N/A
	Odometry	774.51	51.07
2	EKF	119.18	8.76
	PF	138.94	15.74
	Cricket	178.61	N/A
	Odometry	1537.55	20.94
3	EKF	50.36	20.99
	PF	53.41	21.58
	Cricket	111.90	N/A
	Odometry	1320.73	42.68
4	EKF	69.29	17.95
	PF	71.84	19.37
	Cricket	133.69	N/A
	Odometry	2072.62	45.91
5	EKF	57.75	27.55
	PF	60.79	27.40
	Cricket	114.77	N/A
	Odometry	1379.57	56.43

Table 6.1: Pose error for each estimation source with respect to ground truth for each dataset

It is clear from the graphs that the error in odometry grows with path length. The oscillatory behaviour of the curve is due to the robot turning back to ground truth on corner rotations thus reducing the accumulating error. The filter estimates are essentially a smoothed version of the Cricket data where the odometry fulfils the smoothing function. It is clear from Table 6.1 that the error of both the PF and EKF is less than that of the positional sources used in isolation. The MSE's

for all five datasets are shown in Table 6.1. The reduction of positional error due to filtering with the EKF and PF for the different datasets is clearly shown by the MSEs in comparison to the ultrasonic system and wheel encoders.

If filter error defined as MSE in only (x, y) is considered as function of the number of particles, N, it is found that the PF error effectively saturates to the level of the EKF, this is illustrated in Figure 6.17 for dataset 1. Each point on the PF curve is calculated by averaging five runs of the code executed on the RSA which was carried out offline on a PC. It may be concluded, from this result, that the system is sufficiently linear within the system time-step defined by the odometry, that the potential gains offered by the PF are not manifest in the MSE result.

This conclusion was tested by considering the scenario in which the odometry arrives at a slower rate. The encoder data was decimated by a factor of 10 in effect simulating a larger time-step, 10 times greater than the time-step associated with the raw data. The resulting MSE vs N plot is shown in Figure 6.16.

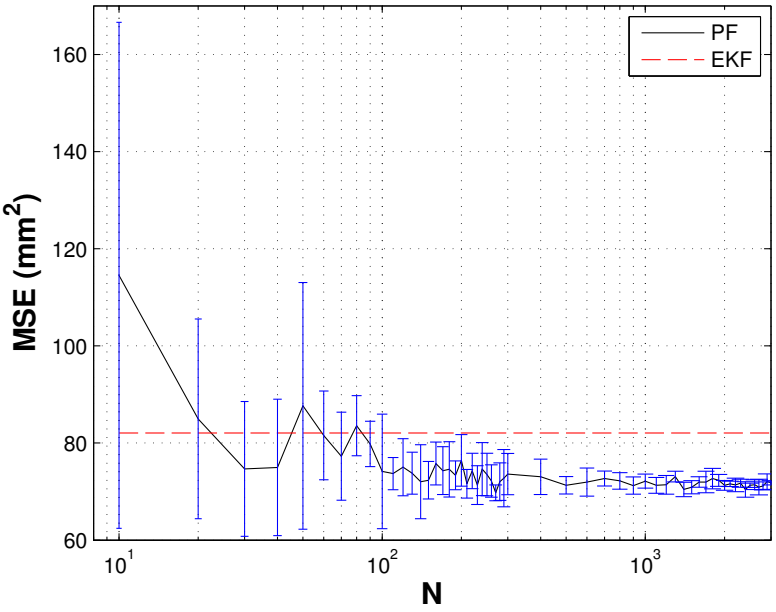


Figure 6.16: PF falls below the EKF line when data is decimated. Error bars are plotted for each PF point showing range of values used to calculate the mean point.

The PF curve now intersects at approximately 50 particles and subsequently saturates below the EKF curve where each PF point is again five runs averaged. The larger time-step means that the EKF has to linearise a more non-linear region of the state-space which gives rise to greater linearisation error and subsequently higher MSE. The saturation of the PF error to the level of the EKF in Figure 6.17 suggests that the EKF is Bayes optimal as when N becomes large the PF becomes Bayes optimal.

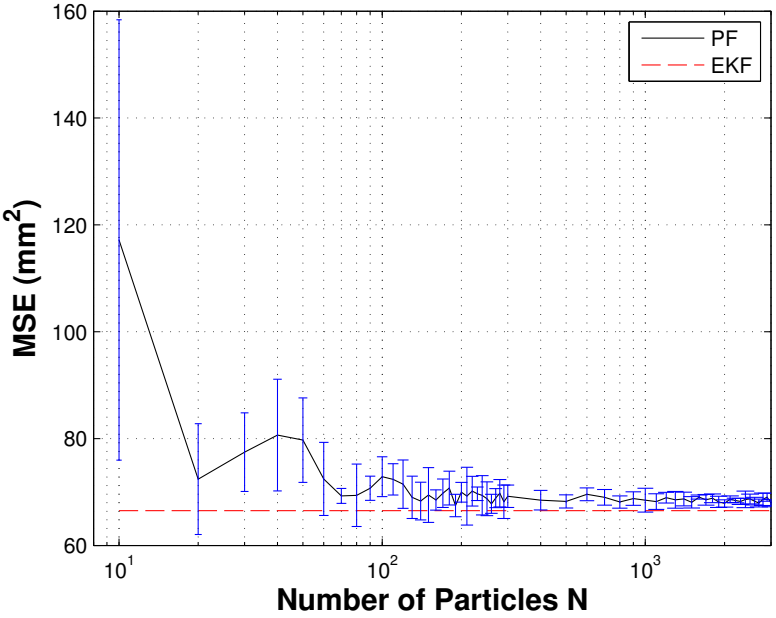


Figure 6.17: PF saturates to the level of the EKF in processing raw data. Error bars are plotted for each PF point showing range of values used to calculate the mean point.

It may be said from Figure 6.16 that the PF is more efficient in the sense that it produces an error in the decimated-data case comparable to the case processing the raw data.

6.5.4 Computational Cost

The computational cost of running the filters onboard the robot is an important factor in practice since the robot has other processing tasks running during operation. The EKF is less of a computational burden in comparison to the PF in which execution time is a function of the number of particles N. Measuring execution time resulting from running a single predict-update cycle for each filter while varying N, yields Figure 6.18.

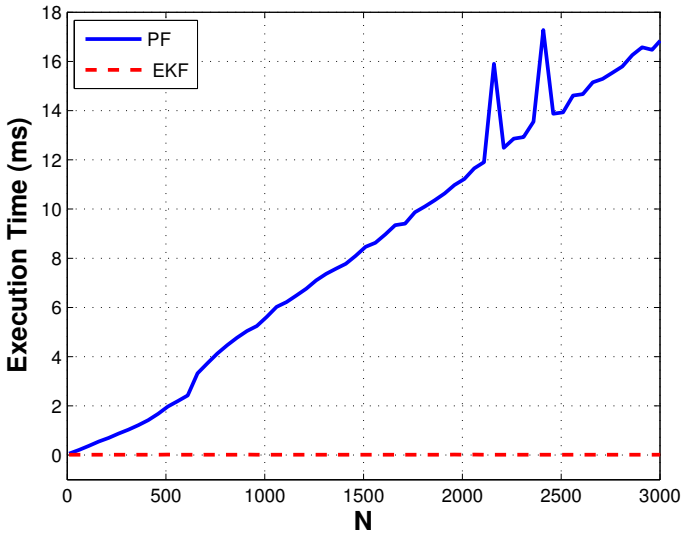


Figure 6.18: On-board execution time for a single predict-update cycle. The spikes at N = 2160 and N = 2410 result from increased CPU activity due to other processes running concurrently and should be considered outliers (but representative that computation time can increase)

The curve is valid for the particular implementation on the specific hardware being used but gives an idea of the trend that would be true given another implementation/hardware combination. The PF curve displays a linear growth with N, reaching a value of approximately 17 ms when using 3000 particles whereas the EKF approximately stays constant with a value of 0.04 ms. The EKF is more suited to real-time operation particularly when more functions are added to extend the robots capabilities.

6.6 Conclusion

This chapter focused upon methods to accurately localise a RSA using noisy positional updates from onboard wheel encoders and an external acoustic based positioning system at 100 Hz and 3 Hz respectively. The former is a relative system that is subject to drift like all such methods of estimation, while for many practical uses the latter provides absolute updates too infrequently for sole use. The use of EKF and PF Bayesian filters was investigated for combining the available positional estimates and an experimental comparison of the performance of these filters was performed.

It was demonstrated that for a typical raster scan as used in NDE, both methods yield lower path error than using either measurement source in isolation. For example in dataset one in Table 6.1 the EKF and PF acheived similar MSEs of 66.53 mm^2 and 67.86 mm^2 respectively while the ultrasonic system gave 146.58 mm^2 and optical encoders produced much higher error due to drift at 774.51 mm^2. The errors corresponding to the remaining datasets follow a similar trend with the EKF and PF having similar performance while the unfiltered sources display much greater error.

The EKF was expected to produce greater path error than the PF due to its requirement for process/measurement model linearisation, however, this was not found to be the case in practice. It is considered that the models are sufficiently linear within the system time step (dictated by the encoders) that the potential benefits of the PF do not become apparent. Indeed it also shown that if the rate of the encoder data is reduced the EKF estimation error increases as a consequence of larger linearisation error. The graph of Figure 6.16 shows that when the encoder data corresponding to dataset one is decimated by a factor of ten that the MSE for the EKF increases from ≈ 66 to ≈ 82 mm^2.

Within the constraints of the described system, the conclusion that can be drawn from this experiment is that there is no benefit in using the PF. This, however, is not true in general and shows that choice of filtering technique is dependent upon both the system setup and the models employed in the algorithms. A practical

aspect of importance for resource limited systems such as the presented system is the computational cost of algorithms running onboard. An attractive benefit of the EKF it is the ability to compute the update in a fixed time period while the cost associated with the PF is proportional to the number of particles used of which the optimal number is not always clear in advance.

Chapter 7

Conclusions and Future Work

7.1 General Overview

Automated NDE where feasible, is highly attractive, and potentially allows inspection of operational plant. The safety, environmental and financial benefits for automating NDE measurements are clear, and applicable across a broad range of NDE technology. Robotic NDE systems can be categorised as systems that are designed for a specific task and those based upon conventional robot work cells such as pick and place type robots employed to carry out repetitive scanning of components. With respect to the former, there exist a wide range of inspection environments giving rise to the development of inspection devices for a broad spread of engineering structures/components including storage tanks, pipe networks, ships' hulls and offshore structures.

The development of a bespoke miniature NDE robot and the integration of NDE payloads was previously investigated by Friedrich [9] and Dobie [14]. The overall concept was to have many of theses robots working together to form a dynamic scanning system providing a flexible and scalable inspection platform able to negotiate complex geometry surfaces. This thesis has focused upon one of the major problems in this concept, namely that of accurate positional estimation. Registering defect information with the position at which the defect was detected

on the structure under investigation is a fundamental requirement of manual and automated NDE. The obvious benefit of such knowledge is in directing subsequent repair/replacement of components. In real world NDE applications there is often a need to monitor a defect to ensure it does not reach a critical state. While the defect is non-critical the associated component/structure can still be used thus averting the shutdown of important operations. The monitoring task requires that the NDE measurements are taken at the *same* position. Positioning data is also of importance to register NDE measurements from different sensors acquired from multiple scans potentially conducted at different times.

This thesis has investigated a range of positioning strategies for a single RSA unit. Throughout this research a photogrammetry system is used to evaluate the performance of the developed algorithms. The accuracy and repeatability of this system for use as a ground truth system is investigated in Chapter 3. An onboard positioning system which extracts motion information from an image stream is developed and evaluated in Chapter 4. A low cost ultrasonic positioning system is characterised in terms of accuracy and repeatability for use in tracking applications in Chapter 5. In addition, several algorithms for ultrasonic ranging are investigated with a view to creating a high accuracy acoustic positioning system. In a real inspection scenario multiple positioning systems would need to be employed in order to deal with the failure modes of each individual system. In Chapter 6, Bayesian filters are implemented for combining multiple sources of noisy positional data to produce an overall estimate of the position of an RSA.

A review of the fundamental sensing techniques and state of the art algorithms from the field of robotics that may be applied to estimate position was presented in Chapter 2. The methods can be reduced to relative positioning, absolute positioning and SLAM based methods. Relative positioning makes use of on-robot sensors to estimate the change in position over a time increment, summing these changes results in an absolute trajectory; the optical encoders and visual odometry are examples of this type of estimation. The major issue with relative positioning is that small errors in the estimate of change accumulate when generating the absolute path potentially leading to gross errors. Absolute positioning requires external sensors with fixed location to be deployed into the operating

environment to enable position tracking; the ultrasonic system, VMCS and LAT are all systems of this type. SLAM methods remove the need to embed beacons into the operating environment through automatically generating beacons from the environment itself.

A photogrammetry system which is used as the ground truth system throughout this thesis was characterised in Chapter 3. It was recognised through empirical observations that the accuracy of positional measurements varied as a function of location within the 6.8 m x 3.8 m x 3.8 m measurement volume. Using a high accuracy/precision metrology instrument, a procedure was developed to evaluate the accuracy and repeatability of the photogrammetry system . New targets for the photogrammetry system were created allowing a common test object to be tracked by both systems while being scanned in four planes of the measurement volume. It was found that the central region of the measurement space contained the most accurate measurements while the those at the edges were least accurate. The error surface for each plane of measurements had a parabolic form. The maximum error recorded in the experiment was approximately 6 mm while the maximum standard deviation was approximately 74 μm. Importantly the results indicate that the system is precise but inaccurate. This inaccuracy is a systematic error that may be minimised through improving the calibration of the system.

In Chapter 4, a visual based relative positioning system was developed and evaluated for vehicles designed for challenging surface conditions in which conventional optical encoder base positioning is not feasible. A camera was mounted such that it observed the surface over which the vehicle was travelling. Acquired images were then transmitted over a wireless connection to a workstation for processing. The aim of the system was to stitch successive surface images and estimate the motion of the vehicle through computing the motion of image features. The mosaic resulting from the stitching process is a direct NDE output of the system in that it could be used in performing visual NDE. The applicability of the feature extraction technique was investigated for materials of the type encountered in NDE. It was shown that the technique is suitable for matte texture rich surfaces such as rusted steel and brick while problematic on surfaces such as glossy alu-

minium. The system was evaluated on several image datasets and compared with encoder based odometry. Across these datasets, the maximum distance error with respect to ground truth was found to be approximately 60 mm.

In Chapter 5, an ultrasonic positioning system, one of the major systems employed in this research was characterised. In contrast to the photogrammetry and laser system, this ultrasonic system is relatively low cost and has a small physical size. In a multi-robot inspection scenario where access limitations preclude use of the laser, the modularity of such a system gives rise to the possibility of utilising a subset of vehicles equipped with ultrasonic positioning modules as the nodes of a positioning system. A calibration procedure carried out over a working linear range of 2000 mm was implemented. After calibration the maximum error was -4.79 mm with a maximumm standard deviation 1.45 mm. Following on, a calibration procedure in two dimensions was developed to remove an observed radial type distortion in positional measurements in the operating plane of the system. In the two dimensional setting the maximum absolute error was approximately 10 mm. With a view to creating a high accuracy version of the ultrasonic system, an experimental comparison of several methods for through air ultrasonic ranging was carried out. The methods comprised of the cross correlation, phase based methods and BIRA, a novel bio-inspired algorithm. The evaluation consisted of testing distance accuracy over a linear range from 100 mm to 900 mm. It was found that the cross-correlation and cross-spectrum methods displayed similar levels of accuracy and repeatability. However, the BIRA yielded greater standard deviation for ranges greater than 300 mm.

In Chapter 6, Bayesian filtering was investigated for combining noisy positional estimates from the ultrasonic positioning system Characterised in Chapter 5 and on-robot encoders. The motivation for this work arose from the fact that all sensors have limited perception capabilities imposed by the underlying measurement principle resulting in different accuracies, uncertainties and failure modes. In robot tracking, data fusion is required to allow continuous estimation of location. A detailed experimental evaluation of an Extended Kalman and Particle Filter was implemented to fuse noisy optical encoders estimates available at 100 Hz and the ultrasonic positioning measurements available at 3 Hz to track the

planar motion of a vehicle. It was demonstrated that for a typical raster scan as used in NDE, both methods yield lower path error than using each measurement source in isolation. It was demonstrated that for a typical raster scan as used in NDE, both methods yield lower path error than using either measurement source in isolation. For example in dataset one of the experiments the EKF and PF acheived similar MSEs of 66.53 mm^2 and 67.86 mm^2 respectively while the ultrasonic system gave 146.58 mm^2 and optical encoders produced much higher error due to drift at 774.51 mm^2. The errors corresponding to the remaining datasets follow a similar trend with the EKF and PF having similar performance while the unfiltered sources display much greater error. The Extended Kalman Filter was expected to produce greater path error than the Particle Filter due to its requirement for process/measurement model linearisation however, this was not found to be the case in practice. It is considered that the models are sufficiently linear within the system time step (dictated by the encoders) that the potential benefits of the Particle Filter do not become apparent. A practical aspect of importance for resource limited systems such as the presented system is the computational cost of algorithms running onboard. An attractive benefit of the Extended Kalman Filter it is the ability to compute the update in a fixed time period while the cost associated with the Particle Filter is proportional to the number of particles used of which the optimal number is not always clear in advance.

7.2 Suggestions for Further Work

The use of SLAM techniques that provide absolute positioning without the need to modify the environment is a very desirable property given the potential for access restrictions in NDE inspections. The environment in which the robots are intended to be used can be particularly testing for the sensors. For example, unpredictable lighting make visual methods unfavourable and slippery surfaces could yield poor encoder estimates. Future work should consider methods for the extraction of robust and distinctive features from different sensors in NDE

environments. A novel approach would be to use defects within the environment itself as the features used in the algorithm for navigation. For example, sub-surface defects from an Eddy current probe could be tracked as the robot moves over the surface. Taking this further, map building due to mulitple robots that share and contribute to a global map of the environment could result in rapid inspections. Future work should also consider the use of previously generated maps to enable a robot to return to a previously visited location which is often a requirement in real world inspections.

References

[1] S. Walke A. Stacey, H. Harte. Asset Integrity Management of Ageing Offshore Installations On The UKCS A Regulatory Perspective. *Offshore Technology International*, July 2010. 1

[2] Jonathan Amos. Gulf Oil Spill Volume Estimated from Video, September 2010. 1

[3] A.H. Sherry. A Review of the UK's Nuclear R&D Capability. Technical report, Technology Strategy Board, 2010. 2

[4] World Nuclear News. Oldbury to Continue Operating into 2009. World Wide Web, December 2008. 2

[5] BBC News. Oldbury Power Station to Stay Open Until 2012. World Wide Web, June 2011. 2

[6] R. Halmshaw. *Non-Destructive Testing, 2nd ed.* Edward Arnold, London, 1991. 2

[7] P. Cawley. Non-destructive Testing-Current Capabilities and Future Directions. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, 215(4):213–223, 2001. 2, 44

[8] Frost & Sullivan. World Nondestructive Test Equipment Markets. March 2009. 3

[9] Markus Friedrich. *Design of Minature Mobile Robots for Non-Destructive Evaluation*. PhD thesis, University of Strathclyde, 2007. 4, 215

[10] M. Friedrich, G. Dobie, C.C. Chan, S.G. Pierce, W. Galbraith, S. Marshall, and G. Hayward. Miniature Mobile Sensor Platforms for Condition Monitoring of Structures. *IEEE Sensors Journal*, 9(11):1439, 2009. 4

[11] M. Friedrich, W. Galbraith, and G. Hayward. Autonomous Mobile Robots for Ultrasonic NDE. In *Proceedings of the IEEE Ultrasonics Symposium*, pages 902–905, 2006. 4

[12] M. Friedrich, L. Gatzoulis, G. Hayward, and W. Galbraith. Small Inspection Vehicles for Non-destructive Testing Applications. *Climbing and Walking Robots*, pages 927–934, 2006. 4

[13] M. Friedrich, SG Pierce, W. Galbraith, and G. Hayward. Data Fusion in Automated Robotic Inspection Systems. *Insight-Non-Destructive Testing and Condition Monitoring*, 50(2):88–94, 2008. 4

[14] Gordon Dobie. *Ultrasonic Sensor Platforms for Non-Destructive Evaluation*. PhD thesis, University of Strathclyde, 2010. 4, 215

[15] G. Dobie, W. Galbraith, M. Friedrich, SG Pierce, and G. Hayward. Robotic Based Reconfigurable Lamb Wave Scanner for Non-Destructive Evaluation. In *IEEE Ultrasonics Symposium, 2007*, pages 1213–1216, 2007. 4

[16] R. Summan, G. Dobie, J. Hensman, SG Pierce, and K. Worden. A Probabilistic Approach to Robotic NDE Inspection. In *AIP Conference Proceedings*, volume 1211, page 1999, 2010. 5

[17] H. Durrant-Whyte and T. Bailey. Simultaneous Localisation and Mapping (SLAM): Part I the Essential Algorithms. *Robotics and Automation Magazine*, 13(2):99–110, 2006. 10, 13, 34, 179

[18] B. Siciliano and O. Khatib. *Springer Handbook of Robotics*. Springer-Verlag New York Inc, 2008. 11, 18

[19] D. Voth. A New Generation of Military Robots. *Intelligent Systems, IEEE*, 19(4):2–3, 2004. 11

[20] M. Maimone, A. Johnson, Y. Cheng, R. Willson, and L. Matthies. Autonomous Navigation Results from the Mars Exploration Rover (MER) Mission. *Experimental robotics IX*, pages 3–13, 2006. 11

[21] M.J. Bakari, K.M. Zied, and D.W. Seward. Development of a Multi-Arm Mobile Robot for Nuclear Decommissioning Tasks. *International Journal of Advanced Robotic Systems*, 4(4), 2007. 11

[22] G. Giralt, R. Chatila, and R. Alami. Remote Intervention, Robot Autonomy, and Teleprogramming: Generic Concepts and Real-World Application Cases. In *Intelligent Robots and Systems' 93, IROS'93. Proceedings of the 1993 IEEE/RSJ International Conference on*, volume 1, pages 314–320. IEEE, 1993. 12

[23] E. Nebot. The Sydney Victoria Park Dataset. 2009. 17

[24] J. Borenstein, HR Everett, and L. Feng. Where am I? Sensors and Methods for Mobile Robot Positioning. *University of Michigan*, 119:120, 1996. 19, 30, 179, 199

[25] D.E. Manolakis. Efficient Solution and Performance Analysis of 3-D Position Estimation by Trilateration. *IEEE Transactions on Aerospace and Electronic Systems*, 32(4):1239–1248, 1996. 20

[26] AM Chandra. *Higher Surveying*. New Age International, 2007. 20

[27] R.I. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997. 21

[28] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*, volume 2. Cambridge Univ Press, 2000. 22, 25, 26, 112, 120

[29] M Varga. *Practical Image Processing and Computer Vision*. John Wiley, 2009. 22

[30] F. Thomas and L. Ros. Revisiting Trilateration for Robot Localization. *IEEE Transactions on Robotics*, 21(1):93–101, 2005. 26

[31] W. Schofield and M. Breach. *Engineering Surveying*. Butterworth Heinemann, 2007. 29

[32] N. Samama. *Global Positioning: Technologies and Performance*, volume 7. LibreDigital, 2008. 30

[33] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005. 31, 40

[34] P. Newman and J. Leonard. Pure Range-Only Sub-Sea SLAM. In *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, volume 2, pages 1921–1926. IEEE, 2003. 32

[35] NM Kwok and G. Dissanayake. Bearing-only SLAM in Indoor Environments using a Modified Particle Filter. In *Australasian Conference on Robotics and Automation*, volume 2003, pages 1–8. Citeseer, 2003. 32

[36] P. Newman, D. Cole, and K. Ho. Outdoor SLAM using Visual Appearance and Laser Ranging. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 1180–1187. IEEE, 2006. 32

[37] R. Smith, M. Self, and P. Cheeseman. Estimating Uncertain Spatial Relationships in Robotics. *Autonomous robot vehicles*, 1:167–193, 1990. 34

[38] D. Wolf and G.S. Sukhatme. Online Simultaneous Localization and Mapping in Dynamic Environments. In *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volume 2, pages 1301–1307. IEEE, 2004. 35

[39] T. Bailey and H. Durrant-Whyte. Simultaneous Localization and Mapping (SLAM): Part II. *Robotics & Automation Magazine, IEEE*, 13(3):108–117, 2006. 36

[40] P.C. Mahalanobis. On the Generalized Distance in Statistics. In *Proceedings of the National Institute of Science, Calcutta*, volume 12, page 49, 1936. 37

[41] J. Neira and J.D. Tardós. Data Association in Stochastic Mapping using the Joint Compatibility Test. *IEEE Transactions on Robotics and Automation*, 17(6):890–897, 2001. 37

[42] P. Newman and K. Ho. SLAM-Loop Closing with Visually Salient Features. In *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, pages 635–642. IEEE, 2005. 38

[43] Mark Cummins and Paul Newman. FAB-MAP: Probabilistic Localization and Mapping in the Space of Appearance. *The International Journal of Robotics Research*, 27(6):647–665, 2008. 38

[44] A. Elfes. Using Occupancy Grids for Mobile Robot Perception and Navigation. *Computer*, 22(6):46–57, 1989. 39

[45] S. Thrun. Learning Occupancy Grid Maps with Forward Sensor Models. *Autonomous robots*, 15(2):111–127, 2003. 39, 41

[46] B. Yamauchi, A. Schultz, and W. Adams. Mobile Robot Exploration and Map-Building with Continuous Localization. In *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volume 4, pages 3715–3720. IEEE, 1998. 39, 43

[47] A.C. Schultz, W. Adams, B. Yamauchi, and M. Jones. Unifying Exploration, Localization, Navigation and Planning Through a Common Representation. In *Robotics and Automation, 1999. Proceedings. 1999 IEEE International Conference on*, volume 4, pages 2651–2658. IEEE, 1999. 39

[48] K.M. Wurm, A. Hornung, M. Bennewitz, C. Stachniss, and W. Burgard. Octomap: A Probabilistic, Flexible, and Compact 3D Map Representation for Robotic Systems. In *Proc. of the ICRA 2010 workshop on best practice in 3D perception and modeling for mobile manipulation*, 2010. 41

[49] P. Stepan, M. Kulich, and L. Preucil. Robust Data Fusion with Occupancy Grid. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 35(1):106–115, 2005. 41

[50] A. Birk and S. Carpin. Merging Occupancy Grid Maps from Multiple Robots. *Proceedings of the IEEE*, 94(7):1384–1397, 2006. 42

[51] B. Schiele and J.L. Crowley. A Comparison of Position Estimation Techniques using Occupancy Grids. *Robotics and autonomous systems*, 12(3-4):163–171, 1994. 42

[52] S. Thrun. Learning Metric-Topological Maps for Indoor Mobile Robot Navigation. *Artificial Intelligence*, 99(1):21–71, 1998. 43

[53] E. Zwicker, W. Zesch, and R. Moser. A Modular Inspection Robot Platform for Power Plant Applications. In *Applied Robotics for the Power Industry (CARPI), 2010 1st International Conference on*, pages 1–6. IEEE, 2010. 45

[54] R. Bogue. The Role of Robotics in Non-destructive Testing. *Industrial Robot: An International Journal*, 37(5):421–426, 2010. 45

[55] JNH Tiratsoo. *Pipeline Pigging Technology*. Gulf Professional Publishing, 1992. 47

[56] J.A. Czyz, C. Fraccaroli, and A.P. Sergeant. Measuring Pipeline Movement in Geotechnically Unstable Areas Using an Inertial Geometry Pipeline Inspection Pig. In *ASME International Pipeline Conference, Calgary*, page 337, 1996. 47

[57] J. Yu, J.G. Lee, C.G. Park, and H.S. Han. An Off-Line Navigation of a Geometry PIG using a Modified Nonlinear Fixed-Interval Smoothing Filter. *Control engineering practice*, 13(11):1403–1411, 2005. 47

[58] H. Schempf, E. Mutschler, A. Gavaert, G. Skoptsov, and W. Crowley. Visual and Nondestructive Evaluation Inspection of Live Gas Mains using the ExplorerTM family of Pipe Robots. *Journal of Field Robotics*, 27(3):217–249, 2010. 47

[59] TS White, R. Alexander, G. Callow, A. Cooke, S. Harris, and J. Sargent. A Mobile Climbing Robot for High Precision Manufacture and Inspection of Aerostructures. *The International Journal of Robotics Research*, 24(7):589, 2005. 48

[60] G. Caprari, A. Breitenmoser, W. Fischer, C. Hurzeler, F. Tâche, R. Siegwart, P. Schoeneich, F. Rochat, F. Mondada, and R. Moser. Highly Compact Robots for Inspection of Power Plants. In *Applied Robotics for the Power Industry (CARPI), 2010 1st International Conference on*, pages 1–6. IEEE, 2010. 48

[61] Fabien Tache. *Robot Locomotion and Localization on 3D Complex-Shaped Structures*. PhD thesis, ETH Zurich, 2010. 48

[62] J. Katrasnik, F. Pernus, and B. Likar. A Survey of Mobile Robots for Distribution Power Line Inspection. *IEEE Transactions on Power Delivery*, 25(1):485–493, 2010. 49

[63] Cyberhawk Innovations, Accessed June 2011. 49

[64] Leica Geosystems Metrology Products. Leica Absolute Tracker: ASME B89.4.19 Specifications. 2006. 55

[65] Leica Geosystems. The Leica Absolute Interferometer: A New Approach to Laser Tracker Absolute Distance Meters. 2008. 55

[66] Vicon Motion System Ltd. T-Series Brochure, 2009. 58

[67] R. Szeliski. Image Alignment and Stitching: A Tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1):1–104, 2006. 107

[68] M. Brown and D.G. Lowe. Automatic Panoramic Image Stitching Using Invariant Features. *International Journal of Computer Vision*, 74(1):59–73, 2007. 107

[69] H.P. Moravec. Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover. *tech report CMURITR8003 Robotics Institute Carnegie Mellon University doctoral dissertation Stanford University*, 1980. 107

[70] S. Lacroix, A. Mallet, R. Chatila, and L. Gallo. Rover Self Localization in Planetary-Like Environments. In *Artificial Intelligence, Robotics and Automation in Space*, volume 440, page 433, 1999. 107

[71] D. Nistér, O. Naroditsky, and J. Bergen. Visual Odometry. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages 1–652. IEEE, 2004. 107, 108

[72] L. Mackenzie. *An Intelligent Inspection System for Non-Ferromagnetic Pipes*. PhD thesis, University of Strathclyde, 2010. 107

[73] J. Ferrer, A. Elibol, O. Delaunoy, N. Gracias, and R. Garcia. Large-Area Photo-Mosaics using Global Alignment and Navigation Data. In *MTS/IEEE Oceans Conference, Vancouver, Canada*, pages 1–9, 2007. 107, 108

[74] P. Hansen, H. Alismail, P. Rander, and B. Browning. Monocular Visual Odometry for Robot Localization in LNG Pipes. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3111–3116. IEEE, 2011. 108, 122

[75] P. Hansen, H. Alismail, B. Browning, and P. Rander. Stereo Visual Odometry for Pipe Mapping. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4020–4025. IEEE, 2011. 108

[76] Point Grey Research, Accessed May 2012. 110

[77] Fujinon DF6HA-1B, Accessed May 2012. 110

[78] libUSB, Accessed May 2012. 111

[79] P. Azad, T. Gockel, and R. Dillmann. *Computer Vision: Principles and Practice*, volume 78. elektor, 2008. 112

[80] A.J. Davison, Y.G. Cid, and N. Kita. Real-Time 3D SLAM with Wide-Angle Vision. In *Proc IFAC Symposium on Intelligent Autonomous Vehicles Lisbon*. Citeseer, 2004. 113

[81] J.Y. Bouguet. Camera Calibration Toolbox for MATLAB. 2004. 113

[82] G. Bradski and A. Kaehler. *Learning OpenCV Computer Vision with the OpenCV Library*. O'Reilly Media, 2008. 115

[83] T. Tuytelaars and K. Mikolajczyk. Local Invariant Feature Detectors: A Survey. *Foundations and Trends in Computer Graphics and Vision*, 3(3):177–280, 2008. 116

[84] D.A. Lisin, M.A. Mattar, M.B. Blaschko, E.G. Learned-Miller, and M.C. Benfield. Combining Local and Global Image Features for Object Class Recognition. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 47–47. IEEE, 2005. 116

[85] C. Schmid and R. Mohr. Local Grayvalue Invariants for Image Retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5):530–535, 1997. 116

[86] D.G. Lowe. Object Recognition from Local Scale-Invariant Features. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 2, pages 1150–1157. Ieee, 1999. 117

[87] K. Mikolajczyk and C. Schmid. A Performance Evaluation of Local Descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, 2005. 118

[88] J. Bauer, N. Sunderhauf, and P. Protzel. Comparing Several Implementations of Two Recently Published Feature Detectors. In *Proc. of the International Conference on Intelligent and Autonomous Systems*, 2007. 118

[89] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded Up Robust Features. *Computer Vision–ECCV 2006*, pages 404–417, 2006. 118

[90] S. Heymann, K. Muller, A. Smolic, B. Frohlich, and T. Wiegand. SIFT implementation and optimization for general-purpose GPU. In *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, 2007. 118

[91] A. Vedaldi and B. Fulkerson. VLFeat: An Open and Portable Library of Computer Vision Algorithms. In *Proceedings of the international conference on Multimedia*, pages 1469–1472. ACM, 2010. 118

[92] M.A. Fischler and R.C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, 1981. 119

[93] K.G. Derpanis. Overview of the RANSAC Algorithm. Technical report, Technical report, Computer Science, York University, May 2010, 2005. 120

[94] A. Nüchter, J. Elseberg, P. Schneider, and D. Paulus. Study of Parameterizations for the Rigid Body Transformations of the Scan Registration Problem. *Computer Vision and Image Understanding*, 114(8):963–980, 2010. 131

[95] K.S. Arun, T.S. Huang, and S.D. Blostein. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):698–700, 1987. 131

[96] A. Lorusso, D.W. Eggert, R.B. Fisher, and University of Edinburgh. Dept. of Artificial Intelligence. A Comparison of Four Algorithms for Estimating 3-D Rigid Transformations. *DAI RESEARCH PAPER*, 1995. 131

[97] J. Borenstein and L. Feng. UMBmark: A Benchmark Test for Measuring Odometry Errors in Mobile Robots. *Ann Arbor*, 1001:48109–2110, 1995. 138, 203

[98] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon. Bundle AdjustmentA Modern Synthesis. *Vision algorithms: theory and practice*, pages 153–177, 2000. 153

[99] Ke-Nung Huang and Yu-Pei Huang. Multiple-Frequency Ultrasonic Distance Measurement using Direct Digital Frequency Synthesizers. *Sensors and Actuators A: Physical*, 149(1):42 – 50, 2009. 156, 169

[100] N.B. Priyantha. *The Cricket Indoor Location System*. PhD thesis, Massachusetts Institute of Technology, 2005. 156, 158, 163

[101] Wen-Yuan Tsai, Hsin-Chieh Chen, and Teh-Lu Liao. High Accuracy Ultrasonic Air Temperature Measurement using Multi-Frequency Continuous Wave. *Sensors and Actuators A: Physical*, 132(2):526 – 532, 2006. 169

[102] J.A. Simmons. A View of the World Through the Bat's Ear: The Formation of Acoustic Images in Echolocation. *Cognition*, 33(1-2):155–199, 1989. 171

[103] D.K. Barton and S.A. Leonov. *Radar Technology Encyclopedia*. Artech House, 1997. 171

[104] F. Devaud, G. Hayward, and JJ Soraghan. P1G-3 Evaluation of a Bio-Inspired Range finding Algorithm (BIRA). In *IEEE Ultrasonics Symposium, 2006*, pages 1381–1384, 2006. 171, 172

[105] I. Matsuo, K. Kunugiyama, and M. Yano. An Echolocation Model for Range Discrimination of Multiple Closely Spaced Objects: Transformation of Spectrogram into the Reflected Intensity Distribution. *The Journal of the Acoustical Society of America*, 115:920, 2004. 171, 172

[106] P.A. Saillant, J.A. Simmons, S.P. Dear, and T.A. McMullen. A Computational Model of Echo Processing and Acoustic Imaging in Frequency-Modulated Echolocating Bats: The Spectrogram Correlation and Transformation Receiver. *Journal of the Acoustical Society of America*, 1993. 171, 172

[107] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics*. MIT Press, 2008. 178, 188

[108] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House Publishers, 2004. 182, 190, 193

[109] Y. Bar-Shalom, X.R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation*. Wiley Online Library, 2001. 182

[110] D.C. Montgomery, G.C. Runger, and N.F. Hubele. *Engineering Statistics*. Wiley, 2009. 182

[111] A. Gelb. *Applied Optimal Estimation*. The MIT press, 1974. 184, 189

[112] R.E. Kalman et al. A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, 82(1):35–45, 1960. 186

[113] S.J. Julier and J.K. Uhlmann. New Extension of the Kalman Filter to Nonlinear Systems. In *AeroSense'97*, pages 182–193. International Society for Optics and Photonics, 1997. 188, 195, 196

[114] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, 2002. 192, 194

[115] D.J.C. MacKay. *Information Theory, Inference, and Learning Algorithms*. Cambridge Univ Press, 2003. 193

[116] F. Gustafsson. Particle Filter Theory and Practice with Positioning Applications. *Aerospace and Electronic Systems Magazine, IEEE*, 25(7):53–82, 2010. 195

[117] F. Van Der Heijden. Consistency Checks for Particle Filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(1):140–145, 2006. 196

[118] Vinay A. Bavdekar, Anjali P. Deshpande, and Sachin C. Patwardhan. Identification of Process and Measurement Noise Covariance for State and Parameter Estimation Using Extended Kalman Filter. *Journal of Process Control*, 21(4):585 – 601, 2011. 197

[119] C.H. Tong and T.D. Barfoot. A Comparison of the EKF, SPKF, and the Bayes Filter for Landmark-Based Localization. In *Computer and Robot Vision (CRV), 2010 Canadian Conference on*, pages 199–206. IEEE, 2010. 197

[120] J.L. Peralta-Cabezas, M. Torres-Torriti, and M. Guarini-Hermann. A Comparison of Bayesian Prediction Techniques for Mobile Robot Trajectory Tracking. *Robotica*, 26(05):571–585, 2008. 197

[121] Nicola Bellotto and Huosheng Hu. People Tracking with a Mobile Robot: A Comparison of Kalman and Particle Filters. In *Proceedings. 13th IASTED International. Conference. Robot. Applied*, pages 388–393, 2007. 198