Analysing and Enhancing the Coarse Registration Pipeline

A thesis submitted in fulfilment of
the requirements for the degree of

Doctor of Philosophy

in Electronic Engineering at
The University of Waikato
by

Robert L. Larkins

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2015

Abstract

The current and continual development of sensors and imaging systems capable of acquiring three-dimensional data provides a novel form in which the world can be expressed and examined. The acquisition process, however, is often limited by imaging systems only being able to view a portion of a scene or object from a single pose at a given time. A full representation can still be produced by shifting the system and registering subsequent acquisitions together. While many solutions to the registration problem have been proposed, there is no quintessential approach appropriate for all situations. This dissertation aims to coarsely register range images or point-clouds of *a priori* unknown pose by matching their overlapping regions.

Using spherical harmonics to correlate normals in a coarse registration pipeline has been shown previously to be an effective means for registering partially overlapping point-clouds. The advantage of normals is their translation invariance, which permits the rotation and translation to be decoupled and determined separately. Examining each step of this pipeline in depth allows its registration capability to be quantified and identifies aspects which can be enhanced to further improve registration performance. The pipeline consists of three primary steps: identifying the rotation using spherical harmonics, identifying the translation in the Fourier domain, and automatically verifying if alignment is correct. Having achieved coarse registration, a fine registration algorithm can be used to refine and complete the alignment.

Major contributions to knowledge are provided by this dissertation at each step of the pipeline. Point-clouds with known ground-truth are used to examine the pipeline's capability, allowing its limitations to be determined; an analysis which has not been performed previously. This examination allowed modifications to individual components to be introduced and measured, establishing their provided benefit. The rotation step received the greatest attention as it is the primary weakness of the pipeline, especially as the nature of the overlap between point-clouds is unknown. Examining three schemes for binning normals found that equiangular binning, when appropriately normalised, only had a marginal decrease in accuracy with respect to the icosahedron and the introduced Fibonacci schemes. Overall, equiangular binning was the most appropriate due to its natural affinity for fast spherical-harmonic conversion. Weighting normals was found to provide the greatest benefit to registration performance. The introduction of a straightforward method of combining two different weighting schemes using the orthogonality of complex values increased correct alignments by approximately 80% with respect to the next best scheme; additionally, point-cloud pairs with overlap as low as 5% were able to be brought into correct alignment. Transform transitivity, one of two introduced verification strategies, correctly classified almost 100% of point-cloud pair registrations when there are sufficient correct alignments.

The enhancements made to the coarse registration pipeline throughout this dissertation provide significant improvements to its performance. The result is a pipeline with state-of-the-art capabilities that allow it to register point-cloud with minimal overlap and correct for alignments that are classified as misaligned. Even with its exceptional performance, it is unlikely that this pipeline has yet reached its pinnacle, as the introduced enhancements have the potential for further development.

Acknowledgements

Endeavouring to complete this thesis has been by far the most strenuous task that I have subjected upon myself, and it would not have been achieved without the support and encouragement of the people around me. Firstly, I would like to thank Michael Cree and Adrian Dorrington. Your supervision and patience throughout this research has allowed me to comprehend and expand upon a range of challenging topics, which when described to others results in glazed eyes. Additionally, your many reviews of my written work has markedly improved my writing ability, though at the detriment of many a red pen.

The collaborative environment of C.1.01, the home base of the Chronoptics Research Group, has provided an enjoyable space in which to work and engage with like-minded peers. It is here that I meet and made friends with Andrew Payne, Richard Conroy, Refael Whyte, John Godbaz, Lee Streeter and others. The camaraderie of this group made it an easy place to ask questions and discuss ideas, allowing the collective knowledge of the group to be drawn upon. I especially enjoyed our games of Settlers of Catan and 500 on Friday afternoons.

My gratitudes go out to Stuart Robinson, Jan Boehm and Mona Hess, your welcoming of me to UCL and making me part of the group, especially outside of the office, made my time in London so much more enjoyable. Without this hospitality, there would be many aspects of London that I would have missed out on, especially the many evening outings with good company.

My gratefulness also goes out to my family, especially my wife Michelle. While you may not have always understood my research, you endured me being a perpetual student. I would not have completed this journey without your motivation and emotional support. You have my whole-hearted appreciation.

Contents

List of Figures

List of Tables

List of Symbols

Variable Types

$\mathbb{N}$	The set of natural numbers beginning at zero
$\mathbb{R}$	The set of real numbers
$\mathbb{R}^3$	Three-dimensional Cartesian coordinate system
$\mathbb{Z}$	The set of integers
$\mathbb{S}^2$	The 2-sphere; the two-dimensional surface of a sphere
$\mathbf{x}$	A vector
$\hat{\mathbf{x}}$	A unit vector; $\lvert\lvert\hat{\mathbf{x}}\rvert\rvert = 1$
$\hat{\mathbf{n}}$	Unit vector normal to a surface comprised as $\hat{\mathbf{n}} = [n_x, n_y, n_z]$
$\mathbf{X}$	A Matrix
$\mathbf{R}$	3×3 rotation matrix
$\hat{x}$	The representation of x in the Frequency domain

Unary Operations

$\bar{x}$	Complex conjugate of x
$\mathbf{X}^\intercal$	Matrix transpose of $\mathbf{X}$
$\lvert x \rvert$	Absolute value of x
$\lfloor x \rfloor$	Floor of x
$\lceil x \rceil$	Ceiling of x
$\lvert\lvert\mathbf{x}\rvert\rvert$	Euclidean norm or magnitude of the vector $\mathbf{x}$

Binary Operations

$\mathbf{A} \circ \mathbf{B}$	Hadamard product of $\mathbf{A}$ and $\mathbf{B}$
$\mathbf{a} \cdot \mathbf{b}$	Dot product of $\mathbf{a}$ and $\mathbf{b}$

Common Variables, Constants and Functions

i	The imaginary unit, given as $\sqrt{-1}$
j	Indexing value for an array or a set of variables
e	The mathematical constant approximately equal to 2.718
α, β, γ	The three Euler angles for constructing a rotation matrix
θ, ϕ	The polar and azimuthal angles, respectively, of a spherical coordinates system on a unit sphere
$\mathbf{C}_R$	The rotation correlation matrix
$\mathbf{C}_T$	The translation correlation matrix
B_t	Transform bandwidth; the maximum bandwidth normals are represented in the Frequency domain
B_c	Correlation bandwidth; the maximum bandwidth when inverting the Frequency domain
$\mathrm{O}(x)$	Big O notation
$\mathrm{Tr}\,(\mathbf{M})$	The trace of the square matrix $\mathbf{M}$

Chapter 1

Introduction

The current development of imaging systems capable of three-dimensional (3-D) data acquisition allows the world to be expressed in a digital form that enables new and novel examination processes to be performed. These range imaging systems (Blais, 2004) are however only able to view a portion of the scene or object as they are frequently restricted to a single pose at a given time. Because of this limitation, the imaging system is shifted relative to the scene to acquire multiple images in order to reveal more of the scene. The task of bringing these individual images together and correctly aligning them in a single coordinate system is known as *registration*; Figure 1.1 provides a simplified demonstration of registration. While many solutions to the registration problem have been proposed (Salvi *et al.*, 2007; Tam *et al.*, 2013), it is still an active area of investigation as there is currently no quintessential approach appropriate for all situations.

Representing each range image in a standard form allows registration algorithms to process data acquired by different types of range imaging systems. Point-clouds provide the most intuitive representation of this data as they are expressed as a set of points in a 3-D Euclidean coordinate system. The structure of the point-cloud is formed from individual surfaces which make up the sampled scene, with each point identifying a location on a surface. Each surface is deemed to be a contiguous set of points which maintain a consistent orientation. The surface orientation at a point is given by its associated normal, which is a vector perpendicular to the surface (see Section 3.1).

Correctly registering together a set of 3-D point-clouds enhances the representation of the object or scene, making it better suited for a wider range of applications than that of a point-cloud acquired from a single pose. Many fields have existing applications that can benefit from registration, reducing

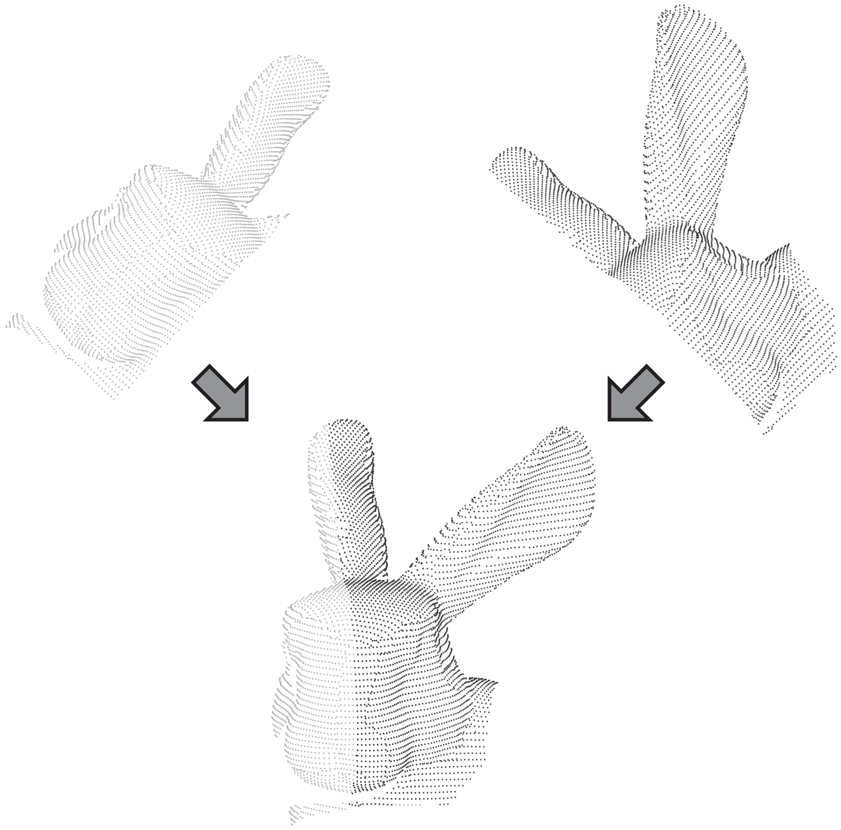


Figure 1.1: Simplified illustration of registration showing two separate halves of the Stanford bunny's head being brought into alignment.

the amount of time, cost or manual intervention required. Bi and Wang (2010) provide a list of industrial applications that include reverse engineering, rapid prototyping, part location and alignment, virtual assembly, and the inspection and measurement of manufactured items (Tian *et al.*, 2009). Simultaneous localisation and mapping (SLAM) uses registration techniques when constructing 3-D maps of an environment (Morell *et al.*, 2012); these techniques are also used for relocalisation, in which an unknown imaging pose is established with a previously generated map (Martínez-Carranza *et al.*, 2013). Grading of fruit and vegetables can also benefit from registration by forming a complete representation of their structure and comparing it with respect to a gold standard (Moreda *et al.*, 2009; Kondo, 2010). This idea also applies to surveying of construction sites to monitor progress (Kim *et al.*, 2011) and possible extensions such as digital mock-up and simulation.

These 3-D representations are also be used for object recognition and classification (Halma *et al.*, 2010), allowing objects to be identified in real world scenes. Virtual reality and gaming can also make use of point-clouds, as real-world objects and scenes can be converted directly into the virtual world.

This is a useful application for locations that may have limited access, such as cultural heritage sites, or museums whose collections are not on permanent display (Blais and Beraldin, 2006). In the medical field, registration can pertain to a wide variety of applications, such as the design and fitting of orthotics (Thabet *et al.*, 2011) and prosthetics (Mahmood *et al.*, 2012), or registering internal body parts to a model for efficient analysis (Gutman, 2013). Registration is not limited to just these applications, as there are an abundant number of new and novel applications that can be developed from the examination of registered point-clouds.

1.1 Background

Range imaging systems with no knowledge of their global position, when acquiring an image, produce a point-cloud whose coordinate system is formed with respect to the imaging system. Therefore, each point-cloud is established with its own coordinate system, meaning that the coordinate system does not intuitively change with the imaging system's pose. The consequence is that simply placing these point-clouds together in the same coordinate system results in them being layered haphazardly on top of each other instead of forming a cohesive alignment; a visual example of this is presented in Figure 1.2. Registering all point-clouds so that they correctly portray the imaged scene requires fixing one point-cloud and transforming the others with respect to it. The transform which registers one point-cloud with another is the combination of a rotation and a translation; scale does not need to be taken into account as range imaging systems produce point-clouds at a consistent scale.

The transformations that align point-clouds together are *a priori* unknown if there is no external system in place for tracking how the imaging system or object has moved. In this situation, the structural elements (such as planar surfaces, edges and corners) of the point-clouds are examined and matched to identify the transformation which provides the best alignment. It is this task which is essential for robustly solving the registration problem. The difficulty of finding the transformation is influenced by point-cloud quality, number of sample points and their distribution, and the percentage of overlap between point-clouds. Point-cloud quality is given by the accuracy and precision of the imaging system and its capability to minimise noise. As quality decreases, the ability to identify common elements between point-clouds also decreases. Increasing the number of sample points improves the characterisation of each

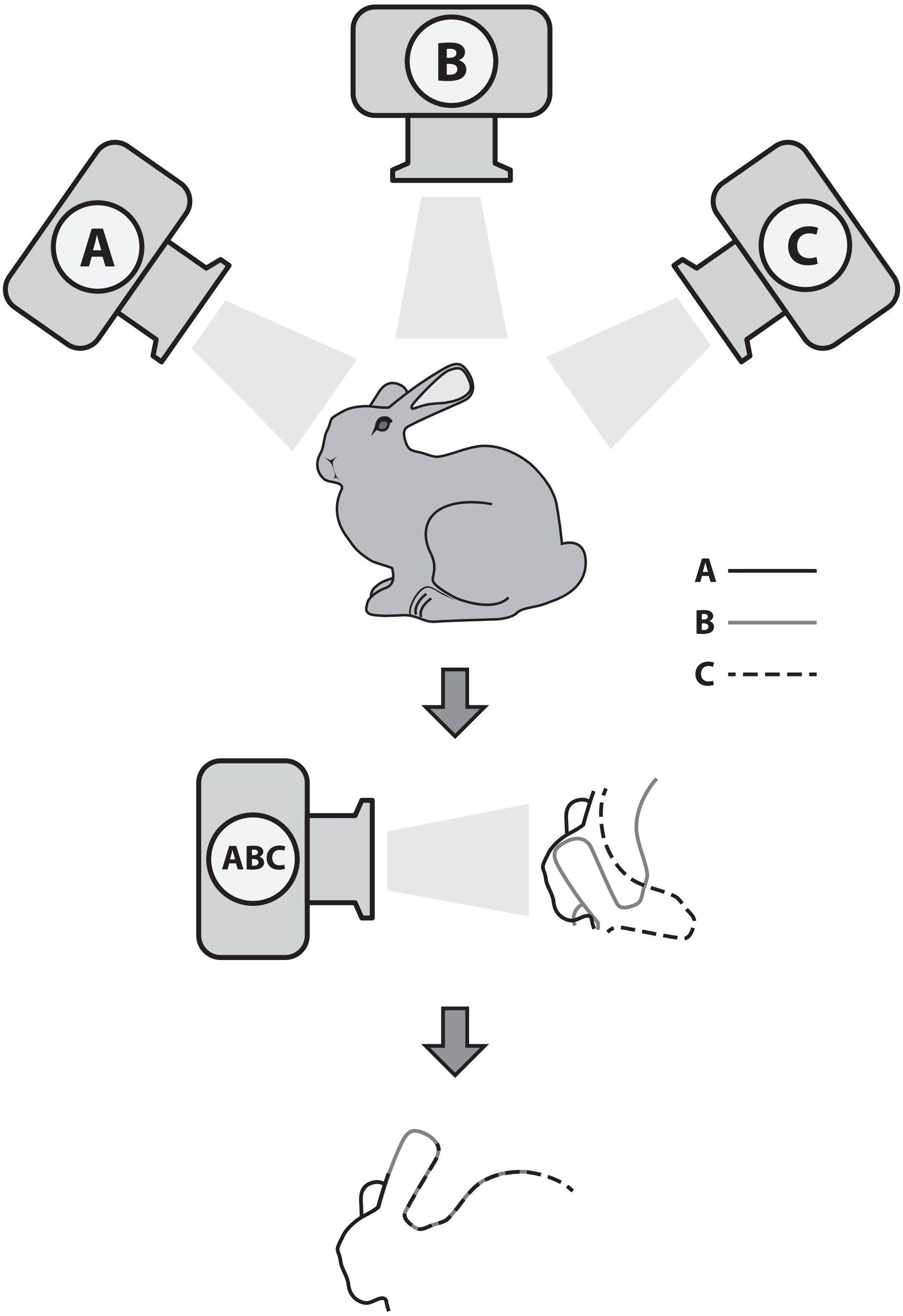


Figure 1.2: Visual example showing different imaging poses (**A**, **B** and **C**) only sampling a portion of the object's surface, and how these samples, when viewed from the camera's perspective, are initially positioned with respect to each other. Registering these samples then aligns them together in the same coordinate system.

structural element. Each imaging system's pose has a unique distribution of sampling points in the scene, meaning that the characterisation of each structural element will vary with pose. The pose also determines point-cloud overlap, thus as overlap decreases, the number of structural elements common to both point-clouds also decreases, escalating registration difficulty.

Point-cloud registration is normally performed in two stages: coarse registration followed by fine registration. Coarse registration algorithms bring point-clouds with any orientation difference into a close approximate alignment. Fine registration algorithms, such as iterative closest point (ICP) (Chen and Medioni, 1991; Besl and McKay, 1992), complete the alignment by providing near optimal minimisation of the misalignment between point-clouds. However, they require the rotation and translation error to be sufficiently small, otherwise the found alignment can become trapped in local minima. The required capability of the coarse registration pipeline is determined by the aforementioned influences, with these influences varying with both the type of imaging system used and the range images they produce.

1.2 Motivation and Objectives

The initial motivation for analysing the registration pipeline was to register range images acquired by simultaneous full-field range imaging systems that use the amplitude modulated continuous wave (AMCW) time-of-flight method (Dorrington *et al.*, 2007). These systems have had an increase in attention (Kolb *et al.*, 2010) as they are able to acquire individual range samples simultaneously across the full field-of-view. Due to both the nature of how they image a scene and their relative infancy, these imaging systems have inherent limitations that can degrade acquisition quality. To minimise these limitations, the registration pipeline needs to be sufficiently robust to increase correct alignments.

The limitations of these imaging systems encompass a variety of systematic aspects that pertain to image quality and how these imaging systems can be used. The resolution of an acquired range image is comparatively sparse with respect to more mature imaging systems, as these systems can produce dense sampling with tens of millions or more points within the same field-of-view. Examples of resolution for three commercially available cameras are 176×144 for the Mesa SwissRanger SR4500 (Mesa Imaging, 2014), 320×240 for the SoftKinetic DS325 (SoftKinetic, 2014) and 512×424 for the Microsoft Kinect

2.0 (Payne *et al.*, 2014). Along with there being fewer sample points, their accuracy and precision quality can also be degraded by various phenomena that relate to how the infrared active illumination is reflected and collected. Two such phenomena are mixed pixels, which occurs when collected light is reflected back by two or more surfaces, and multipath interference, which is when the path of light bounces off of two or more surfaces before returning. Various other phenomena such as subsurface scattering, pixel saturation, intralens scattering, blurring due to depth-of-field and undesirable electronic effects can also occur (Foix *et al.*, 2011; Godbaz, 2012). These imaging systems are also limited by how they can be used. The modulation frequency of their light source determines the range to which they can measure before ambiguity becomes an issue; a 30MHz modulation frequency equates to five metres unambiguous range. The colour of an object within a scene also influences its signal to noise ratio, with white objects imaging well, and black objects typically producing noise due to insufficient light being returned (Dorrington *et al.*, 2007). Motion blur can also corrupt an image if the imaging system is moved (or the scene has motion) during image acquisition.

The objective of this thesis is to analyse and enhance the coarse registration pipeline proposed by Makadia *et al.* (2006); a comprehensive analysis of this pipeline, which has not been performed previously, is carried out to determine its capabilities and limitations. The pipeline performs registration by correlating spherical harmonics to find the rotational alignment between two point-clouds independently of the translational alignment. The registration is coarse as the correlation only tests a discrete set of rotations or translations, therefore, some degree of misalignment will invariable remain. The pipeline then attempts to automatically classify whether the found alignment is correct by examining the relationship between the point-clouds.

While this pipeline has had minimal investigation since its conception, it shows a lot of promise for the registration problem: Makadia *et al.* showed that it can register point-clouds with overlap as low as 45%. The only input this pipeline requires are partially overlapping point-clouds (with any orientation difference) that have a normal associated with each point. As the pipeline uses all the points and normals of a point-cloud, all of its structural elements are used. This differs from other registration approaches, such as key-point matching (see Section 2.3.2), that only use a subset of points.

The specific focus is on coarsely registering partially overlapping sparse point-clouds (as opposed to point-clouds with dense sampling or are full models) which may be afflicted by the aforementioned limitations of the AMCW range imaging systems. Establishing the capability of the pipeline using only the point-cloud overlap requires that no forms of scene preparation are introduced (see Section 2.2); scene preparation includes measuring or restricting the imaging system's pose by using external apparatus, such as accelerometers, as they reduce the registration search space. The analysis of the pipeline deems the rotational alignment correct if its misalignment is 10° or less, as fine registration algorithms, such as ICP, have been shown capable of handling this amount of misalignment (Larkins *et al.*, 2010; Minguez *et al.*, 2006). The translational alignment is deemed correct based on the sampling density of the tested datasets (see Section 7.2).

Herein each step of the pipeline is analysed to provide a reliable quantification of its performance, with potential improvements being proposed and contrasted to establish whether they enhance registration capabilities. The goal of these enhancements is to produce a pipeline whose registration capability is more robust than its predecessor, while only requiring minimal overlap between point-clouds.

1.3 Thesis Organisation

The organisation of this dissertation is structured across 10 chapters in a manner akin to that of the coarse registration pipeline itself. Figure 1.3 is a visual representation of the pipeline as a flow diagram, with each major step identified. Each major step that is analysed is contained within its own chapter, though the fundamental mathematics of the rotation and translation are presented in a separate preliminaries chapter. Where applicable, the following descriptions of each chapter identify the step in the flow diagram to which they correspond.

Chapter 2 contains the literature review pertaining to the current state of the registration field. The types of systems capable of range image acquisition are briefly described first, as these produce the images that need registration. The algorithms presented in literature for both coarse and fine registering are then critiqued, with their strengths and weaknesses identified. Different strategies for automatically verifying correct registration are then presented.

Chapter 3 outlines the mathematical preliminaries required for the coarse registration pipeline. This primarily covers spherical-harmonic conversion and correlation for finding the rotation, and the binning of points and correlating

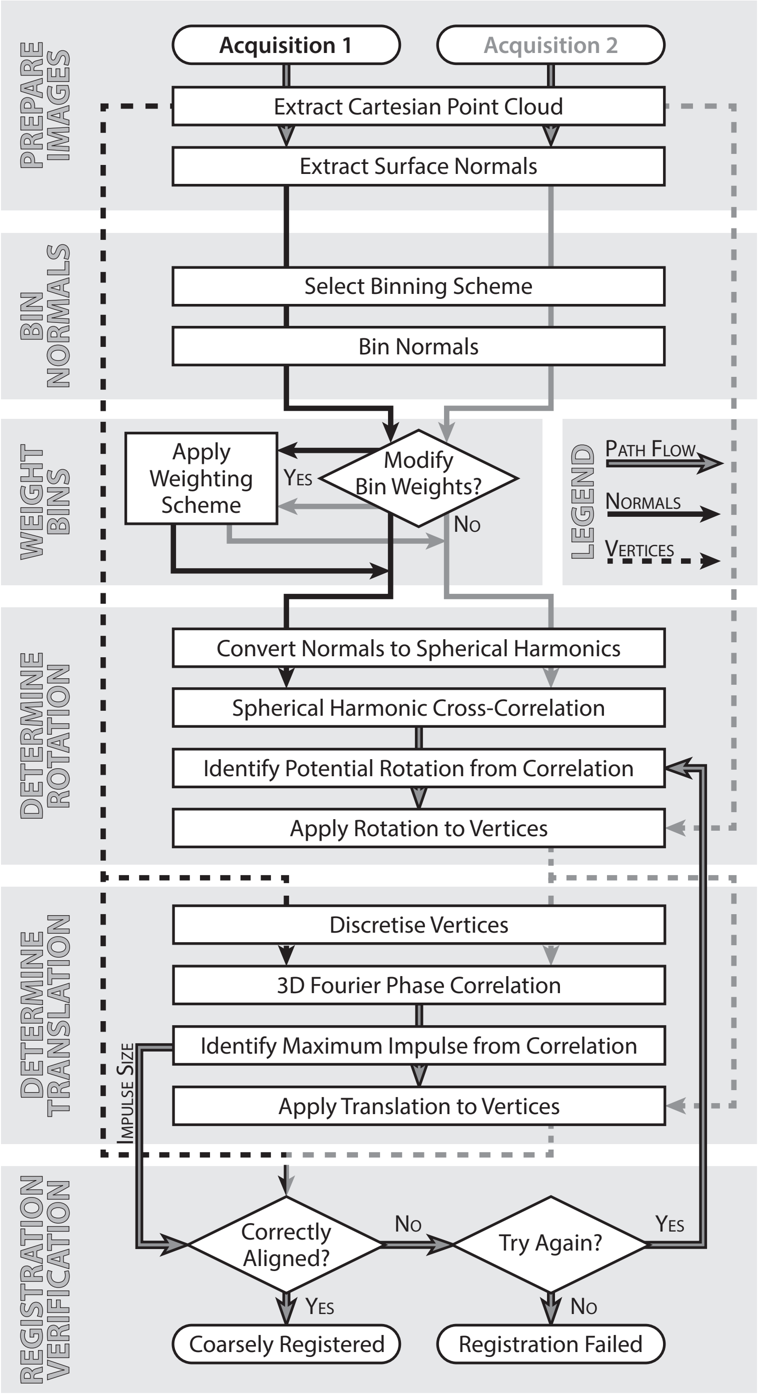


Figure 1.3: The flow diagram of the coarse registration pipeline for registering two partially overlapping range image acquisitions. The flow diagram is broken into major steps, each of which (except *Prepare Images*) are investigated throughout the body of this dissertation.

their histograms in the frequency domain to find the translation. The effects that these mathematical approaches have on registration are also discussed. The mathematics behind both the *Determine Rotation* and *Determine Translation* steps of the flow diagram are covered in this chapter.

Chapter 4 establishes the three ground-truth datasets used for analysing each step of the pipeline. This chapter also details how the ground-truth data is used to assess the overall registration performance.

Chapter 5 tests three schemes for binning normals to determine what effect the distribution of bins on the unit sphere has on registration performance. A new algorithm for binning normals on a Fibonacci bin distribution is introduced and the appropriate normalisation of equiangular binning is established. This chapter corresponds to the *Bin Normals* step in the flow diagram.

Chapter 6 investigates whether different schemes for reweighting either the normals or the bins improves the representation of each point-cloud for correlation. A representation that is conducive to correlation increases the likelihood of identifying the correct rotation. A novel method of combining two weightings together is introduced and shown to significantly improve registration accuracy. This chapter forms the *Weight Bins* step of the flow diagram.

Chapter 7 analyses the capability of translational registration with respect to how frequently it achieves correct translation when correct rotation has been identified. This chapter constitutes the analysis of the *Determine Translation* step in the flow diagram.

Chapter 8 details three approaches for verifying whether correct alignment has been achieved between two or more point-clouds. The ability of alternative rotation selection is also tested to establish the amount of benefit that it provides if an incorrect rotation is selected. This chapter is the *Registration Verification* step, which is the last step in the flow diagram, and as such, it completes the coarse registration pipeline.

Chapter 9 takes the produced coarse registration pipeline and concisely evaluates its ability to correctly align point-clouds from real datasets. Where possible, this pipeline is also tested on datasets that have been used with other registration algorithms, allowing the comparative ability of the pipeline to be established.

Chapter 10 summarises both the key findings and the contributions made to the field of 3-D registration. This dissertation is then concluded by discussing the future directions in which this research could be further progressed.

1.4 Publications

The following publications were prepared and presented during the course of this research:

- Larkins, R.L., M.J. Cree, and A.A. Dorrington. Evaluating equiangle binning for spherical harmonic registration. In: *Image and Vision Computing New Zealand (IVCNZ), 29th International Conference*, pp. 230–235 (2014).

- Larkins, R.L., M.J. Cree, and A.A. Dorrington. Analysis of weighting of normals for spherical harmonic cross-correlation. In: *Three-Dimensional Image Processing (3DIP) and Applications*, volume 8650 of *Proceedings of SPIE* (2013).

- Larkins, R.L., M.J. Cree, and A.A. Dorrington. Verification of multi-view point-cloud registration for spherical harmonic cross-correlation. In: *Image and Vision Computing New Zealand (IVCNZ), 27th International Conference*, pp. 358–363 (2012).

- Larkins, R.L., M.J. Cree, and A.A. Dorrington. Analysis of binning of normals for spherical harmonic cross-correlation. In: *Three-Dimensional Image Processing (3DIP) and Applications II*, volume 8290 of *Proceedings of SPIE* (2012).

- Larkins, R.L., M.J. Cree, and A.A. Dorrington. Analysis of ICP variants for the registration of partially overlapping time-of-flight range images. In: *Image and Vision Computing New Zealand (IVCNZ), 25th International Conference* (2010).

- Larkins, R.L., M.J. Cree, A.A. Dorrington, and J.P. Godbaz. Surface projection for mixed pixel restoration. In: *Image and Vision Computing New Zealand (IVCNZ), 24th International Conference*, pp. 431–436 (2009).

- Cree, M.J., J.P. Godbaz, R. Larkins, W.H. Round, L. Streeter, A.A. Dorrington, R. Künnemeyer, A.D. Payne, and D. Worsley. Computer vision and image processing at the University of Waikato. In: *Image and Vision Computing New Zealand (IVCNZ), 25th International Conference* (2010).

Chapter 2

Review of Literature

The extent of literature within the field of 3-D registration is numerous. This literature encompasses a broad range of approaches and algorithms that are often designed to handle a particular type of registration problem, or register 3-D data produced from a specific source. This chapter begins by detailing the different techniques used by range imaging systems to produce 3-D data and what effect they have on registration. The different approaches for registration are then reviewed by firstly describing scene preparation, which is a manual solution. This is followed by the main categories of registration algorithms in the published literature. The final section touches on different strategies proposed for verifying the alignment is correct.

2.1 Range Imaging Systems

The last three decades have seen increased growth of systems capable of acquiring range images that allow an object or scene to be digitally represented in 3-D. Blais (2004) and Bi and Wang (2010) provide in-depth reviews detailing different types of range imaging systems that have been developed. These range imaging systems are classified based on the techniques behind their 3-D acquisition, which fall into two main categories: passive or active acquisition.

Passive range imaging measures the scene without using an active illumination light source. A variety of techniques for passively determining range information are shape-from-shading (Zhang *et al.*, 1999; Xiong *et al.*, 2013), photometric stereo (Herbort and Wöhler, 2011), structure-from-motion (Oliensis, 2000; Corsini *et al.*, 2013), passive stereo vision (Tippetts *et al.*, 2013), depth from defocus (Sun *et al.*, 2013), coded aperture (Levin *et al.*, 2007) and plenoptic cameras (Perwaß and Wietzke, 2012). These forms of acquisition are less sensitive to constraints in the scene, such as reflective surfaces or surface material, and are more suited to handling motion, such as on a mobile platform. This reduced sensitivity is due to passive acquisition not requiring the scene to be actively illuminated for a measurement to be produced. The principles underlying passive acquisition techniques do have their own limitations which affect ability and accuracy. The primary limitation is the difficulty to infer absolute depth when there is an absence of surface lighting or texture, as parts of the scene at different ranges appear the same (Levin *et al.*, 2007; Bi and Wang, 2010). Stereo has the additional limitation of decreased range accuracy if its two views are not aligned accurately (Tippetts *et al.*, 2013).

Range imaging systems that actively illuminate the scene measure range by emitting and collecting light in a manner that pertains to their underlying measurement principle. Active illumination techniques include time-of-flight, interferometry, and active triangulation approaches, such as structured lighting and 3-D laser scanning; Blais (2004), Bi and Wang (2010) and Payne (2008) describe these in detail. The advantage of active illumination systems is that they can achieve sub-millimetre accuracy (Dorrington *et al.*, 2007) without needing texture information in the scene. Additionally, they can operate in environments which have no external light sources. Active illumination techniques do have limitations that affect range measurement. If their technique requires multiple measurements to establish range, then camera or scene movement can induce motion distortions, which can be difficult to detect and fix (Lee, 2014). Other factors, such as external light sources, scene structure and surface materials, can also corrupt or distort the returned light, reducing measurement accuracy (Foix *et al.*, 2011; Kolb *et al.*, 2010; Godbaz, 2012).

The range imaging system appropriate for a particular task is dependent upon its strengths and weaknesses. The technique used to measure range determines acquisition quality, which for registration affects its alignment accuracy. Different systems solve the registration problem by incorporating a hardware based solution into their set-up to provide an exact measure of pose. These include mechanical arms (Hexagon Metrology, 2014; GOM, 2014), fixed rigs (Arius 3D, 2014), inertia sensors (Pribanic *et al.*, 2013) and external reference apparatus (Penman *et al.*, 2008). While these solutions help mitigate the need for additional registration algorithms, they are not always successful if the object or additional apparatus need repositioning to complete the acquisition process. The coarse registration pipeline analysed herein can process any range imaging data as long as each sample point has an associated normal.

2.2 Scene Preparation

The registration problem is often handled by introducing markers or targets into the scene prior to acquisition, allowing the relationship between multiple images to be easily determined (Bornaz *et al.*, 2002). This is achieved for two overlapping images by identifying three or more targets that are common to both images; if there are fewer than three targets, then alignment ambiguity occurs, and a rigid transform cannot be calculated. Additionally, the position of targets determines the necessary overlap between images, as a reduction in overlap requires an increase in the number of targets, so as to ensure that the prerequisite of three common targets is met. Because the targets are identified algorithmically from the acquired images, they need to be relatively immune to acquisition angle, and adequately sized to sufficiently collect enough sample points, otherwise the available positions for imaging the scene will be limited (Valanis and Tsakiri, 2004; Franaszek *et al.*, 2009). Becerik-Gerber *et al.* (2011) found that spherical targets performed the best when compared against paddle and paper based targets.

Preparing the scene with targets can provide registration alignment with high accuracy, but this approach does have disadvantages. The primary of which is that the targets introduced to the scene may adversely modify its representation due to the targets adding to or obscuring its true shape. The nature of the scene may also make the use of targets infeasible if the scene covers an extensive area, or if access into the scene area is limited or restricted (Guarnieri *et al.*, 2011). The scene preparation step can also be very time consuming, as this is an additional step that requires choosing optimal marker locations, as well as having to actively add and remove targets. An example of this is when using the Creaform portable 3-D scanner (Creaform, 2014), which is an imaging system that requires extensive placement of retro-reflective markers in the scene to achieve registration. The coarse registration pipeline presented in this thesis does not require scene preparation as it registers point-clouds using only their overlapping surface structure.

2.3 3-D Registration Algorithms

Solving the registration problem is not limited to hardware based solutions, as software algorithms are also capable of performing registration. There is an abundance of proposed algorithms which cover a variety of distinct registration approaches. Currently there is no single algorithm suitable for all situations.

This is due to the operating limitations of each algorithm and their particular data requirements. Salvi *et al.* (2007), Blumenthal *et al.* (2011), Castellani and Bartoli (2012), Tam *et al.* (2013) and Díez *et al.* (2015) provide a broad survey of many of these algorithms. Because there is an extensive range of literature available, it is infeasible to review all algorithms. Instead, this section provides a concise review of the most successful algorithms using up-to-date literature.

2.3.1 Iterative Closest Point Algorithm

The iterative closest point (ICP) algorithm (Besl and McKay, 1992; Chen and Medioni, 1991) and its subsequent variants are the predominant set of algorithms used for registering point-clouds whose rotational and translational misalignment is minimal; this form of registration is known as *fine registration*. ICP operates by calculating the transform which minimises the root-mean-squared distance of corresponding points between two point-clouds. By iteratively repeating this process, the transform is revised and the distance minimised. Once the distance is below a prescribed threshold, or the maximum number of allowable iterations is reached, the iterative process is stopped. The disadvantage of ICP is that its iterative process often gets trapped in a local minimum, especially if the initial misalignment between the two point-clouds is large (Castellani and Bartoli, 2012). The overall alignment accuracy is dependent upon a variety of aspects, with initial alignment, point-cloud quality and applicability of algorithm variants having the most influence. Depending on the application, finding a good initial alignment is achieved by acquisition set-up, manual registration, or by a coarse registration algorithm.

The ability of ICP to produce a tight fit between point-clouds (once a good initial alignment has been established) has made it the dominant fine registration algorithm. Due to this success, the majority of subsequent fine registration algorithms follow an iterative scheme akin to ICP (Salvi *et al.*, 2007). Rusinkiewicz and Levoy (2001) reviewed different variants proposed for each step of the ICP algorithm to determine their convergence speed and alignment accuracy. The primary variants found to benefit ICP are the rejection of point-pairs containing boundary points, point-to-plane correspondence instead of point-to-point and avoiding the rejection of point-pairs based on their distance.

Because hundreds of variants have been published since the inception of ICP, Pomerleau *et al.* (2013) constructed a testing framework to compare how different variants perform when applied to real-world data. They found through their framework that improvements to the current ICP variants or new methods are needed to improve ICP when registering natural, unstructured or information-deprived scenes. Larkins *et al.* (2010) ascertained that the errors that occur in range images acquired by an AMCW range imaging system have minimal influence on the ability of ICP.

2.3.2 Key-Point Matching

Registration based on key-point or feature matching operates by using image descriptors that identify aspects (such as shape or texture) in the scene that are recognisable from multiple view-points. A variety of descriptors have been proposed and evaluated in literature (Dutagaci *et al.*, 2012; Tombari *et al.*, 2013). Like scene preparation (see Section 2.2), this approach requires sufficient overlap to allow enough common key-points between overlapping point-clouds to be detected. Once established, the corresponding key-points between point-clouds need to be filtered and matched to allow a rigid transform to be calculated. This matching process can be achieved in different ways; the combination of ICP and the key-points is one such approach (Ekekrantz *et al.*, 2013; Liu *et al.*, 2014) or by constructing a voting procedure to find pairs of key-points that correspond to each other (Kulkarni and Kumar, 2012; Pham *et al.*, 2013). Alternatively, each key-point can be provided with a unique descriptor, such as a spin image (King *et al.*, 2005; Torre-Ferrero *et al.*, 2011; Fantoni *et al.*, 2012) or shape contexts (Price *et al.*, 2012), to reduce the search space for performing the matching. While the matching process is an active area of research, key-point identification is often the primary focus.

Individual range imaging systems provide different types of data that contain a variety of aspects from which key-points can be calculated. Systems capable of acquiring colour or intensity images in conjunction with the range data provide additional means of obtaining key-points (King *et al.*, 2005; Seo *et al.*, 2005). However, as this auxiliary data is not always available or it changes with view-point, key-points are primarily found at structural elements of the point-cloud. Gradient features, such as those calculated by scale invariant feature transform (SIFT) and speeded up robust features (SURF), are detectors adopted from 2-D registration (Houshiar *et al.*, 2013). They can perform particularly well (Bonarrigo *et al.*, 2011; Fantoni *et al.*, 2012; Segundo *et al.*, 2012), but their ability is influenced by sample density. Increasing sample density improves registration success, but computational cost also increases as more key-points are produced that need to be filtered and matched. Hołowko and Sitnik (2013) uses the Harris corner detector which identifies key-points by calculating the shape variation around a vector using the distribution of normals in its local neighbourhood. This detector requires high density point-clouds so that well defined edges and corners are formed, allowing reliable key-points to be produced; because of this it also has a high computational cost. Features such as surface planes (Theiler and Schindler, 2012; Xiao *et al.*, 2012) and object boundaries (Steder *et al.*, 2011) can also be used to produce key-points, but these rely on the scene having a particular structure, and are therefore not always available or reliable.

The advantage of key-point based registration is that if enough key-points are acquired and sufficiently defined, they can accurately align point-clouds. The disadvantage is that the approach used to identify features or key-points often need to be tailored to a number of factors. These include scene structure, data quality, arrangement of sample points and the types of information acquired, while being repeatable with view-point changes. Tombari *et al.* (2013) surveys a selection of current state-of-the-art 3-D key-point detectors to determine their repeatability and distinctiveness under a variety of testing conditions. The performance evaluation provided by Tombari *et al.* confirmed that a detector should be chosen based on the desired task and data quality. Two open issues also highlighted as requiring future research were the efficiency of all existing key-point detectors and improving their robustness to changes in sampling density and scene dimensionality; scene dimensionality relates to point-cloud overlap and the amount of data available about the scene. Using key-point based registration with respect to AMCW imaging systems would be a challenging task due to their sparse point-clouds and potentially degraded quality at object boundaries due to mixed pixels and multipath interference.

2.3.3 Evolutionary Algorithms

Evolutionary algorithms provide a means of robustly solving optimisation problems by drawing inspiration from the process of natural evolution (Eiben and Smith, 2003). Applying evolutionary algorithms to registration allows the desired alignment to be refined over successive generations. Genetic algorithms are the primary type of evolutionary algorithm used for registration but different evolutionary models have been proposed (Santamaría *et al.*, 2011). Genetic algorithms work by encoding the rotation and translation as a string of numbers that represents the genetic make-up of each individual, where each individual is a potential solution (Jacq and Roux, 1995). The initial generation is formed by randomly creating a set of individuals, with the fittest individuals reproducing to form the next generation. The fitness is measured by using an appropriate function, such as the root-mean-squared distance between point-clouds (De Falco *et al.*, 2013). The new generation is produced by combining the genetic make-up of pairs of individuals by using random crossover and by introducing mutations. Once the stopping condition is reached, such as when the allowable number of generations has been produced or fitness has plateaued, then the fittest individual of the last generation is used as the solution.

The primary benefit of using evolutionary algorithms for registering point-clouds is that they develop multiple solutions in parallel. This allows them to perform global optimisation without being trapped in local minima. Because of this, evolutionary algorithms are capable of performing both the coarse and fine registration steps, though their strength is coarse registration. Santamaría *et al.* (2011) performs a comparative study of different algorithms and found that their accuracy varies substantially based on their chosen evolutionary model. This is due to these algorithms needing to be finely tuned to the particular registration problem to perform well. The fine tuning relates to both how the initial population is generated and how the subsequent generations are produced (Santamaría *et al.*, 2013). Because the fitness test is performed for every individual at each generation, computational cost is one of the greatest limitations of using evolutionary algorithms, especially as the number of points in each point-cloud increases (Salvi *et al.*, 2007). The continual development of evolutionary algorithms for registration will reduce the need to tailor an algorithm to each registration problem (Valsecchi *et al.*, 2013). Computational performance can also be aided by using additional approaches, such as sub-sampling (Santamaría *et al.*, 2013) or inertia sensors (Hrgetić and Pribanić, 2013).

2.3.4 Spin Images

Spin images (SIs) were initially introduced by Johnson (1997) to provide a set of rotation and translation invariant 2-D representations of a 3-D point-cloud, which allow for object recognition or point-cloud registration. A complete SI model is produced for a point-cloud by generating an individual SI for every point. The process for generating the SI for a given point, called the orientated point, begins by calculating its surface normal and forming a plane that is orthogonal to the normal and passes through the orientated point. A SI is produced by discretely histogramming all points in the point-cloud, where each point is placed into a bin with respect to the orientated point. The two values that determine the bin are the perpendicular distance from the normal vector to a point and the signed perpendicular distance between the plane and a point. Once two or more complete SI models are created, registration is performed by determining which SIs match each other via correlation and using the method by Horn (1987) to finding the transform which brings them together.

Johnson (1997) described the ability of SIs for registering relatively simple point-clouds, and since their inception, work has been undertaken to improve both their registration and recognition capability. Brusco *et al.* (2005) improved SI registration by histogramming the luminance texture information from the image at different levels, instead of the points. The addition of luminance improved the correct matching of SIs, particularly as overlap decreased, though the overlap limit was not identified. A disadvantage of this approach is that it does not cater for active illumination systems or lighting changes, as it requires that luminance levels remain relatively consistent between views. Zhang *et al.* (2012) instead proposed replacing either axis of the SI with the signed angular difference between the normals of the orientated point and each normal of the neighbouring points. The sign of the angular difference indicates whether the neighbouring normal points towards or away from the normal of the orientated point, which corresponds to the surface being concave or convex, respectively. Zhang *et al.* used this approach to identify and match complete models, and as such, they did not analyse its ability with respect to registration.

The limitation of forming and comparing SI models is the computation cost; as the number of point-clouds or their points increases, more SIs are formed, requiring additional correlation comparisons. Different strategies have been introduced to help alleviate this computational burden by making the comparisons more efficient or by limiting which points produce SIs. These include compressing SIs using principal component analysis (Johnson and Hebert, 1999), point-cloud decimation (Guarnieri *et al.*, 2011) to reduce the number of points, or randomly selecting a subset of points (Zhang *et al.*, 2012). Ho and Gibbins (2008) suggested parametric surface fitting at each point to measure surface shape, allowing a set of salient key-points to be identified at locations with high shape variation. SIs can then be formed at these key-points, minimising the number of SIs that need to be made. While this process may improve overall efficiency, using locations with high shape variations are not always pertinent if the point-clouds are formed primarily from planar surfaces. Dinh and Kropac (2006) varied SI resolution to construct a SI pyramid, SI pairings that had a high enough correlation were promoted to the next resolution, with the rest being discarded. This process was repeated at each resolution from lowest to highest, what remained was a set of SI pairings which had the best correspondence between two point-clouds.

Even with these modifications to improve efficiency, SI based registration is still computationally expensive. This is shown by Guarnieri *et al.* (2011) who used SI pyramids to register 12 point-clouds with a combined count of 3 000 000 points. Each point-cloud was filtered to remove isolated points with this number being further reduced via decimation. The complete SI model for each point-cloud was produced by randomly selecting 10% of the points to construct SIs, with the resulting alignments being refined by ICP. The registration accuracy was not quantified, though it was stated that a minimum of 30% overlap was required. The reported execution time required for the registration process was approximately five hours when using a 64-bit workstation with a quad-core 2.53GHz processor and 6GB of RAM. This reveals that SI based registration still needs further improvements to boost its efficiency.

2.3.5 Random Sample Consensus

Random sample consensus (RANSAC) is an iterative algorithm introduced by Fischler and Bolles (1981) that robustly fits models to data that may be corrupted by both noise and outliers. This is achieved by randomly selecting the minimum number of points needed to rigidly fit the model to the data. The quality of this fit is determined by measuring how many points, called inliers, are within the models threshold distance. The more times this random selection and fitting is performed, the greater the probability of finding a close fit. Once the predefined number of iterations is reached, the model with the most inliers is selected and refined by performing least-squares fitting using only the inliers. A variety of algorithms proposed to improve both efficiency and accuracy are evaluated by Choi *et al.* (2009). Their conclusion was that merging algorithms or investigating RANSAC further is required to better balance its accuracy and efficiency.

Applying RANSAC to the registration of 3-D point-clouds is performed by selecting a set of points from the first point-cloud to be used as a model. A minimum of three points are required to ensure that a rigid transform can be obtained (Chen *et al.*, 1999). The structure of the model is defined by the distance between each of its points, with one of these points being selected as the primary. A potential alignment is then tested by matching the primary point of the model to a point in the second point-cloud. If the structure of the model has a sufficiently close match with this point and its neighbours, then the alignment is stored. This process is repeated for as many points in the second point-cloud as possible. The number of inliers for each alignment is counted as the number of points that correspond between the two point-clouds. The alignment with the greatest number of inliers is selected as the most suitable transform (Fortenbury and Guerra-Filho, 2012). If a suitable transform cannot be found, then an alternative model is needed.

RANSAC is able to provide very robust registration, even in presence of noise and outliers (Fortenbury and Guerra-Filho, 2012). Its primary disadvantage though is that the computational cost of its iterative process can be very expensive, especially as the number of points increases or if an exhaustive search is performed (Salvi *et al.*, 2007). The efficiency of the RANSAC algorithm can be improved by using objects (Yang *et al.*, 2010) or planes segmented from the scene instead of or in conjunction with points to reduce the number of matches that need to be performed (Yao *et al.*, 2011; Taguchi *et al.*, 2013). Alternatively, RANSAC can be used with key-points extracted from the point-clouds instead of only using the initial points (Houshiar *et al.*, 2013; Hołowko and Sitnik, 2013; Ekekrantz *et al.*, 2013). While these modifications can benefit RANSAC, identifying key-points has its own computational overhead, additionally, their use is often scene dependant (see Section 2.3.2).

2.3.6 Principal Component Analysis

Principal component analysis (PCA) is a statistical technique that is able to find patterns in data by identifying the orthogonal axes in this data which correspond to the greatest variance. These axes are the principal components, which are ordered from largest to smallest based on their variance, with there being as many principal components as there are dimensions in the original data (Jolliffe, 2002). The principal components are calculated by first converting the data into a covariance matrix. Applying either eigendecomposition or singular value decomposition to the covariance matrix allows its eigenvalues and eigenvectors to be extracted. Each eigenvector determines the direction of a principal component, while its variance amount is specified by the associated eigenvalue.

Applying PCA to a 3-D point-cloud gives three principal components which correspond to its primary structural shape. Repeating this process for a second point-cloud will give principal components that have the same or similar structure as the first point-cloud if their overlap is sufficient. The two point-clouds are coarsely registered by orientating the principal components into the same alignment from largest to smallest. The primary advantage of PCA is its efficiency, allowing it to rapidly calculate a transform between two point-clouds. However, because the principal components are calculated based on its overall structure, its ability quickly degrade as overlap decreases or if there is an increase in noise or outliers (Salvi *et al.*, 2007). Due to this, PCA is best suited for registering whole objects (Tam *et al.*, 2013). Liu and Ramani (2009) proposed a robust version of PCA which is able to better handle outliers and noise by following a least median of squares based approach. Their approach operates by taking a subset of points that are outlier free and iteratively adding points to this subset that are not considered to be outliers; this subset is deemed the major region, with the minor region comprising of outliers. Calculating the principal components using the major region provides greater registration accuracy, though a high level overlap between each point-cloud's major region is still required to ensure the registration is reliable.

2.3.7 Frequency Domain Correlation

Converting point-clouds or a particular aspect of them into a frequency domain representation permits them to be correlated. Correlation efficiently measures the similarity between two point-clouds using a large range of discrete alignments. For 2-D images, frequency domain correlation has the capability of calculating translation, rotation and scale (Reddy and Chatterji, 1996). Applying correlation to 3-D data also allows rotation and translation to be found; scale is not necessary as range imaging systems produce point-clouds at a consistent scale. A suitable representation in the frequency domain is one that is conducive to the calculated shift. Spherical harmonics provide the most natural representation for finding the rotation as they produce an orthogonal system on the surface of a unit sphere. Translation is calculated using the Fourier series, as they convert data contained in Euclidean space. Because frequency domain correlation is a major component of this thesis, it is reviewed in greater detail in Section 2.4 below.

2.4 Registration in the Frequency Domain

Literature that uses the frequency domain for achieving registration is reviewed in this section, which is divided into two parts. The first focuses on spherical harmonic based approaches that have been used for finding the rotation that orientates two closed objects. The second part reviews how the frequency domain can be used to calculate the rotation and translation that registers partially overlapping point-clouds. Operating in the frequency domain also permits other novel applications to be performed that are not directly pertinent for registration. These include finding the bilateral symmetry plane, which is the best plane for splitting a symmetric object (Kakarala *et al.*, 2013) and 3-D model categorisation and retrieval (Makadia and Daniilidis, 2010).

2.4.1 Orientation of Closed Objects

The rotation which orientates a closed object (an edgeless 2-manifold) with either a ground-truth model or an object of the same type can be calculated by using spherical harmonics to represent the object. This section reviews literature that pertains to the prominent approaches for finding the orientation of closed objects when using spherical harmonics.

Burel and Henocq (1995) proposed the idea of finding the aligning orientation by directly solving for the three rotation angles using spherical harmonics. Their goal of orientating 3-D vertebrae (the bones composing the spinal column) with respect to a model was performed by forming a 3-D voxel matrix of a vertebra and placing its centroid at the origin of the coordinate system. The vertebra's spherical-harmonic representation was produced by mapping it onto the unit sphere by taking the distance between the origin and the farthest point on the vertebra in the direction of the sample point on the sphere. The rotations were found as those that brought the spherical-harmonic coefficients within a predefined set of values. Burel and Henocq identified that if the centroid of the vertebra matched that of the model, then each angle of rotation usually had less than half a degree of error. However, if the centroid locations did not match, then a translational difference occurs, which changed the mapping and increased the rotational error. A spherical-harmonic degree of two was used for finding the rotation, with accuracy being further improved by fine tuning with a degree of four. The spherical-harmonic degree is the degree used to calculate the Legendre polynomial (see Section 3.2.1).

A commonly used approach, referred to as SPHARM (Gerig *et al.*, 2001;

Huang *et al.*, 2005; Shen and Chung, 2007), for representing a 3-D object in spherical harmonics was introduced by Brechbühler *et al.* (1995), in which the vertices of the exterior voxels are mapped onto the surface of the unit sphere. A continuous mapping was achieved by diffusing the surface between vertices, with distortions being minimised by a nonlinear optimisation to preserve the area of the original surface elements. The axes of the first order ellipsoid (FOE) of this mapping were then used to orientate the object into a fixed position by aligning them to the axes of the coordinate system. The limitations of using FOE for registration have been outlined previously (Shen *et al.*, 2007; Gutman *et al.*, 2008), with these being caused by the ellipsoid's axes having similar lengths.

Huang *et al.* (2005) attempted to overcome the limitations of the FOE by minimising the Euclidean distance between two SPHARM surfaces. The distance between the two surfaces is calculated in the frequency domain up to a desired spherical-harmonic degree. This approach is broken into two steps, with the first using brute force to independently calculate the distance for a number of rotations. These rotations were generated by uniformly creating rotation axes and testing with different rotation angles. Having identified the rotation with the minimum distance, the second step used the iterative Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm (Fletcher, 1987) to refine the rotation to further minimise the distance. Shen *et al.* (2007) reduced the number of rotations examined by basing them on icosahedral placements; a hierarchical scheme was then used in which the density of these placements was increased around the best candidates. The BFGS algorithm was also dropped in preference for using an ICP based approach to complete the registration. The results of Shen *et al.* were shown using a spherical-harmonic degree of 12, and were quantified using the Euclidean distance between corresponding surface parts. While this approach was shown to provide a better alignment than FOE, its efficiency is still its primary limitation as the distance is independently calculated for each rotation, with this cost only increasing if higher spherical-harmonic degrees are desired.

Gutman *et al.* (2008) showed that both FOE and the approach by Shen *et al.* (2007) are special cases of spherical-harmonic correlation, and as such, correlation will provide better registration accuracy. Gutman *et al.* initialised the correlation by mapping a 3-D images of the hippocampus (the memory area of the brain) onto the unit sphere using the approach by Gu *et al.* (2004), before being converted to the frequency domain at a bandwidth of 64. Performing the correlation and inverting it into the impulse domain allowed the rotation which maximises the alignment between hippocampi to be identified. Gutman *et al.* attempted to reduce the cost of the correlation by expressing the rotation in the frequency domain as the product of two rotations, allowing β, the rotation angle about the y-axis, to be fixed as $\pi/2$. The idea behind this elegant approach was that the Wigner d-matrix only needs to be calculated and cached once for each spherical-harmonic degree and order instead of being recalculated for each β. While Ritchie and Kemp (1999) found this caching scheme to be slower than the direct evaluation, the stated computational cost of $\mathrm{O}\big(B^4\big)$ is the same as the efficient algorithm proposed by Kostelec and Rockmore (2008).

Rotation estimation between two objects using a closed-form solution was proposed by Althloothi *et al.* (2013), in which the separating rotation is identified as the one which minimises the difference between the two sets of spherical-harmonic coefficients. This was achieved by first expressing objects on the unit sphere using a modified version of the approach by Shen and Makedon (2006), allowing objects with a torus like nature (objects of genus-1 and greater) to also be represented. This modification works by performing hole-filling, in which the objects are made to be genus-0. This reconstruction, however, can misconstrue the representation, such as when filling in the handle of a coffee mug. Suboptimal representations of an object can be detrimental as it limits the ability of the algorithm to accurately determine the correct rotation. By performing eigendecomposition on the cross-covariance matrix of the spherical-harmonic coefficients, the eigenvector of the maximum eigenvalue allowed the rotation which minimises the rotation error to be estimated. This approach was tested using a degree of 15 to calculate the spherical harmonics of both an object and the rotated version of itself. The results showed that as the rotational misalignment between the two models increased, the rotation error also increased, with this increase rapidly rising when misalignment was 80° or more. While this closed-form solution can perform well, the increase in rotation error indicates that its capability is limited; rotational misalignment should not affect the resulting rotation error when spherical harmonics are used, as they can handle any rotational difference.

2.4.2 Orientation of Surfaces

The registration capability of spherical-harmonic correlation permits it to be expanded beyond solely using it for closed objects, though currently there has been minimal research that examines it for partially overlapping point-clouds.

Makadia *et al.* (2006) provided the first and primary piece of literature that applies spherical-harmonic correlation to this task. The approaches used for representing an object's orientation on a unit sphere do not intuitively apply to a point-cloud that is open, as opposed to being fully closed. Therefore, an alternative means of representing surface orientation on the unit sphere is used. This representation is achieved by extracting the surface normals from the point-cloud to form a Gaussian sphere, which is a unit sphere with points on its surface indicating the orientation of each normal. Makadia *et al.* (2006) refers to these Gaussian spheres as extend Gaussian images (EGI) (Horn, 1984), but this is a misnomer as the formed Gaussian spheres do not take into account the surface area associated with each normal when projecting them onto the sphere. The advantage of using surface normals is that their relationship with each other is invariant to both rotation and translation.

Converting the Gaussian sphere into spherical harmonics first requires that a discrete representation of the sphere surface be produced to allow a computer implementation to transform the normals into the frequency domain. Makadia *et al.* (2006) achieved this by uniformly sampling both the polar and azimuthal coordinate directions using 256 bins. This binning procedure is known as equiangle binning, which efficiently places each normal into a bin, with these bins being collectively referred to as an *orientation histogram*. It was stated that the best orientation histogram would be comprised of bins which have the same surface area and shape, but this was not focussed on due to their requirement for fast correlation. Makadia *et al.* (2006) did not examine how different binning schemes or bin sizes for collecting normals would affect registration performance.

Wang and Qian (2008) proposed an alternative means of producing the Gaussian sphere, in which a range image is converted into a 2-D phase-encoded map and then transformed into the Fourier domain. An EGI was then produced by performing a coordinate transform from the frequency domain into spherical coordinates. While this is a novel approach, it is less efficient than directly binning normals and it can only operate on a range image acquired from a single pose of a range-imaging system at a time; therefore, this procedure is not applicable to all systems, such as those that use a mechanical arm. This procedure also requires a uniform sample distribution; non-uniform sampling can be made uniform via interpolation, but this adds to the computational overhead.

Makadia *et al.* (2006) introduced an additional step for converting the orientation histogram into a *constellation image*, by setting all bins that have a *dominant peak* to a value of one, and the rest to zero. However, it was not stated what criteria determines a dominant peak, though it is assumed that it is based on each bin collecting a sufficient number of normals. The advantage provided by a constellation image is that each surface region is treated as being equally sized, improving the likelihood that the same surface region from different views will be matched together. This is important as the same surface region may have different sampling densities, or portions of it may be outside the field-of-view when imaged from different poses.

The registration pipeline used by Makadia *et al.* (2006) first identified the rotation between two point-clouds by converting their formed constellation images into spherical harmonics using the algorithm by Driscoll and Healy (1994). A bandwidth of 128 was used for this conversion process, which corresponds to a 256×256 histogram; the bandwidth specifies the upper limit for the calculation of spherical harmonics (see Section 3.2.2). Correlating the two sets of spherical-harmonic coefficients at this bandwidth and inverting the correlation into the impulse domain allowed 2 097 152 different rotations to be efficiently tested simultaneously. The size of each resulting impulse provided a measure of how closely the two constellation images match each other at the associated rotation. The largest impulse response corresponds to the rotation that brings the two point-clouds into rotational alignment, if they are conducive to correlation. Calculating the registration in this manner decouples the rotation from the translation, allowing them to be calculated independently. The mathematics behind both the rotation and translation are outlined with greater depth in Chapter 3.

Calculating the rotation and translation independently means that these two steps only have three degrees of freedom each, instead of the original six. Because the translation does not accommodate rotational misalignment, it must be calculated after the rotation, otherwise the found translation will be incorrect. Makadia *et al.* (2006) calculated the translation in a manner similar to the rotation, except the points of the point-cloud were instead binned and converted into Fourier coefficients. A process similar to the rotation was performed, with the coefficients being correlated in the frequency domain and inverted to the impulse domain. The translation corresponding to the largest impulse was then selected and used to complete the registration.

In order to automate this pipeline Makadia *et al.* incorporated two alignment verification strategies (see Section 2.5). If the found alignment was classified as incorrect, then the rotation associated with next largest impulse was applied and the translation and verification steps reperformed. The number of available rotations limits the effectiveness of this approach, as the next selected rotation can be very similar to one previously selected. This pipeline was shown capable of coarsely registering point-clouds with overlap as low as 45%. However, as this pipeline was formed to show that spherical-harmonic correlation can be applied to partially overlapping point-clouds, the capabilities and limitations of each step were not evaluated in detail. The analysis performed throughout this dissertation characterises each step with much greater depth.

In literature there have been alternative pipelines proposed which use techniques similar to those presented by Makadia *et al.* (2006). Buchholz *et al.* (2012) collected normals from a scene by acquiring an image using structured lighting. An EGI was formed from these normals by using equiangle binning and taking into account the area of their corresponding face to overcome changes in sampling density. Weighting the normals in this fashion does not take into account how the visible area of each surface region changes with different imaging poses. The orientation of the image was then found via spherical-harmonic correlation with respect to a known model. Having rotated the image, the model was projected onto a plane from the image's acquisition pose, 2-D scale invariant image correlation was then performed to complete the registration, allowing the acquired image to be compared against the model. This approach for finding the scale and translation was used as the acquisition process used by Buchholz *et al.* did not provide depth data for each sample point.

The pipeline by Bülow and Birk (2013) avoided using spherical harmonics for calculating the rotation between two point-clouds by instead resampling them into structures which allowed the yaw to be determined separately from the roll and pitch. The yaw was determined by representing two point-clouds as 2-D images using a combination of spherical coordinates at different radii and polar resampling. The shift between these images, found by a phase correlation, then corresponded to the yaw angle. The roll and pitch were determined by projecting the spherical coordinates of each point-cloud at a given radii onto a plane, which was discretised into an image. Once again applying a 2-D phase correlation to these images produced a shift, in which the x and y translations corresponded to pitch and roll, respectively. The translational alignment was then found by a 3-D phase correlation of the voxelised points. This pipeline was shown to efficiently register point-clouds with overlap as low as 29%. However, the nature of the used algorithm limits both the roll and pitch angles to ±35° each, it is therefore unable to register all imaging poses, restricting the applications to which it can be applied.

2.5 Verification Strategies

The 3-D registration algorithms reviewed above do not guarantee correct alignment. Incorporating a verification strategy into the registration pipeline provides a means of confirming whether the found point-cloud alignment is correct or not. This is achieved by testing whether the alignment abides by predefined criteria which stipulate how two or more point-clouds should relate to each other. If the found alignment does not meet the criteria, then this indicates that an alternative alignment should be sought. The simplest strategy is to measure the root-mean-squared distance between point-clouds, with this distance pertaining to the accuracy of the alignment (Dorai *et al.*, 1998). If this distance is below a given threshold, such as the noise limit of the sensor, then this alignment can be considered correct (King *et al.*, 2005). While this strategy provides a form of verification that indicates correct alignment, incorrect alignments can still be classified as correct if their error is within the threshold. Applying different or more complex strategies to the pipeline help limit an incorrect alignment being chosen.

When producing a registration pipeline, the inclusion of a verification strategy is often overlooked. Therefore, there has been a minimal number of verification strategies proposed and investigated in literature. One strategy is to maintain the line-of-sight between the imaging system and the acquired point-cloud. If the registration brings a point-cloud into an alignment that obscures this line-of-sight, then this alignment cannot occur and is classified as incorrect (Huber and Hebert, 2003; King *et al.*, 2005; Makadia *et al.*, 2006). This strategy can work well, but it is not appropriate for all registration problems as it requires an imaging system with a fixed sensor position. If the point-cloud is constructed by a range sensor that moves, such as those that are on a mechanical arm, then the line-of-sight changes and it is unknown whether it has been obscured. The allowable separation between overlapping point-clouds (before they are deemed obscured) needs to account for noise level, and the amount of error that remains if only coarse registration is performed.

Makadia *et al.* (2006) presented a strategy which verifies an alignment by establishing whether the surface orientations between overlapping point-cloud regions are consistent. This was based on the assumption that the normals of points in the same vicinity have the same orientation. Makadia *et al.* did not explore the accuracy of this strategy, but as voxelisation is needed to measure the difference in orientation, there is a limit to its classification ability. This is because each given voxel can collect normals that correspond to more than one surface causing their mean orientation to be shifted. When using this strategy, the allowable orientation error needs to account for these shifts, as well as any noise or alignment error if used with a coarse registration algorithm.

The transform transitivity between three or more point-clouds has previously been used for improving their overall registration accuracy (Pooja and Govindu, 2010). The transitive relation stipulates that when an element in a set is related to a second element it is also related to a third element if the third element is related to the second element. The transitive nature of the transforms between point-clouds also has the potential of being used as a verification strategy by determining whether the found transforms are consistent with each other. Both the orientation of normals and transform transitivity are examined in Chapter 8. The lack of proposed verification strategies limits the ability to automatically determine if correct alignment has been achieved. Further investigation into this step of the pipeline will help ensure that the found alignment is the one that is desired.

Chapter 3

Mathematical Preliminaries

The points of a point-cloud and their associated normals are formed in the spatial domain, which is the coordinate system specifying their individual locations. The coarse registration pipeline analysed in this dissertation operates primarily in the frequency domain. The frequency domain provides an alternative representation of the spatial domain by measuring how many points or normals occur within a particular frequency band at a range of frequency rates. Cross-correlating two sets of points or normals in the spatial domain is performed by representing and multiplying them together in the frequency domain; the correlation measures their similarity when one set is discretely shifted relative to the other. Inverting the correlation to the impulse domain represents each shift as an impulse, where the position of an impulse corresponds to a specific shift between the two sets of points or normals within the spatial domain. The size of an impulse is the measure of their similarity at this shift.

This chapter outlines the mathematics used to transform between the respective domains when calculating the rotation and translation that aligns two overlapping point-clouds. The application of the mathematics is necessary as the coordinate system of each acquired point-cloud is attached to the range imaging system. Specifically, the origin of the local coordinate system of an acquired point-cloud is located at the imaging system's sensor. Therefore, simply placing these point-clouds into the same coordinate system does not bring them into alignment, but will haphazardly place them all together within the same field-of-view of the imaging system. The rotation and translation used for alignment are determined independently, but both follow a sequence of steps similar to those initially outlined.

3.1 Surface Normals

When acquiring a point-cloud of an object or scene, the surface orientation at each sample point is represented by its surface normal. A normal is a unit vector associated with a point and perpendicular to the surface. Because each point is only a location in 3-D Cartesian space, its normal is not inherently given, instead the normal is calculated using the relationship between the sampled point and its neighbours (Klasing *et al.*, 2009). If an imaging system samples the scene in a uniform manner, then the neighbouring sample points are intrinsically given. However, if the imaging system does not sample the scene in a uniform manner, then the relationship between points can be determined algorithmically (Newman and Yi, 2006; Isenburg *et al.*, 2006).

The procedure for bringing two overlapping point-clouds into coarse rotational alignment by way of spherical-harmonic correlation is reliant upon the calculated normals. The normals maintain the same directional relationship relative to each other because the shape of an object (to which the normals pertain) is invariant to the coordinate system, even when the coordinate system changes with imaging system pose; it must be noted that in practice, sample point noise and changes in sampling distribution between acquisitions can affect this relationship. Extracting the normals from a point-cloud and placing them at the origin of the coordinate system means they are no longer associated with a point and can instead be thought of as identifying a location on a unit sphere. This coordinate space on the sphere is defined as

$$\mathbb{S}^2 = \left\{x \in \mathbb{R}^3 : ||x|| = 1\right\} \tag{3.1}$$

and is known as a Gaussian sphere (Horn, 1984).

A point-cloud's normals can be calculated using a number of algorithms (Jin *et al.*, 2005; Klasing *et al.*, 2009). The choice of algorithm will affect the rotational alignment, but this aspect is not examined as part of this research.

3.2 Rotation Registration

This section details the process involved for calculating the rotation which provides coarse rotational alignment. This is achieved by converting two sets of surface normals extracted from overlapping point-clouds into the frequency domain. These two sets are then correlated and inverted to give the rotation which maximises their rotational similarity.

3.2.1 Spherical Harmonics

Spherical harmonics arise as the angular solutions to partial differential equations when these equations are expressed in spherical coordinates and solved by the separation of variables method. The Laplace, Helmholtz and Schrödinger equations are such examples (Arfken *et al.*, 2013). The usefulness of spherical harmonics in the context of range-image registration comes from their capability to provide an orthonormal system that allows a function formed on the surface of a unit sphere to be represented in the frequency domain (Driscoll and Healy, 1994). Spherical harmonics are given in the form

$$Y_l^m(\theta,\,\phi) = \begin{cases} \sqrt{\dfrac{2l+1}{4\pi}\dfrac{(l-m)!}{(l+m)!}}\,P_l^m(\cos\theta)e^{im\phi} & m \geq 0, \\ (-1)^m\overline{Y_l^{|m|}}(\theta,\,\phi) & m < 0, \end{cases} \tag{3.2}$$

where i is the imaginary unit, $\bar{x}$ is the complex conjugate of x, and l and m, the respective degree and order, are defined as

$$l \in \mathbb{N}, \tag{3.3}$$

and

$$-l \leq m \leq l, \quad m \in \mathbb{Z}. \tag{3.4}$$

The P_l^m component is the Legendre polynomial of degree l and order m expressed as

$$P_l^m(x) = \frac{(-1)^m}{2^l l!}(1-x^2)^{\frac{m}{2}}\frac{d^{l+m}}{dx^{l+m}}(x^2-1)^l. \tag{3.5}$$

Visual representations of the spherical harmonics up to degree three are constructed by sampling across the entire sphere surface; the squared real components are presented in Figure 3.1, while the magnitude of the spherical-harmonic complex values are shown in Figure 3.2.

3.2.2 Spherical-Harmonic Transform

Correlating two sets of normals in the Frequency domain requires their spherical-harmonic representation. The spherical-harmonic transform (SHT) provides this conversion from the spatial domain to the frequency domain, though it usually samples a continuous function on the unit sphere. While normals have discrete locations, this issue is circumvented by binning the normals (see Chapter 5), with each bin being formed around the SHT sample points.

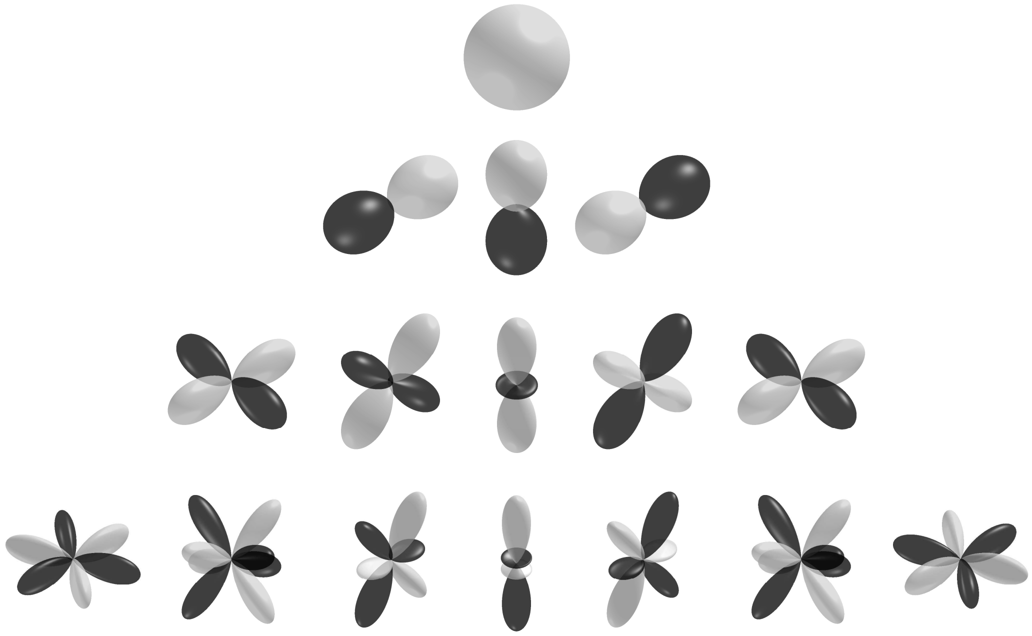


Figure 3.1: Visualisation of the squared real components of the spherical harmonics up to the third degree; expressed mathematically as $\mathrm{Re}\left(Y_l^m(\theta,\phi)\right)^2$. Each row shows a particular degree of a spherical harmonic, along with all of its orders. The lighter colouring is where the spherical-harmonic function is positive, and the darker colouring is where the function is negative.

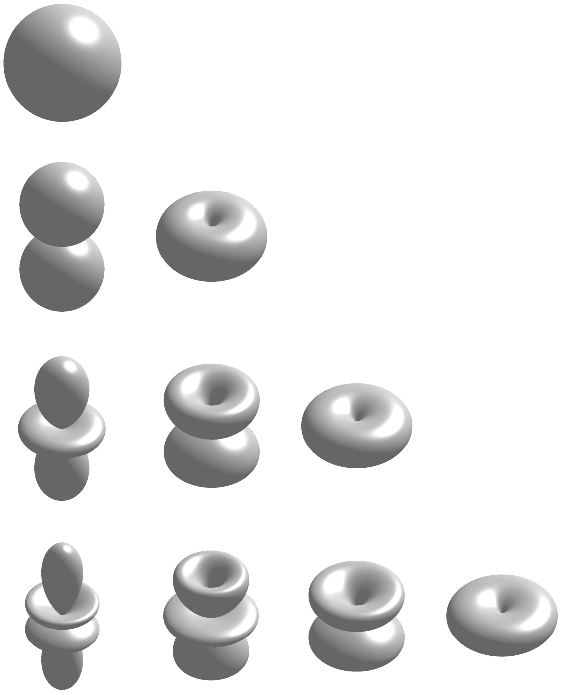


Figure 3.2: Visualisation of the magnitude of the spherical harmonics up to the third degree. Each row shows a particular degree of a spherical harmonic, along with all of the positive orders.

Performing the transform requires that a bandwidth, B_t, specify the upper limit of the calculated frequencies, in which B_t is a positive integer. For these frequencies to represent the function defined on the unit sphere, the surface must be sampled (i.e. discretised) in a sufficient manner. Any number of sample points can be used for the transform to the frequency domain, but if an insufficient number of samples are used, then ambiguous aliasing of the harmonics will occur. The sufficient sample count is related to the bandwidth and is dependent upon the chosen SHT algorithm. There are a variety of SHT algorithms, two of which are used in this work; these are the fast SHT by Healy *et al.* (2003) and least-squares regression, both of which are described below. The difference in their efficiency is then outlined.

Fast Spherical-Harmonic Transform

The fast SHT is one of the most efficient SHT algorithms, however it is limited to equiangular sampling. This limitation is one aspect that permits the fast SHT to be fast as it allows the integrals defining the spherical-harmonic coefficients to by efficiently evaluated by the weighted sums of the samples. Additional improvements to efficiency come from the manner in which the base mathematics are implemented, as the Legendre polynomials can be decomposed into smaller sub-problems and solved recursively. The work by Driscoll and Healy (1994) and subsequently Healy *et al.* (2003) describe in detail the efficient mathematical derivations of the fast SHT; these derivations are not presented here as they contain many nuances that unnecessarily complicate the algorithm description. Instead, the algorithm is expressed in its discrete form, allowing its key components to be focussed upon.

The initial step is to sample the unit sphere, as it is the value at these sample points that are converted to the frequency domain. The fast SHT samples points on the unit sphere in an equiangular fashion, with their locations being given in spherical coordinates. Sufficient sampling is achieved when $4B_t^2$ sample points are distributed about the sphere, with the placement of these points specified by

$$\theta_j = \frac{(2j+1)\pi}{4B_t} \qquad j = 0, 1, \ldots, 2B_t - 1, \tag{3.6}$$

and

$$\phi_k = \frac{k\pi}{B_t} \qquad k = 0, 1, \ldots, 2B_t - 1. \tag{3.7}$$

As the sampling has greater density at the poles, the transform needs to accommodate this change in sampling distribution. This is achieved by applying Chebychev weights to the transform which are analogous to the $\sin(\theta)$ factor present in the integral depiction of the transform (Healy *et al.*, 2003). The Chebychev weights are given as

$$w_j^{(B_t)} = \frac{2\sin\theta_j}{B_t}\sum_{k=0}^{B_t-1}\frac{\sin\left((2k+1)\theta_j\right)}{2k+1}, \tag{3.8}$$

and are only dependent upon the transform bandwidth and the polar angle position as specified by j.

Sampling at the specified locations gives the value attributed to the function defined on the sphere at that point. If the function f (which can be complex) is given in the spatial domain, its representation in the frequency domain is $\hat{f}$, which is the set of spherical-harmonic coefficients. Having now outlined these components, the formula for the transform at each degree $l < B_t$ and order $|m| \leq l$ is

$$\hat{f}_l^m = \frac{\sqrt{\pi}}{B_t}\sum_{j=0}^{2B_t-1}\sum_{k=0}^{2B_t-1} w_j^{(B_t)} f(\theta_j,\phi_k)\overline{Y_l^m}(\theta_j,\phi_k), \tag{3.9}$$

in which $\sqrt{\pi}/B_t$ is simply a scaling factor, that is unnecessary for the following spherical-harmonic correlation.

Inverting from the frequency domain back to the original form of the function in the spatial domain is unnecessary for identifying the rotation which maximises the correlation. It is provided here purely for completeness, and is given as

$$f(\theta_j,\phi_k) = \sqrt{\pi}\sum_{l=0}^{B_t-1}\sum_{m=-l}^{l}\hat{f}_l^m Y_l^m(\theta_j,\phi_k). \tag{3.10}$$

Least-Squares

Least-squares regression is one approach that has been commonly used to transform a set of points sampled on a sphere into the frequency domain (Blais and Soofi, 2006; Brechbühler *et al.*, 1995). While least-squares is a computationally expensive SHT algorithm, it is capable of transforming any distribution of sample points. This is achieved by constructing the matrix $\mathbf{B}$ which contains the spherical-harmonic basis functions in the form

$$\mathbf{B}_{s,k} = Y_l^m(\theta_s,\phi_s), \tag{3.11}$$

where each row and column is indexed by

$$s = 0, 1, \ldots, n - 1 \tag{3.12}$$

and

$$k = l(l + 1) + m, \tag{3.13}$$

respectively, in which n is the total number of sample points. The value at each sample point is stored in the vector $\mathbf{x}$ which is indexed by s. The least-squares equation is then constructed as

$$\mathbf{c} = (\mathbf{B}^{\intercal}\mathbf{B})^{-1}\mathbf{B}^{\intercal}\mathbf{x}. \tag{3.14}$$

The vector $\mathbf{c}$ then contains the spherical-harmonic coefficients. If however, $\mathbf{B}^{\intercal}\mathbf{B}$ is ill-conditioned or singular, then it cannot be inverted and the least-squares solution cannot be determined.

Least-squares via singular value decomposition (SVD) is able to determine a solution even if it cannot be found by the prior approach. This is achieved by expressing $\mathbf{B}$ in its SVD form, given as

$$\mathbf{B} = \mathbf{U}\mathbf{S}\mathbf{V}^{\intercal}, \tag{3.15}$$

when the null space has been removed (making $\mathbf{S}$ a square diagonal matrix). Least-squares is then calculated by rearranging the SVD equation to

$$\mathbf{c} = \mathbf{V}\mathbf{S}^{-1}\mathbf{U}^{\intercal}\mathbf{x}. \tag{3.16}$$

The fast SHT and least-squares produce the same spherical-harmonic coefficients when using the same equiangle sample points and corresponding values.

Computational Efficiency

The computational cost of performing the fast SHT is significantly less than the least-squares approach, both in the necessary processing and the required storage. The work presented by Driscoll and Healy (1994) states that the cost of performing fast SHT is dependent upon the specified bandwidth B_t. The computational cost for transforming the sample points to the frequency domain is then $\mathrm{O}\left(4B_t^2 \log^2\left(4B_t^2\right)\right)$. An array of size B_t^2 is then needed to store the spherical-harmonic coefficients.

Computing least-squares by way of SVD requires that the matrix $\mathbf{B}$, which is of size nB_t^2, be decomposed into **U**, **S** and **V**; if $n = 4B_t^2$ then the size of **B** is $4B_t^4$. The combination of **U**, **S** and **V** has the size $n^2 + nB_t^2 + B_t^4$. The total computational cost is difficult to measure due to the SVD, the matrix inversion of **S** and the two matrix multiplications.

Because there is a significant difference in the computational cost between the fast SHT and least-squares, least-squares is unable to perform at the same bandwidth as the fast SHT for any given computer system. While there are other approaches for calculating least-squares regression, SVD is the most robust. Least-squares is more versatile than the fast SHT as it handles any distribution of sampling points; the benefit of this versatility with respect to registration performance is compared with the fast SHT in Chapter 5.

3.2.3 Spherical-Harmonic Correlation

Sampling two sets of normals, given as f and g, and converting them to the frequency domain by way of SHT gives $\hat{f}$ and $\hat{g}$, respectively. Correlating these two sets of spherical-harmonic coefficients is the next step, which is achieved by

$$\hat{h}^l_{m\,m'} = \hat{f}_l^{-m}\overline{\hat{g}_l^{-m'}}, \tag{3.17}$$

in which $\hat{h}$ is the set of correlated spherical harmonics, l is the degree and $-l \leq m, m' \leq l$ are the orders. Because every order between $\hat{f}$ and $\hat{g}$ are correlated, $\hat{h}$ has a size of $(4B_t^3 - B_t)/3$, which is greater than B_t^2, the size of either $\hat{f}$ or $\hat{g}$.

3.2.4 Correlation Inversion

Having produced the correlation $\hat{h}$ in the frequency domain, $\hat{h}$ is then inverted into the rotation domain as specified by the rotation group $SO(3)$. The bandwidth used for the forward transform is termed the transform bandwidth, but it is not necessary for the inverse transform. Instead, the bandwidth, B_c, used for the inverse, is termed the correlation bandwidth and is specified as

$$B_c = 0, 1, \ldots, B_t. \tag{3.18}$$

The storage requirements needed for the inverse transform is greater than the forward SHT when B_t and B_c are the same. Therefore, using a transform bandwidth larger than the correlation bandwidth is beneficial; these benefits are covered in detail in Section 3.4 below.

The inversion to the rotation domain uses the Wigner D-matrix and its subsidiary the Wigner (small) d-matrix. The Wigner D-matrix is the frequency domain equivalent to the rotation matrix in the spatial domain. The Wigner D-matrix is given as

$$D^l_{m\,m'}(\alpha,\beta,\gamma) = e^{-i(m\alpha+m'\gamma)} d^l_{mm'}(\beta) \tag{3.19}$$

and the Wigner d-matrix as

$$\begin{aligned} d^l_{m\,m'}(\beta) = \sum_{t=\max(0,\,m'-m)}^{\min(l+m',\,l-m)} (-1)^t \frac{\sqrt{(l+m')!(l-m')!(l+m)!(l-m)!}}{(l+m'-t)!(l-m-t)!(t+m-m')!t!} \\ \times \left(\cos\frac{\beta}{2}\right)^{2l+m'-m-2t} \left(\sin\frac{\beta}{2}\right)^{2t+m-m'}. \end{aligned} \tag{3.20}$$

The rotation angles α, β and γ correspond to a ZYZ rotation matrix in the spatial domain (see Section 3.2.5). The correlation value of a particular set of rotation angles is given by

$$h(\alpha,\beta,\gamma) = \sum_{l=0}^{B_c-1} \sum_{m=-l}^{l} \sum_{m'=-l}^{l} \hat{h}^l_{m\,m'} \overline{D^l_{m\,m'}}(\alpha,\beta,\gamma). \tag{3.21}$$

Typically, the closer the tested rotation is to the desired rotation, the greater the correlation value. The work by Kostelec and Rockmore (2008) provides an efficient approach for testing a discrete set of rotations that encompass the entire rotation domain; therefore, it is not necessary to test every rotation (this is infeasible as rotations angles are real numbers). The rotation angles used for each rotation from the discrete set are calculated as

$$\alpha = \frac{\pi a_1}{B_c}, \tag{3.22}$$

$$\beta = \frac{\pi(2b_1+1)}{4B_c}, \tag{3.23}$$

$$\gamma = \frac{\pi c_1}{B_c}, \tag{3.24}$$

where

$$a_1, b_1, c_1 = 0, 1, \ldots, 2B_c - 1, \tag{3.25}$$

which restricts the rotation angles to

$$0 \leq \alpha,\, \gamma < 2\pi, \tag{3.26}$$

and

$$0 \leq \beta \leq \pi. \tag{3.27}$$

The inverse transform has a processing cost of $\mathrm{O}\left(B_c^4\right)$, with this size being the combined total of the $8B_c^3$ tested rotations. The correlation value of each of these rotations is stored in the rotation correlation matrix, $\mathbf{C}_R$, which is indexed by a_1, b_1 and c_1. This correlation matrix has a storage cost equal to the total rotation count, which is $8B_c^3$.

3.2.5 *ZYZ* Rotation Matrix

Identifying the rotation which provides coarse alignment is simply a case of locating the element in the rotation correlation matrix, $\mathbf{C}_R$, that has the largest value. The position of this element specifies a_1, b_1, and c_1, each of which correspond to an individual axis. Having determined a_1, b_1 and c_1, the rotation angles are given by inserting them into equations (3.22), (3.23) and (3.24), respectively. These rotation angles correspond to a post-multiplication *ZYZ* rotation matrix for use in a right-hand coordinate system, where all three rotations are counter-clockwise (when viewing along the axis towards the origin).

If each rotation angle is dealt with individually, the rotation matrix which they each form are given as

$$\mathbf{R}_Z(\alpha) = \begin{bmatrix} \cos\alpha & \sin\alpha & 0 \\ -\sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.28}$$

$$\mathbf{R}_Y(\beta) = \begin{bmatrix} \cos\beta & 0 & -\sin\beta \\ 0 & 1 & 0 \\ \sin\beta & 0 & \cos\beta \end{bmatrix}, \tag{3.29}$$

$$\mathbf{R}_Z(\gamma) = \begin{bmatrix} \cos\gamma & \sin\gamma & 0 \\ -\sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.30}$$

Multiplying these three rotation matrices together gives the complete *ZYZ* rotation matrix as

$$\mathbf{R}_{ZYZ}(\alpha,\beta,\gamma) = \mathbf{R}_Z(\alpha)\mathbf{R}_Y(\beta)\mathbf{R}_Z(\gamma) =$$
$$\begin{bmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & \sin\alpha\cos\gamma + \cos\alpha\cos\beta\sin\gamma & -\cos\alpha\sin\beta \\ -\cos\alpha\sin\gamma - \sin\alpha\cos\beta\cos\gamma & \cos\alpha\cos\gamma - \sin\alpha\cos\beta\sin\gamma & \sin\alpha\sin\beta \\ \sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{bmatrix}. \tag{3.31}$$

The two sets of sample points, f and g, are produced from the two sets of normals extracted from the overlapping point-clouds. Applying this rotation matrix to g will (in suitable situations) bring g into coarse rotational alignment with f. Subsequently applying this rotation to the points from which g was extracted will bring them into coarse rotational alignment with the points from which f was extracted. Having achieved the desired coarse rotational alignment, the coarse translational alignment can be obtained, completing the coarse registration between these two point-clouds.

3.3 Translation Registration

The rotational alignment achieved in the previous section does not complete the registration, as translational alignment is also needed. The coarse translational shift which aligns the two point-clouds is calculated using a three-dimensional phase correlation in the Fourier domain. If the two point-clouds do not have correct rotational alignment, the accuracy of the translation will be limited as the translation correlation simply attempts to maximise the translational similarity between two point-clouds. It is for this reason that the translation is performed subsequent to the rotation.

The translation registration goes through a similar sequence of steps as the rotation registration. These are binning, transformation to the frequency domain, correlation, inversion and applying the identified translation. These steps are described in detail below.

Voxelisation

The two point-clouds in their initial form are continuous in nature, but for a computer implementation to transform them into the frequency domain, a discrete representation is required. Converting a point-cloud into a discrete representation is achieved by segmenting the Cartesian coordinate space in a process called voxelisation. To ensure that the occurrence of redundant empty space is minimised, each point-cloud is first translated so that their centroid is placed at the origin of the coordinate system. The voxelisation process is then carried out by centring a cube at the origin of the coordinate system and making it large enough to encompass all point of both point-clouds. Because there is no scale difference between the two point-clouds, it is best to use the same sized cube for both. The length of each side of the cube is given as ℓ, which is determined as twice the maximum x, y or z point coordinate within the point-cloud. This length is then equally divided by v voxels, with each voxel acting as a bin.

The position of each of the v^3 voxels are specified by the integers a_2, b_2 and c_2, which respectively correspond to the x, y and z axes. The voxel that collects a given point, $\mathbf{p}$, is determined by

$$a_2 = \left\lfloor v\left(\frac{\mathbf{p}_x}{\ell} + \frac{1}{2}\right)\right\rfloor, \tag{3.32}$$

$$b_2 = \left\lfloor v\left(\frac{\mathbf{p}_y}{\ell} + \frac{1}{2}\right)\right\rfloor, \tag{3.33}$$

$$c_2 = \left\lfloor v\left(\frac{\mathbf{p}_z}{\ell} + \frac{1}{2}\right)\right\rfloor. \tag{3.34}$$

A minor issue arises if $\mathbf{p}$ is on a positive edge of the cube, as this situation combined with these equations causes one or more of the position values to be set to v, which is undesired as

$$a_2, b_2, c_2 = 0, 1, \ldots, v-1. \tag{3.35}$$

This situation is resolved by setting the position value of $\mathbf{p}$ to $v-1$. Having discretised the two point-clouds into three-dimensional histograms, they are correlated, allowing the translation which maximises the correlation to be determined.

Translation Correlation

The translation correlation is carried out by transforming the two point-cloud histograms into the Fourier domain. Let two histograms of points be $\mathbf{F}$ and $\mathbf{G}$, and their Fourier representations be $\hat{\mathbf{F}}$ and $\hat{\mathbf{G}}$, respectively. The correlation, known as a phase correlation, is given by

$$\hat{\mathbf{C}}_T = \frac{\hat{\mathbf{F}} \circ \overline{\hat{\mathbf{G}}}}{\|\hat{\mathbf{F}} \circ \overline{\hat{\mathbf{G}}}\|}, \tag{3.36}$$

where the matrix $\hat{\mathbf{C}}_T$ is the normalised correlation in the frequency domain and $\mathbf{F} \circ \mathbf{G}$ is the Hadamard product or entry-wise product between two matrices. Inverting $\hat{\mathbf{C}}_T$ gives the normalised correlation in the spatial domain as $\mathbf{C}_T$. The position of the maximum impulse response is once again indexed by a_2, b_2 and c_2, which identifies the coarse translation which maximises the translational similarity between the two point-clouds. The translational shift is performed along each axis, where a_2, b_2, and c_2 correspond to the x, y, and z directions, respectively, and are found as,

$$x = \begin{cases} a_2\ell/v & a_2 \leq \lfloor v/2 \rfloor \\ (a_2 - v)\ell/v & \text{otherwise} \end{cases}, \tag{3.37}$$

$$y = \begin{cases} b_2\ell/v & b_2 \leq \lfloor v/2 \rfloor \\ (b_2 - v)\ell/v & \text{otherwise} \end{cases}, \tag{3.38}$$

$$z = \begin{cases} c_2\ell/v & c_2 \leq \lfloor v/2 \rfloor \\ (c_2 - v)\ell/v & \text{otherwise} \end{cases}, \tag{3.39}$$

where ℓ/v gives the side length of each individual voxel.

Increasing the number of voxels or decreasing the side length improves the precision of the resulting translational alignment, as the sampling density will increase. Setting the voxel count, v, along each side to a power of two allows the fast Fourier transform to have the best efficiency.

3.4 Practical Issues

Both normals and points, due to their nature, cannot be sampled in a manner that satisfies the sampling assumption of the spherical harmonic or Fourier transforms; this is because, unlike a function, it is not possible to sample them at a point. Instead, sampling is performed by averaging about the point, which is achieved by histogramming. Histogramming provides a set of samples that represent either the normals or points at the sample locations required for the discrete SHT. Histogramming acts like a convolution of a square function that has the same width as the histogram bin, which in the 1-D Fourier domain is a multiplication of the sample by a sinc function instead of a delta spike. The spherical-harmonic domain and the 3-D Fourier domain, each have their own function that has an effect analogous to the sinc function. The effect of these functions is that some of the high frequency information is lost. While histogramming reduces the representation of the normals or points in the frequency domain, the impact on correlation is minimal. This is because the effect of noise is mitigated if the same bin still collects a given point or normal, which will occur if the noise is insignificant. Additionally, the goal is coarse registration, for which a histogrammed representation is sufficient.

The discrete spherical-harmonic mathematics in Section 3.2 are presented in a form that slightly differs from that in the literature by Driscoll and Healy (1994) and Healy *et al.* (2003). The reason for this is to help present the mathematics in an uncomplicated manner. The difference that occurs is the scaling of the individual equations, which for the rotation registration is inconsequential as the scaling factor does not change the maximum element in the rotation correlation matrix. Therefore, as the same element in the matrix is still the maximum, the final selected rotation remains unaffected. While the presented mathematics shows the steps for producing the correlation matrix, the actual algorithmic implementation used to achieve the correlation is provided by the SOFT library.[1]

Because the SOFT library is independently called for each SHT, the spherical harmonics are recalculated every time. It may be possible to construct SHT algorithms in a manner which allows either the Legendre, Wigner or spherical-harmonic values to be precomputed and stored. If the fast SHT is conducive to having these values precomputed, then its processing efficiency will be further improved. While improving the processing efficiency is beneficial, it is inconsequential when compared with the exorbitant memory requirement. To store the two sets of spherical harmonics, their correlation in the frequency domain and its inversion requires $2(2B_t^2 + (4B_c^3 - B_c)/3 + (2B_c)^3)$ elements of type double (this includes both the real and imaginary components). If $B_t = B_c = 256$, then this is a memory requirement of 2.4 gigabytes; which with modern computing is not significant (nor is it insignificant), but does not include any required overhead. Checking the true memory usage showed that Matlab R2011a (which can be memory inefficient) with the SOFT library at times required up to 8.5 gigabytes of memory. The primary means of limiting

---

[1] SOFT is a collection of spherical-harmonic routines provided by P.J. Kostelic and D.N. Rockmore (`http://www.cs.dartmouth.edu/~geelong/sphere/`). Version 2.0 is used herein.

this cost is directly related to the used transform and correlation bandwidths. If aspects of the algorithm can be precomputed, their cost will also be determined by the chosen bandwidth, and will themselves require memory space. Additionally, any precomputed values will need to be recalculated if the bandwidth changes.

The chosen transform bandwidth specifies the total number of spherical harmonics used to represent the set of normals in the frequency domain. Because the computational cost of this transform is related to the transform bandwidth, there exists an upper limit for what is feasible given the available computational resources. When the fast SHT is used, the cost of transforming the normals from the spatial domain to the frequency domain is lower than the cost of inverting the correlation to the impulse domain. Therefore, the forward transform can be performed at a higher bandwidth than the inverse transform given equal computational resources. The advantage of increasing the transform bandwidth is that higher harmonics are calculated and distinguished. This is beneficial even when a lower correlation bandwidth is used as the spherical-harmonic degrees between B_t and B_c are isolated and discarded. Discarding these harmonics ensures that they cannot be aliased on to the lower harmonics. However, aliasing will still occur as the harmonics above B_t remain; but as they will be distributed across all calculated harmonics, some of this aliasing will disappear with the discarded harmonics.

Chapter 4

Ground-Truth Datasets

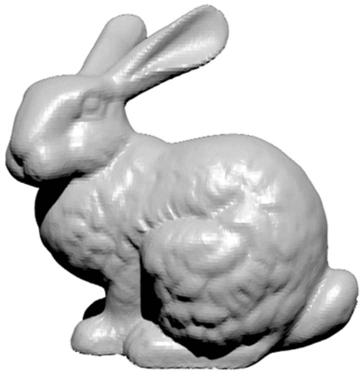
Evaluating each step of the registration pipeline requires that the ground-truth of each tested dataset be known. The purpose of using datasets with ground-truth is that it allows each step to be accurately assessed. In particular, the performance of individual modifications within each step can be compared with each other in terms of both their accuracy and robustness with respect to overlap. This comparison establishes the capability and limitations of each modification, revealing those that provide the most benefit to the registration pipeline. This chapter details the pre-processing used to form the three datasets and how they provide ground-truth. Additional aspects such as overlap and noise generation are also outlined.

4.1 Base Models

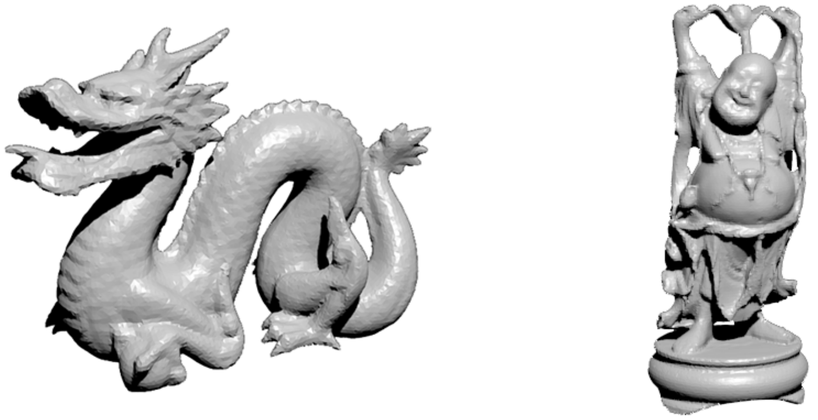
The three datasets used for evaluating the registration pipeline are each formed from a base model that is broken into overlapping segments. The models used for this process are the Stanford bunny, Dragon and Buddha, each of which comes from the Stanford 3D scanning repository[1] and are shown in Figure 4.1; these models contain 35 947, 22 982, and 32 328 points, respectively. While the number of points in these models is comparatively low compared to the achievable dense sampling of modern scanners, it is representative of the sampling density that a low resolution range imaging camera would produce.

---

[1] The Stanford 3-D scanning repository is located at `http://graphics.stanford.edu/data/3Dscanrep/`.



(a) Stanford Bunny



(b) Dragon

(c) Buddha Statue

Figure 4.1: Three models from the Stanford 3-D scanning repository. The models are shown in their initial state before segmentation.

4.2 Forming the Datasets

Creating datasets by segmenting complete models gives the exact rotation and translation between segments. The result of this process is that overlapping regions of different segments each have the exact same points and normals. While this would not occur in a real scene due to changes in sampling distribution and noise, it ensures that neither sampling distribution nor noise are factors that influence the registration results. Additionally, because the relationship between points remains the same for each segment, the exact overlap between segments can be measured. Both of these aspects allow the produced results to accurately reflect how different algorithmic choices compare to each other, and with respect to changes in overlap.

4.2.1 Calculating Normals

The calculation of normals is performed before a model is segmented to ensure that if the same point appears across multiple segments it always has the same associated surface normal. Because there are any number of algorithms available for calculating normals, and as this is a solved problem, MeshLab[2] is used for calculating the normals for each model. This ensures that a reliable algorithm is used for this task without needing to produce a separate implementation. While in practice the reliability of the chosen algorithm will influence the registration performance, it will not have any influence on these datasets as the overlapping regions between segments have the same normals.

4.2.2 Model Segmentation

A dataset is formed from each of the three models by breaking it into 120 overlapping segments, where a segment is the section of the model visible from a single imaging pose. Using uniform random deviates $u_1, u_2 \in [0, 1)$, each pose, A, given in spherical coordinates is randomly placed about the model by

$$A_{\theta\phi} = \left[2\pi u_1,\ \cos^{-1}(2u_2 - 1)\right], \tag{4.1}$$

with its optical axis directed towards the model. This viewing orientation is then further varied by randomly rotating the imaging pose about the optical axis. This process is performed using Blender[3], which is able to extract from the model only the points that are visible to the imaging pose.

Each extracted segment is a point-cloud, with its points and normals maintain their original values. In order for an individual point-cloud to provide a better representation of the capturing process, it needs to be orientated with respect to the imaging pose. To achieve this, a transform is applied to the point-cloud so that it is brought into the same coordinate system as the imaging pose. The translation component of this transform shifts the imaging pose to the origin of the coordinate system, while the rotation orientates the imaging pose so that its optical axis is placed on the z-axis, and its top is aligned with the y-axis. The resulting dataset mimics how the point-clouds would be presented if the model was imaged using an imaging system that only has itself as the reference point.

---

[2] MeshLab is an open-source mesh processing system (`http://meshlab.sourceforge.net/`). Version 1.3.2 is used here.

[3] Blender is an open-source 3-D graphics editor (`http://www.blender.org/`). Version 2.62.0 r44136 is used here.

4.2.3 Noise Generation

Extracting segments from the same base model means that there is no difference in the distribution of normals or points that are common between two overlapping point-clouds. In order to add noise that may occur in a real range imaging system, the position of each normal or point is independently varied. Because the normals are not used in conjunction with the points throughout the registration pipeline, any variations to the normals do not need to be reflected in the points, or vice-versa. This section outlines how noise is generated for each normal. Noise in the points only affects the translational alignment, therefore the manner in which noise is generated for points is contained within the Translational Alignment chapter, in particular Section 7.2.

Spherical-harmonic correlation is carried out using the normals extracted from the two point-clouds, because noise in a real imaging system is independent to each sample, the noise is generated and added to every normal individually. This is achieved by creating a random normal, $\hat{\mathbf{r}}$, near the sphere's pole, with its random placement being given as

$$\hat{\mathbf{r}} = \left(\left|v\sigma\right|, 2\pi u_3\right), \tag{4.2}$$

where v is a normally distributed random number, with σ specifying the one standard deviation angle, and $u_3 \in [0, 1)$ being a uniform random deviate. This noise is then applied to $\hat{\mathbf{n}}$ by performing the rotation

$$\hat{\mathbf{n}}' = \hat{\mathbf{r}}\mathbf{R}, \tag{4.3}$$

where $\mathbf{R}$ rotates the pole to $\hat{\mathbf{n}}$. The rotation axis of $\mathbf{R}$ is perpendicular to both the pole and $\hat{\mathbf{n}}$, with the rotation angle being $\hat{\mathbf{n}}_\theta$; care must be taken to ensure that the rotation direction is correct.

While this approach of applying noise to normals is some-what convoluted, it allows the noise handling capability of spherical-harmonic correlation to be revealed. The level of noise is easily modified as it has a direct relationship with the value of σ; if σ is increased, so does the noise level.

4.3 Quantifying Registration Performance

Registration performance is quantified by testing every possible segment pairing, including registration of an image with itself. This produces 7 260 pairs for

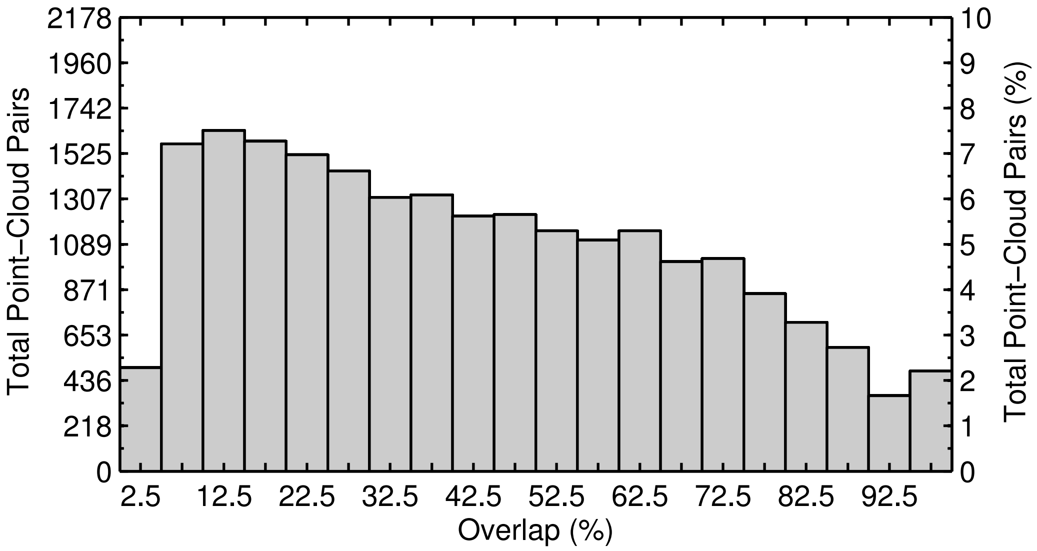


Figure 4.2: The distribution of overlap between the 21 780 point-cloud pairs that are formed from the Bunny, Buddha and Dragon models.

each model, forming a total of 21 780 pairs across the three datasets. Because the rotation and translation of each point-cloud with respect to the original model are known, the exact registration accuracy, for either rotation or translation, between two registered point-clouds can be measured. This accuracy can also be thresholded to establish a percentage of point-cloud pairs which have achieved correct alignment.

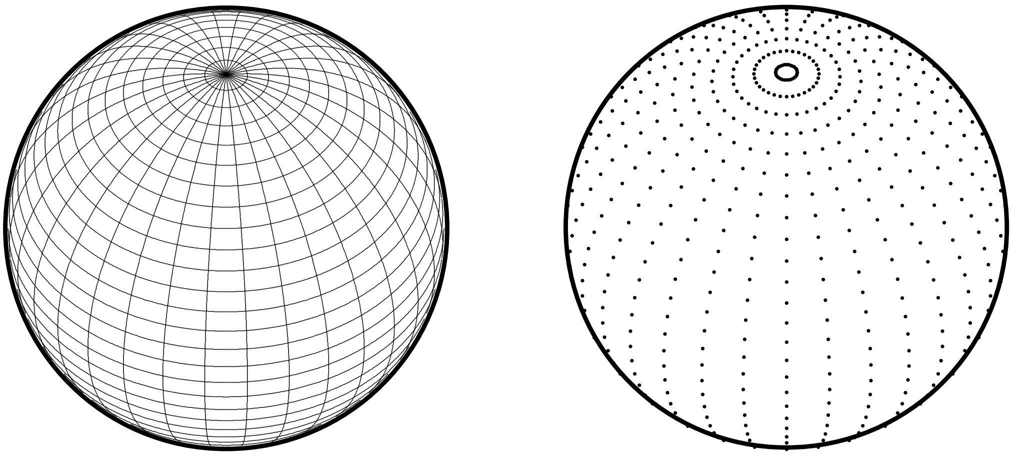
A major factor which influences registration performance is the percentage of overlap between two point-clouds. While it could be expected that overlap is directly related to the rotational difference between two imaging poses, this is not a consistent measure. An example of this is if the imaging pose is rotated by 180° about is optical axis, the overlap would remain at 100%, but the rotational difference is maximised. Instead, overlap is calculated by uniquely ascribing an index to each point of a model, the number of points common to two point-clouds are then counted. However, as two point-clouds do not always have the same number of points (such as if one point-cloud is a subset of another), the overlap percentage is determined by normalising the number of common points by the total number of points in the larger point-cloud. The overall distribution of overlap between the 21 780 point-cloud pairs is presented in Figure 4.2, which shows that the testing will use more point-cloud pairs that have less than 50% overlap.

Chapter 5

Binning Normals

Range imaging systems are capable of producing point-clouds that can contain millions of points. As a normal is generated for each point, there will be an equivalent number of normals. The computational cost of directly transforming this number of normals into spherical harmonics will be immense. This cost is significantly reduced by binning the normals and converting the bin centres to spherical harmonics instead. Having binned the normals, each bin centre is weighted by the total number of normals that it has collected, thus providing a representation of its portion of the surface area. Bins are formed in $\mathbb{S}^2$ by either uniformly dividing the polar and azimuthal angles of a unit sphere or by forming bins about points in $\mathbb{S}^2$ that have a near uniform distribution. Attempting to produce a uniform distribution is referred to as the Fekete problem or the Thomson problem, and has many proposed solutions (Saff and Kuijlaars, 1997; Gorski *et al.*, 2005; Teanby, 2006; Williamson, 2007; Koay, 2011a,b).

The manner in which normals are binned plays a role in registration accuracy, in particular, the more uniform the bin distribution, the better the accuracy. Because bin size and shape affect the relationship between normals, three binning schemes are thoroughly investigated throughout this chapter to determine how they impact the rotational registration capability of spherical-harmonic correlation. This chapter describes the three binning schemes and their implementation, then evaluates their performance with respect to accuracy, efficiency and noise.



(a) Equiangle Bins

(b) Equiangle Bin Centres

Figure 5.1: Orthographic projection of the equiangular binning scheme. A bandwidth of 16 is used to give 1024 bins, which are distributed around the sphere. The spheres are tilted 45° making the north pole visible. Figures 5.1a and 5.1b show the distribution of bin boundaries and their centres, respectively.

5.1 Binning Schemes

In this section, three schemes for subdividing the surface of the unit sphere and binning normals are described in detail. These schemes are equiangular binning, geodesic subdivision using the icosahedron Platonic solid, and binning based on the Fibonacci spiral. This description covers how the bins are positioned and formed, the procedure that is followed for allocating normals to bins and the algorithm variants that influence the registration.

In the following, let the unit sphere be defined by a spherical coordinate system in which $\theta \in [0, \pi]$ is the polar angle (or colatitude) and $\phi \in [0, 2\pi)$ is the azimuthal angle. Each surface normal, $\hat{\mathbf{n}}$, and bin centre, $\hat{\mathbf{b}}$, is expressed in terms of these coordinates.

5.1.1 Equiangle Grid

The equiangle grid, shown in Figure 5.1a, is the simplest approach for subdividing the surface of a sphere. This grid is the same as the fast spherical-harmonic transform sample points (see Section 3.2.2), as such, it was used by Makadia *et al.* (2006) for binning normals in preparation for spherical-harmonic registration. The grid is formed using the equiangle sampling scheme outlined by Healy *et al.* (2003), in which the polar angle θ and the azimuthal angle ϕ are divided into $2B_t$ equal sections, with B_t being the specified transform bandwidth. These polar and azimuthal sections, indexed by

$$j = 0, 1, \ldots, 2B_t - 1, \tag{5.1}$$

and

$$k = 0, 1, \ldots, 2B_t - 1, \tag{5.2}$$

respectively, form a total of $4B_t^2$ bins. The bin that each normal belongs to is directly calculated by

$$j = \begin{cases} \left\lfloor \dfrac{\hat{\mathbf{n}}_\theta 2B_t}{\pi} \right\rfloor & 0 \leq \hat{\mathbf{n}}_\theta < \pi \\ 2B_t - 1 & \hat{\mathbf{n}}_\theta = \pi \end{cases}, \tag{5.3}$$

$$k = \left\lfloor \frac{\hat{\mathbf{n}}_\phi B_t}{\pi} \right\rfloor. \tag{5.4}$$

The bin weight is incremented by one for each normal that it captures.

Using a (j, k) index pair, the coordinate of the (j, k) bin centre is given as

$$\theta_j = \frac{(2j+1)\pi}{4B_t}, \tag{5.5}$$

$$\phi_k = \frac{k\pi}{B_t}. \tag{5.6}$$

Figure 5.1b illustrates the distribution of these bin centres on the unit sphere. Note that the equiangular grid has decreased bin density at the equator.

Weighting by Bin Surface Area

Equally dividing the polar and azimuthal angles to form the equiangle grid distinctly affects the surface area of each equiangle bin: the closer a bin is to the sphere's equator, the larger its area for capturing normals. Figure 5.2 shows that the difference in surface area between the smallest and largest bins increases with bandwidth. At a bandwidth of 128, the largest bin is 163 times the size of the smallest bin; not the approximate 10 times difference stated by Makadia *et al.* (2006). If this area is not taken into account, the influence that each bin has on spherical-harmonic correlation is solely based on the number of normals that it collects. It is incorrect to simply assume that the resulting bin weight is just a reflection of its capturing power. This is because the fast SHT normally samples a continuous function, whereas the distribution of normals is discrete. Therefore, correct sampling in each bin is formed by dividing its

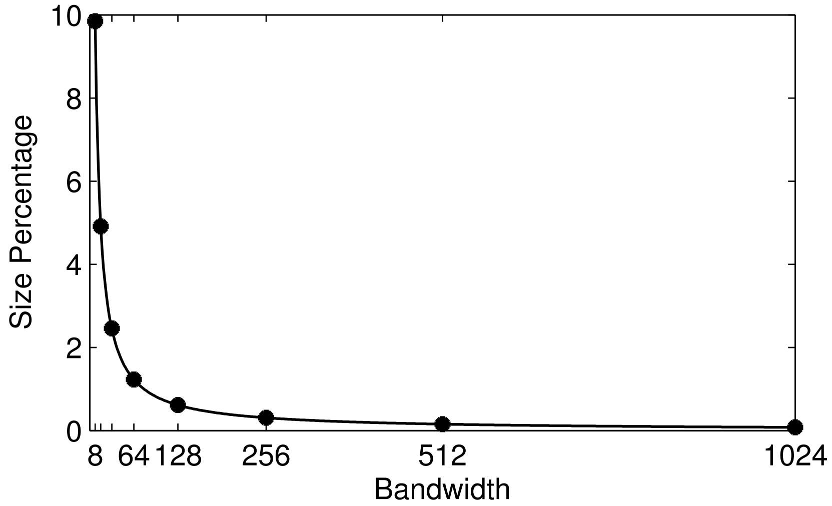


Figure 5.2: The percentage of area that the smallest equiangular bin encompasses of the largest equiangular bin, with respect to bandwidth. As bandwidth increases so does the difference in area between the smallest and largest bins.

count of normals by the size of its surface area. By sampling in this fashion, any bias towards the larger bins is removed.

The bin area is independent of the azimuth angle, thus with the polar index j and the bandwidth B_t the fractional bin area is given by

$$A_{B_t}(j) = \frac{\cos\left(\dfrac{\pi j}{2B_t}\right) - \cos\left(\dfrac{\pi(j+1)}{2B_t}\right)}{4B_t}, \tag{5.7}$$

that is, it has been normalised by 4π, the total surface area of a unit sphere. To confirm that dividing each bin's weighting by its surface area is the correct form of sampling, it and two other approaches of reweighting are investigated. The other two approaches are composed by multiplying each bin's weight by its surface area and not performing any reweighting. Multiplying by surface area emphasises the larger bins, while not rescaling by surface area weights each bin purely on its ability to capture normals.

Orientation of Normals

Many range imaging systems are only capable of viewing a scene from a single pose at any given time. Because of this, the angle between a surface normal and the camera's optical axis is less than 90°. Therefore, when the normals are collected together, they only occur on half of the unit sphere, instead of being distributed about the entire unit sphere. The manner in which the hemisphere

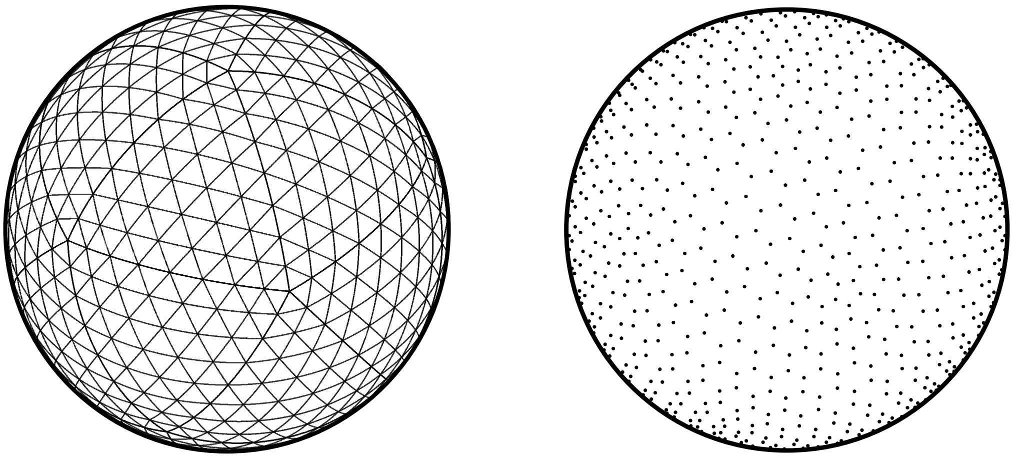

(a) Icosahedron Bins (b) Icosahedron Bin Centres

Figure 5.3: Orthographic projection of icosahedron binning, with 1280 bins distributed around the sphere. The spheres are tilted 45° making the north pole visible. Figures 5.3a and 5.3b show the distribution of bin boundaries and their centres, respectively.

containing the normals is orientated with respect to the equiangle grid changes which bins collect normals. If the chosen orientation aligns the optical axis with the equator of the grid, the largest bins are placed in the centre of the hemisphere, resulting in a poorer sampling of the normals. Alternatively, if the orientation aligns the optical axis with the pole of the grid, the smallest bins are placed in the centre of the hemisphere, providing a denser sampling of the normals.

Changing the orientation of the normals with respect to the equiangle grid is achieved by simply rotating the normals before binning them. The registration accuracy of both orientations is determined as part of this investigation.

5.1.2 Geodesic Subdivision

Geodesic subdivision is achieved by centring a Platonic solid at the origin of the coordinate system and projecting each face on to the surface of a unit sphere to form a bin (Williamson, 2007). The icosahedron is used as it has the most faces of the five Platonic solids. As each face of the icosahedron is an equilateral triangle, more bins can be formed by breaking each face into four subsequent equilateral triangular regions, with this process repeated to a preferred depth; Figure 5.3a shows the icosahedron with a depth of three.

The twelve vertices that define an icosahedron are arranged on a unit sphere using the golden ratio,

$$\tau = \frac{1+\sqrt{5}}{2}, \tag{5.8}$$

where if a vertex is aligned to each pole are given as (in Cartesian coordinates):

$$\begin{array}{c} \begin{array}{ccc} \quad x \quad & \quad y \quad & \quad z \quad \end{array} \\ \begin{bmatrix} 0 & 0 & \pm 1 \\ \pm 2/\sqrt{5} & 0 & \mp(2/(\tau^2+1)-1) \\ \pm 1/(\tau^2+1) & \pm 1/\sqrt{\tau^{-2}+1} & \pm 1/\sqrt{5} \\ \pm 1/(\tau^2+1) & \mp 1/\sqrt{\tau^{-2}+1} & \pm 1/\sqrt{5} \\ \mp 1/(\tau^{-2}+1) & \mp 1/\sqrt{\tau^2+1} & \pm 1/\sqrt{5} \\ \mp 1/(\tau^{-2}+1) & \pm 1/\sqrt{\tau^2+1} & \pm 1/\sqrt{5} \end{bmatrix} \end{array}. \tag{5.9}$$

These vertices are presented in pairs that are antipodal to each other on the sphere. The 20 faces defined by these vertices are projected onto the sphere, creating the initial bins. By specifying a depth, d, more bins can be produced by subsequently dividing each face into four smaller equilateral triangles and projecting these on to the sphere surface. Each face creates 4^d triangles, with the total number of bins being

$$n = 20 \times 4^d. \tag{5.10}$$

The effect of this projection is that the size and shape of each bin changes. Bins closer to the centre of a face have greater surface area than those at the edge, thus there is a higher density of bin centres at the edges. This change in density impacts the implementation described below as the best branch to traverse at a lower depth may not contain the closest bin centre to the point, instead placing the normal in an adjacent bin. It has been hypothesized by Teanby (2006) that for most practical purposes this slight bias is insignificant; this claim is confirmed true by Larkins *et al.* (2012), an analysis previously performed to investigate the binning of normals.

Binning Procedure

Binning normals using geodesic subdivision is performed in two stages. First, a *forest* of trees is built that stores the bins at each depth, and second, the forest is searched to find the closest bin.

The forest is constructed as 20 individual trees, one for each face, with each face defined by three vertices from those listed in equation (5.9). The

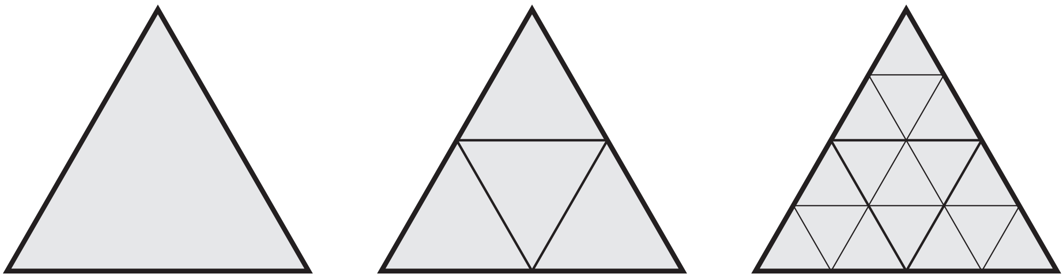


Figure 5.4: Demonstration of how a triangular face is divided into four subsequent triangles. These new triangles can be further divided in the same fashion.

centre of a face is projected onto the sphere surface and stored as a spherical coordinate in the root node of the corresponding tree. For every new level, each node has four children, one for each of the four subsequent triangles created; as demonstrated in Figure 5.4. The centres of these new triangles are again projected onto the sphere surface and stored. This constructs a quadtree formed as a linear array, with the index of each child node, j_c, found from the index of the parent node, j_p, by

$$j_c = 4j_p + c, \qquad c \in \{1, 2, 3, 4\}. \tag{5.11}$$

Each tree contains t nodes, given by

$$t = \frac{4^{d+1} - 1}{3}, \tag{5.12}$$

where d is the depth of the tree. A linear array is composed that only stores the leaf nodes and their weights from all 20 trees, in which the index, j_b, of a bin in this array is given by

$$j_b = 4^d \left(f - \frac{1}{3} \right) + j_l + \frac{1}{3}, \qquad f \in \{0, \dots, 19\}, \tag{5.13}$$

where f is the face and j_l is the index of the leaf node in the above quadtree.

Binning a normal begins by identifying which of the 20 faces is closest to it; these faces are the zero depth bin. As the normals and bins represent a position on a unit sphere, the great circle distance (GCD) is the same as the angle between the normal, $\hat{\mathbf{n}}$, and the centre of a bin, $\hat{\mathbf{b}}$, calculated as

$$\psi = \cos^{-1} (\sin \hat{\mathbf{n}}_\theta \sin \hat{\mathbf{b}}_\theta \cos (\hat{\mathbf{n}}_\phi - \hat{\mathbf{b}}_\phi) + \cos \hat{\mathbf{n}}_\theta \cos \hat{\mathbf{b}}_\theta). \tag{5.14}$$

The layout of the vertices of the icosahedron produces pairs of root nodes that are on direct opposite sides of the unit sphere. The distance therefore is calculated for only ten of the nodes, and the distance of a node's counterpart is given by $\pi - \psi$. If a single shortest distance is found, then the normal goes to the corresponding root node and the search down the tree for the wanted bin begins.

Given the closest root node, the distance between its four children and the normal is found using equation (5.14). This process of finding the closest node and going down its branch is repeated until a leaf node is reached; with the spherical coordinate of the leaf node being the closest bin. Using equation (5.13) the weight of this bin is then incremented. Throughout this process, both for finding the root node and for searching down the quadtree, there is the unlikely possibility that a normal is located where multiple nodes have the same shortest distance, in this situation one of these closest nodes is arbitrarily chosen and used.

5.1.3 Fibonacci Spiral

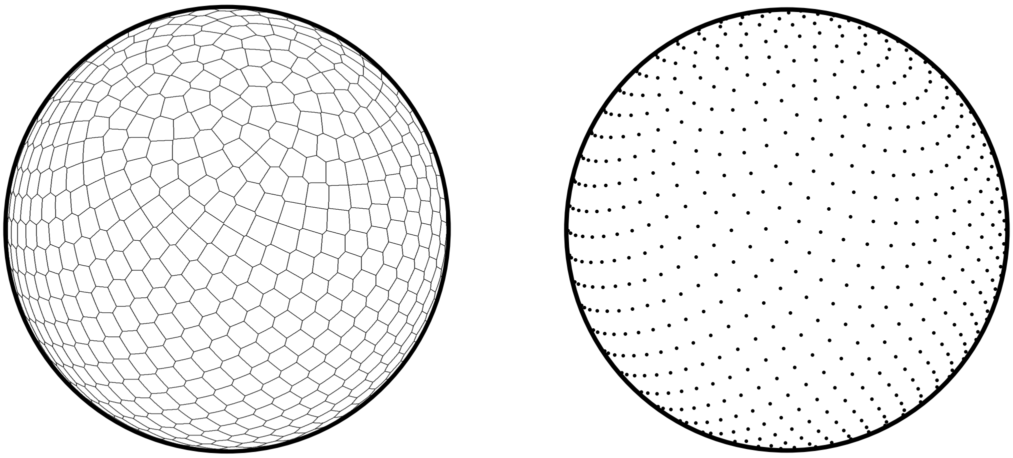
The Fibonacci spiral (González, 2010) is a point distribution method that we propose as an alternative approach for binning normals (Larkins *et al.*, 2012). A spiral is created around the sphere from the north to south pole, with each point placed at equal increments along the spiral, creating a near uniform distribution of bin centres around the sphere, as shown in Figure 5.5b. The bin boundaries are formed around each point, as shown in Figure 5.5a. An odd number of points must be along the spiral to ensure both hemispheres contain the same number of bin centres. Using the number of bins, n, the bin centres are then found as spherical coordinates at

$$\hat{\mathbf{b}} = \left[\sin^{-1}\left(\frac{2d}{n}\right) + \frac{\pi}{2}, \frac{2\pi}{\tau}(d \pmod{\tau}) \right], \tag{5.15}$$

where

$$d = \frac{1-n}{2}, \ldots, \frac{n-1}{2} \tag{5.16}$$

uniquely indexes each point along the spiral and τ is the golden ratio from equation (5.8).



(a) Fibonacci Spiral Bins (b) Fibonacci Spiral Bin Centers

Figure 5.5: Orthographic projection of Fibonacci spiral binning, with 1025 bins distributed around the sphere. The spheres are tilted 45° making the north pole visible. Figures 5.5a and 5.5b show the distribution of bin boundaries and their centres, respectively.

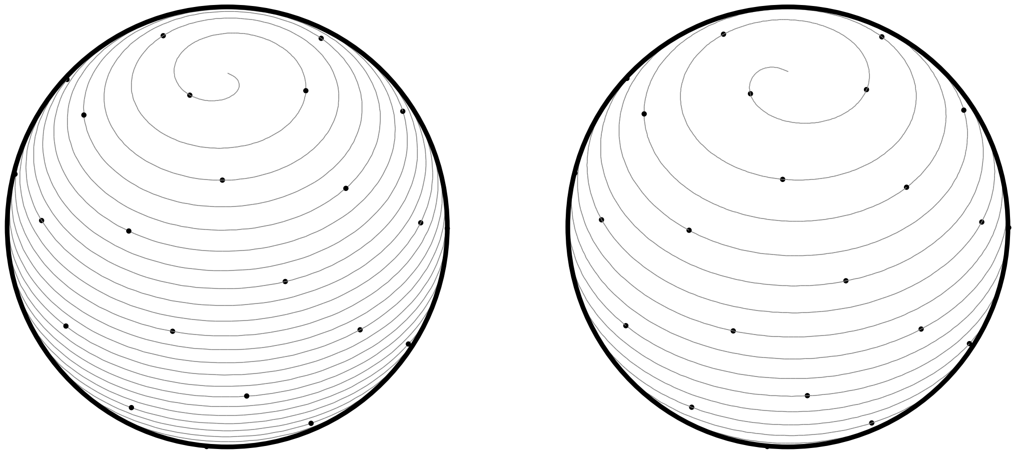
Binning Procedure

Binning normals on a Fibonacci spiral distribution can be achieved in a variety of ways, such as brute force calculation or by a form of Delaunay triangulation. A new algorithm is introduced here that does not require a search or storage structure for finding the closest bin; it instead identifies which bin a normal belongs to by using the turns of the spiral. Because of this, there is no initial construction needed before performing the search, and the number of evaluations of equation (5.14) is reduced. The intersection between each turn and the constant longitudinal line on which the normal is situated is used to locate the closest bins. Using equation (5.15) the bins are distributed on a spiral that is formed clockwise around the sphere. Equation (5.15) can instead be expressed as

$$\hat{\mathbf{b}} = \left[\sin^{-1}\left(\frac{2d}{n}\right) + \frac{\pi}{2}, \frac{2\pi}{\tau^2}(-d \pmod{\tau^2})\right], \tag{5.17}$$

which produces the exact same bins, but on a spiral that is formed counter-clockwise around the sphere. The counter-clockwise spiral decreases the turn density required to place the bins, benefiting the proposed binning procedure as there are fewer turns to test. Figure 5.6 illustrates the difference in turn density produced by these two spirals when $n = 45$.

Determining the intersect locations between the spiral and the constant



(a) Clockwise Spiral (b) Counter-Clockwise Spiral

Figure 5.6: The two Fibonacci spirals that give the same uniform distribution of points on the sphere. These spirals distribute 45 points around the sphere; the clockwise spiral produces a much greater turn density than that of the counter-clockwise spiral.

longitudinal line defined by $\hat{\mathbf{n}}_\phi$ is the first step in identifying the bin closest to $\hat{\mathbf{n}}$. These intersects occur when the ϕ value of the spiral equals $\hat{\mathbf{n}}_\phi$; this requires d being treated as a real value instead of an integer, and allows the following equation rearrangement

$$\begin{aligned} \frac{2\pi}{\tau^2}(-d \pmod{\tau^2}) &= \hat{\mathbf{n}}_\phi, \\ -d - \left\lfloor \frac{-d}{\tau^2} \right\rfloor \tau^2 &= \frac{\hat{\mathbf{n}}_\phi \tau^2}{2\pi}, \\ \left\lfloor \frac{-d}{\tau^2} \right\rfloor &= -\frac{\hat{\mathbf{n}}_\phi}{2\pi} - \frac{d}{\tau^2}. \end{aligned} \tag{5.18}$$

Rearranging for d is not possible as the floor does not permit a direct solution. The floor is instead dealt with by substituting in the variable z (due to the nature of the floor, z is always an integer), in which each turn of the spiral is labelled with a unique integer value z; allowing d to be expressed as

$$\begin{aligned} z &= -\frac{\hat{\mathbf{n}}_\phi}{2\pi} - \frac{d}{\tau^2}, \\ -\tau^2 \left(z + \frac{\hat{\mathbf{n}}_\phi}{2\pi} \right) &= d. \end{aligned} \tag{5.19}$$

Placing this equation for d into equation (5.17), gives the spherical coordinate of each intersect point as

$$\hat{\mathbf{p}}_z = \left[-\sin^{-1}\left(\frac{z+v_2}{v_1}\right) + \frac{\pi}{2},\ \hat{\mathbf{n}}_\phi \right], \tag{5.20}$$

but as the number of turns are finite, only the select range of $z = z_{\text{min}}, \ldots, z_{\text{max}}$ gives legitimate locations on the spiral, where

$$z_{\text{min}} = \lceil -v_1 - v_2 \rceil, \tag{5.21}$$

and

$$z_{\text{max}} = \lfloor v_1 - v_2 \rfloor, \tag{5.22}$$

given

$$v_1 = \frac{n}{2\tau^2}, \tag{5.23}$$

and

$$v_2 = \frac{\hat{\mathbf{n}}_\phi}{2\pi}, \tag{5.24}$$

constant variables common throughout these equations. The bin centre on turn z closest to $\hat{\mathbf{p}}_z$ is found via equation (5.15) when

$$d = \text{round}\left(-\tau^2(z+v_2)\right). \tag{5.25}$$

The z value of the normal $\hat{\mathbf{n}}$ is initially a real value (as it is likely to be between the turns of the spiral) given by

$$z = v_1 \sin\left(\hat{\mathbf{n}}_\theta - \frac{\pi}{2}\right) - v_2, \tag{5.26}$$

with the integer identifiers of the turns either side of $\hat{\mathbf{n}}$ being the floor and ceiling of this z value. The bins are searched by first testing these two turns; the bin centre with the smaller distance, calculated from equation (5.14), is then stored. Each successive turn away from $\hat{\mathbf{n}}$ is tested, both stepping up and down the sphere. If the distance from $\hat{\mathbf{n}}$ to $\hat{\mathbf{p}}$ is greater than the current smallest distance, the bin centre on this turn is tested and the stepping in this direction is stopped.

Because the turns of the spiral have a greater spacing towards the poles, the above approach may miss the closest bin centre, therefore, when the normal is above the spiral or there is only one turn above it, that is, when $\hat{\mathbf{n}}_\theta < \hat{\mathbf{p}}_\theta\left(z_{\text{min}}+1\right)$, all bin centres from the closest turn location up must be tested. Once this is done, searching down the spiral is carried out as described above. This process is repeated by testing all bin centres down the spiral when $\hat{\mathbf{n}}_\theta > \hat{\mathbf{p}}_\theta\,(z_{\max} - 1)$, followed by testing each turn up the spiral.

This implementation requires two full turns of the spiral to operate correctly, which occurs when there are seven or more bins. When there are fewer than seven bins, directly calculating the distance to every bin is feasible, however, a far greater number of bins are used, hence the need for the search.

5.2 Methodology

The analysis performed for testing the three binning schemes uses the Dragon, Buddha statue and the Bunny datasets, each of which are described in Chapter 4. These three datasets, constructed by segmenting each model into 120 overlapping segments, enable a detailed analysis as the percentage of overlap and the angle of separation between every point-cloud is known *a priori*. The rotation which coarsely aligns the two sets of normals from a point-cloud pair is selected from a correlation matrix (whose size is determined by the correlation bandwidth) calculated over the spherical domain. Unless the selected rotation perfectly matches the rotational difference between the two point-clouds, an error in rotational alignment will remain due to the discrete sampling. This remaining rotation error is measured as the angle between the found position of the point-cloud and its known true position.

Determining whether the rotational alignment is correct is a case of applying a threshold to the resulting rotation error. If the rotation error is equal to or lower than the threshold, then the alignment is classified as correct, otherwise it is incorrect. As there is no prescribed fine registration algorithm for use in conjunction with spherical-harmonic correlation, varying the threshold provides a means of measuring how the percentage of correct alignments changes. If a particular fine registration algorithm is applied, and its capability of handling rotation error is known, then the results give a good indication of how well this fine registration algorithm will perform. In most cases the rotation error threshold is tested at 0.2° increments from 0° to 12°, but as this is not always practical, such as in a table, rotation error thresholds of 1°, 2°, 5°, 10° and 15° are used instead.

The ability of spherical-harmonic correlation to accurately register a point-cloud pair is largely influenced by how well the normals are represented in the frequency domain. The transform bandwidth, B_t, determines the level of

Table 5.1: The number of bins used at each bandwidth for the three binning schemes.

Binning Scheme	Transform Bandwidth					
	16	32	64	128	256	512
Equiangle Grid	1 024	4 096	16 384	65 536	262 144	1 048 576
Icosahedron	1 280	5 120	20 480	–	–	–
Fibonacci Spiral	1 025	4 097	16 385	–	–	–

representation as it is the maximum degree of spherical harmonic calculated. Given a transform bandwidth B_t, $4B_t^2$ sample points are specified over the unit sphere as this allows the fast SHT to avoid ambiguous aliasing. As the fast SHT uses equiangle sampling, an equiangle bin is formed around each sample point. Three approaches for reweighting each equiangle bin by its surface area are compared. This reweighting is not performed on the icosahedron and Fibonacci binning schemes as their bins are more uniform both in size and distribution. Converting the icosahedron and Fibonacci bins to spherical harmonics is accomplished via least-squares. This conversion can be achieved with any number of sample points, but a minimum of $4B_t^2$ bins are used to avoid an underdetermined system. The number of bins required for each of the three binning schemes are listed in Table 5.1. Due to the manner in which bins are distributed by the icosahedron and Fibonacci schemes, having the exact number of required bins is not possible. Therefore, the number of bins used for these two schemes at each transform bandwidth is the closest sampling distribution above $4B_t^2$ that they can each achieve. Performing a least-squares inversion has a high computational cost which limits these two schemes to a bandwidth of 64.

The transform and correlation bandwidths are often set to the same value, but the transform bandwidth can be greater. The computational resources required by the fast SHT for converting from the spatial to frequency domains is lower than that required for constructing the correlation matrix. It is for this reason that a higher transform bandwidth is possible, even though only the spherical-harmonics up to the correlation bandwidth are used. The primary advantage of having a higher transform bandwidth is covered in Section 3.4. The secondary advantage is that there is a greater sampling density on the unit sphere, meaning that the lower spherical harmonics are provided with a more accurate representation of the normals. Both the correlation and transform bandwidths are tested at powers of two. When the transform bandwidth is greater than 128, the correlation bandwidth is limited herein to 128 due to the aforementioned computational cost. The exception to this is in the bandwidth and overlap analysis in which the best performing equiangle binning scheme is tested with a correlation bandwidth of 256.

The three models that are used for this analysis are initially provided by the Stanford 3D repository in a form that has been reconstructed from individual scans. It is unlikely that this process was perfect, thus an insignificant amount of noise will have been introduced to each model; but as the overlapping segments are extracted from the same model, their overlapping points and normals are identical. The initial testing is performed without introducing any noise to the normals, allowing each binning scheme to be quantified using perfect scenarios. Once the best scheme or schemes are identified, they are further analysed to determine how robust they are to noise. The manner in which Gaussian noise is added to each normal is outlined in Section 4.2.3. The variability of the noise is specified by a base angle which is set at one standard deviation from the normal: five such base angles are used for introducing noise, with these being 0.5°, 1°, 2°, 5° and 10°.

5.3 Results

The results presented reveal how registration accuracy and efficiency are affected by the manner in which normals are binned. Additionally, the robustness of spherical-harmonic correlation for registration is quantified by adding varying levels of noise to the normals.

5.3.1 Equiangle Accuracy

This section analyses different algorithm choices and aspects of the point-clouds and how their relationship to each other impacts registration. As such, this analysis is broken into three subsequent sections: bin surface area and orientation of normals, bandwidth and overlap.

Bin Surface Area and Orientation of Normals

The process of binning normals on an equiangle grid can take a variety of approaches. The results presented here investigate six approaches, each of which are formed as a combination of reweighting each bin by its surface area and the global orientation of the normals. Figure 5.7 shows the performance of each approach over a variety of bandwidths. The manner in which each bin is weighted with respect to its surface area size has the greatest impact on registration ability. Dividing the count of normals in each bin by its surface area provides the best representation of the normals in the frequency domain. This is followed by giving every bin an equal area weighting, that is, a bin's weight is only determined by the total count of normals that it collects. The poorest performance occurs when each bin is multiplied by its surface area. The ability of the three approaches for reweighting each bin remains consistent across all tested bandwidths, however, their individual ability can be highly influenced by the orientation of the normals.

Two global orientations of the normals are tested: the first places the normals so that binning occurs at the pole of the equiangle grid, and the second orientates the normals so that binning occurs at the equator of the grid. The influence that a given orientation has on the percentage of correctly aligned point-cloud pairs is dependent on the bin reweighting approach. The combination of binning at the pole and dividing by the surface area has the highest consistent performance, especially at lower bandwidths. However, when the bandwidth is equal to or greater than 128 and the bins are divided by area, the orientation has little influence on the percentage of correct alignments. The other four combinations fluctuate extensively with respect to each other at the lower bandwidths. At the higher bandwidths these four combinations remain stable with respect to each other, with a distinct separation in ability occurring when bin area is not incorporated and each bin is multiplied by its area. These two weighting approaches also differ with orientation as binning at the equator is more conducive to the bins being multiplied by their area, while binning at the pole is better when bin area is not incorporated.

Increasing the rotation error threshold on the best performing combination (binning at the pole and dividing by bin size), is presented in Figure 5.8. This graph shows that after the initial spike of aligned point-cloud pairs, the improvement is only gradual. The threshold at which this spike plateaus is dependent upon the chosen bandwidth, with this occurring below the 10° threshold for all bandwidths except 16, and at approximately 2° when the bandwidth is equal to or greater than 128. It also shows that a large percentage of correct alignments only occur in the last 10° of the rotation error threshold. The influence that the bandwidth has on this combination is covered with more detail in the following section.

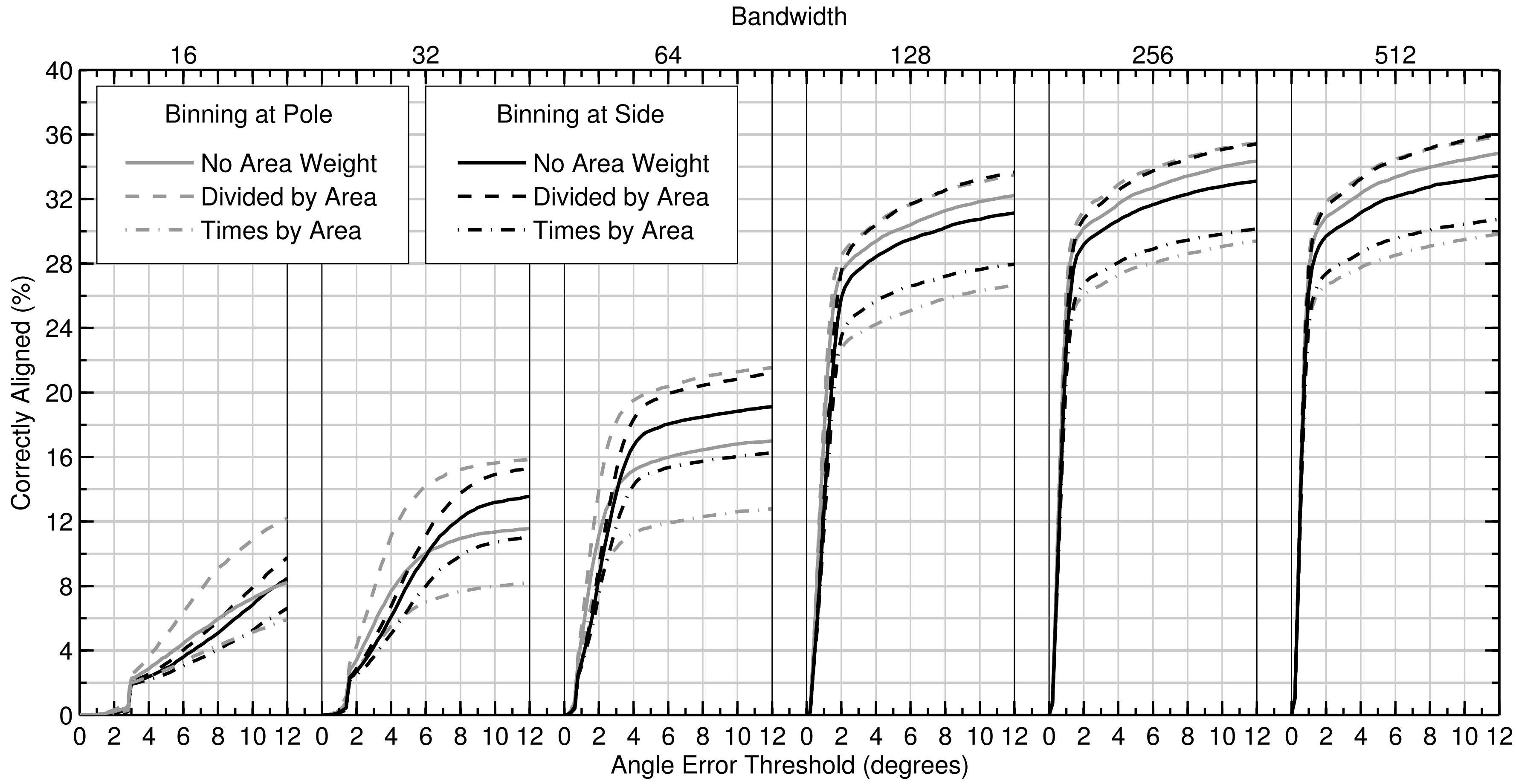


Figure 5.7: Graph comparing six equiangle binning variations, where a rotation error threshold specifies whether a point-cloud pair has correct alignment. The graph also shows how these results differ at six different bandwidths (bandwidths 256 and 512 use a correlation bandwidth of 128).

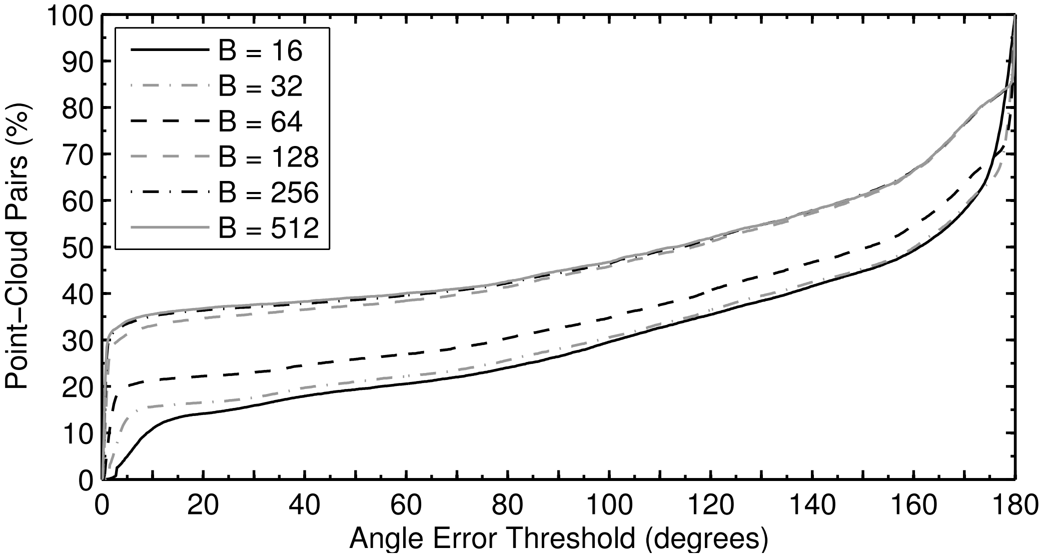


Figure 5.8: The percentage of correctly aligned point-cloud pairs when a threshold specifies the maximum allowed rotation error between two point-clouds. In this graph, the normals were binned at the pole of the equiangle grid, with each bin being divided by its surface area.

Bandwidth

Because spherical-harmonic correlation produces a discrete rotation, it is improbable that the desired rotation will exactly match the discrete rotation; therefore some degree of rotational misalignment will invariably remain. The level of discretisation is a product of the chosen transform and correlation bandwidths. In this section, the likelihood of achieving correct alignment is gauged when the transform and correlation bandwidths are varied, which allows the trade-off between efficiency and accuracy to be adjusted. Although reducing the bandwidth reduces the computational resources required, there is a limit to the minimum bandwidth needed to achieve acceptable results. The results presented in Figure 5.8 show the extent to which bandwidth influences the accuracy when using the best equiangle binning combination. These results along with transform bandwidths of 8 and 1024 are shown with specific values in Table 5.2 at five rotation error thresholds. The lowest presented bandwidth is that of 8, which at a threshold of 10° is capable of correctly aligning approximately 5% of the point-cloud pairs. Each subsequent bandwidth successively improves the percentage of correct alignments, with a bandwidth of 64 bringing 21% of the point-cloud pairs into alignment, more than quadrupling that of bandwidth 8. In comparison, the jumps from bandwidth 64 to 128 and 128 to 256 each provide significant improvements that correctly align 33% and 46%

of the point-cloud pairs, respectively.

The results show that registration accuracy has the greatest improvement when using higher correlation bandwidths, however, higher correlation bandwidths are restricted by their computational cost (see Section 3.2.2). Alternatively, an additional but smaller boost to performance is achieved by fixing the correlation bandwidth and increasing the transform bandwidth; increasing the transform bandwidth improves the representation of the normals across all spherical-harmonic degrees. The results in Table 5.2 show how the improved representation benefits the registration accuracy even though only the spherical harmonics up to the correlation bandwidth are used. The boost to registration accuracy when only increasing the transform bandwidth (the correlation bandwidth is limited to 128) tapers off the further it is increased. For example, when the transform bandwidth is 128, accuracy is 33.1% at a rotation error threshold of 10°, increasing the transform bandwidth to 256 gives an additional 2%, while a transform bandwidth of 512 is only an additional 0.3%. Interestingly, increasing the transform bandwidth further to 1 024 has a marginal adverse impact on registration accuracy.

Increasing the correlation bandwidth not only improves the registration accuracy, but it also decreases the minimum rotation error that occurs between two aligned point-clouds. This is seen in Figure 5.9 where the initial spike of correct alignments tapers off at lower rotation error thresholds. At a correlation bandwidth of 128 and 256, the spike tapers off at a rotation error threshold of approximately 2° and 1°, respectively. Reducing the rotational misalignment benefits the fine registration algorithm, especially if it is restricted to a maximum rotational misalignment that it is capable of handling.

Overlap

The ability of spherical-harmonic correlation to correctly align two point-clouds is primarily linked to the amount of overlap that occurs between them. Figure 5.10 shows how the percentage of correctly aligned point-cloud pairs varies with differing amounts of overlap when the rotation error threshold is set to 10°. The results presented by this graph confirm that as overlap increases so does the percentage of correct alignments. This increase in accuracy is tied to the bandwidth, with the bandwidths of 64 and below starting to achieve correct alignment when overlap is 60%. For bandwidths 128 and above, the starting point is closer to 30%, with approximately half of all point-cloud pairs that have 50% overlap achieving correct alignment. The separation in ability be-

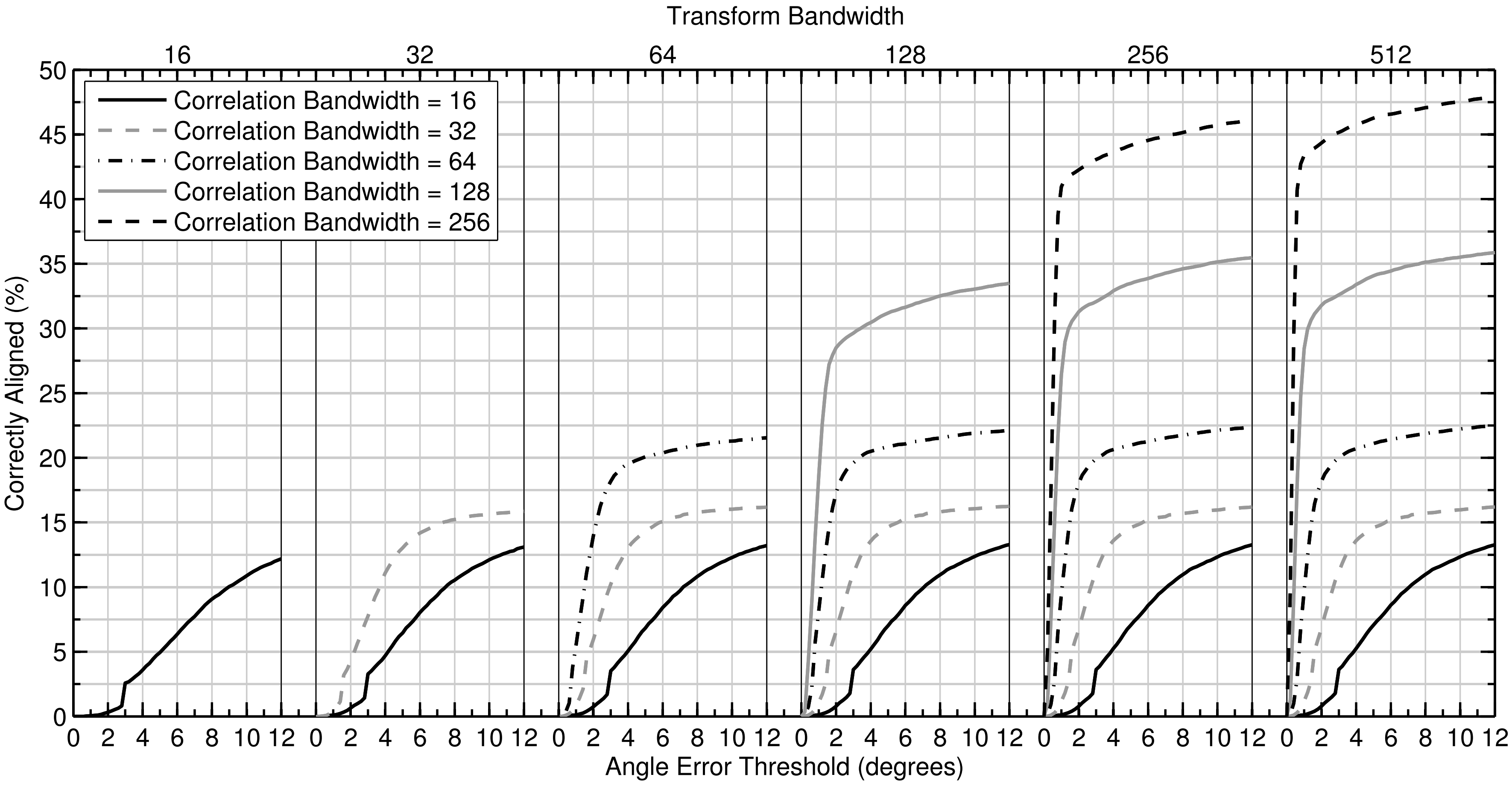


Figure 5.9: Graph comparing how differing transform and correlation bandwidths affect the percentage of correct alignments.

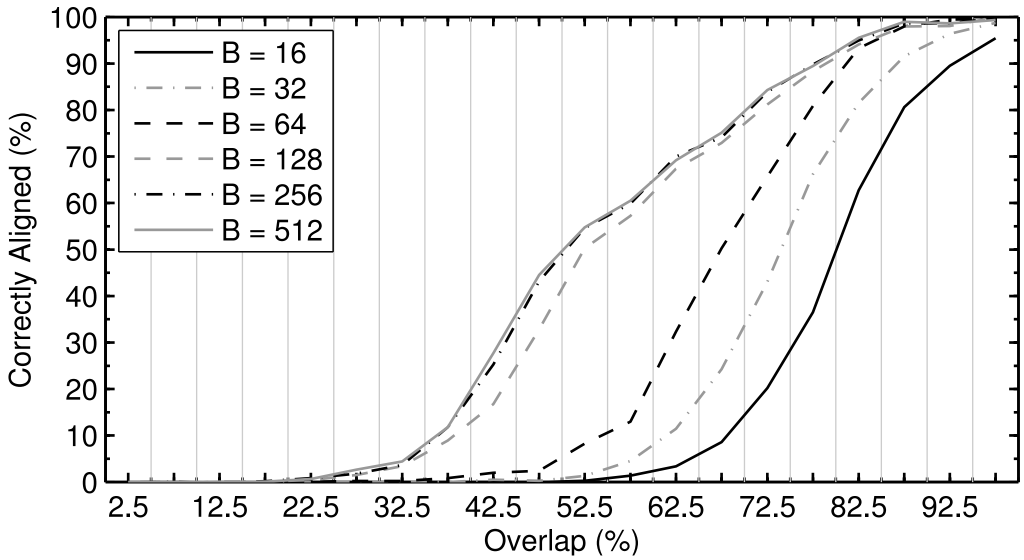


Figure 5.10: The percentage of correctly aligned point-cloud pairs versus their overlap. The overlap is broken into twenty 5% steps, where the correctly aligned percentage is the mean of all point-cloud pairs in a given step. Alignment is deemed correct if its rotation error is 10° or less.

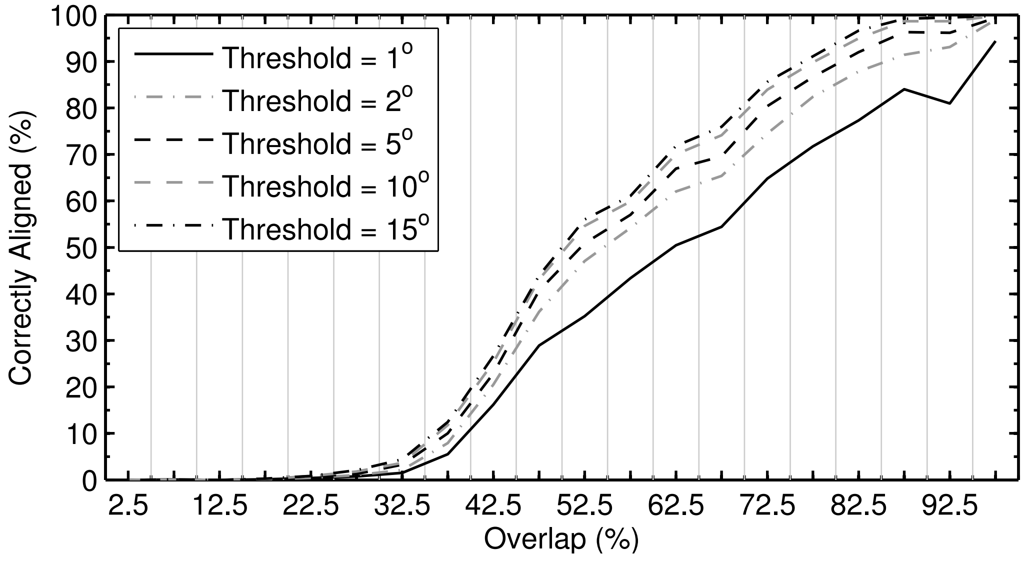


Figure 5.11: The percentage of correctly aligned point-cloud pairs versus their overlap when the transform bandwidth is 256 and the correlation bandwidth is 128. The overlap is broken into twenty 5% steps, where the correctly aligned percentage is the mean of all point-cloud pairs in a given step.

Table 5.2: The percentage of correctly aligned point-cloud pairs at different rotation error thresholds when the normals are orientated with the pole of the equiangle grid and each bin is divided by its surface area.

Bandwidth		Rotation Error Threshold				
Transform	Correlation	1°	2°	5°	10°	15°
8	8	0.0	0.0	0.4	4.6	8.0
16	16	0.1	0.3	4.9	10.9	13.3
32	32	0.4	4.3	13.0	15.6	16.2
64	64	5.5	13.9	20.1	21.3	21.8
128	128	18.6	28.5	31.2	33.1	34.0
256	128	26.4	31.3	33.5	35.2	35.9
512	128	28.4	31.8	34.1	35.5	36.3
1024	128	28.4	31.5	33.7	35.3	36.1
256	256	41.0	42.3	44.2	45.7	46.7
512	256	43.4	44.3	46.3	47.6	48.4

tween bandwidths 64 and 128 is quite distinct, while the difference between 128 and 256 may appear minimal (remembering that the correlation bandwidth is restricted to 128), there is up to 10% improvement in registration performance in the 40% to 50% overlap range. Further increasing the transform bandwidth to 512 barely provides any benefit over 256.

The examination of bandwidth 256 in Figure 5.11 provides insight into how the chosen rotation error threshold permits the alignment of point-cloud pairs to be classified as correct with respect to overlap. Limiting the threshold to 1° restricts many potentially correct alignments, as seen by the consistent 10% improvement when the threshold is 2°. Using the larger thresholds increases the number of correct alignments, but this occurs at diminishing levels due to the discrete nature inherent of the correlation matrix. These results reflect those seen previously in Figure 5.7, in which the majority of correct alignments occur within the initial spike. Figure 5.11 also shows that there is no correlation between the overlap percentage and the rotation error threshold. Therefore, the resulting rotation error does not decrease as overlap increases. The dip that occurs for the 1° threshold at an overlap of 92.5% is explained in the discussion (see Section 5.4).

The minimum overlap capable of achieving correct alignment is presented in Table 5.3, and is shown for each bandwidth at five rotation error thresholds. An initial observation appears to show that the minimum overlap required improves when increasing the rotation error threshold, however, as previously

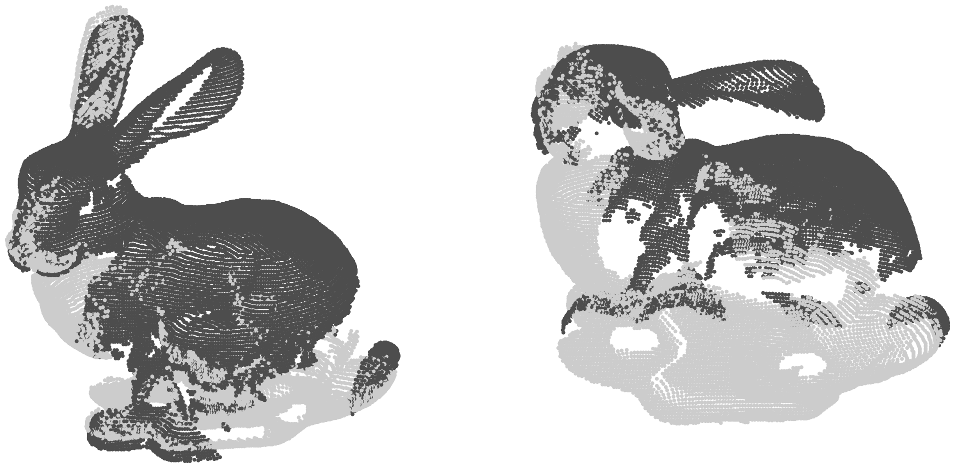
Table 5.3: The minimum overlap necessary to correctly align a point-cloud pair for a range of bandwidths and rotation error thresholds.

Bandwidth		Rotation Error Threshold				
Sampling	Correlation	1°	2°	5°	10°	15°
8	8	–	78.2	75.2	63.4	58.2
16	16	78.4	68.0	60.9	51.4	47.8
32	32	63.9	57.8	40.8	40.8	40.8
64	64	47.2	35.1	13.3	13.3	13.3
128	128	21.5	21.5	19.5	19.5	19.5
256	128	9.2	9.2	9.2	9.2	9.2
512	128	13.7	13.7	13.7	13.7	13.7
1024	128	13.7	13.7	13.7	13.7	13.7
256	256	10.8	10.8	10.8	10.8	10.8
512	256	15.2	15.2	15.2	15.2	15.2

stated, there is no correlation between overlap and the remaining rotation error (when the best available coarse rotation is selected). The apparent improvement is purely due to the number of correctly aligned point-cloud pairs increasing with the rotation error threshold, which in turn increases the likelihood of point-cloud pairs with low overlap being correctly aligned. The results in Table 5.3 also reveal that increasing the two bandwidths will initially decrease the required overlap, but this does not remain consistent when they are greater than 128. Figure 5.12 shows for each of the three datasets an example of the minimum overlap required to correctly align a point-cloud pair. Due to their differing shapes, the Stanford bunny requires the least amount of overlap at 9.2%, followed by the Dragon and Buddha statue at 27.7% and 35.6% overlap, respectively.

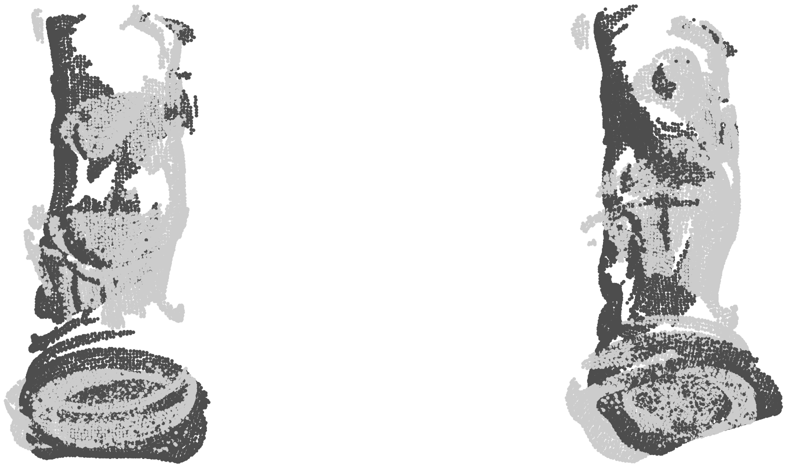
5.3.2 Icosahedron and Fibonacci Accuracy

The icosahedron and Fibonacci binning schemes improve the uniformity of the bins across the unit sphere, allowing them to provide a better representation of normals. Comparing these two schemes with equiangle binning determines if their near uniform binning provides any advantage. Due to the high computational cost of the least-square method of converting to spherical harmonics, the bandwidth is limited to 64. Figure 5.13 shows that at all presented bandwidths, the icosahedron and Fibonacci schemes get a higher percentage of correct alignments at lower rotation error thresholds than equiangle. However, as



(a) Stanford Bunny; 9.2% overlap.

(b) Dragon; 27.7% overlap.



(c) Buddha Statue; 35.6% overlap.

Figure 5.12: Examples of the minimum overlap needed to correctly align point-clouds for each of the three models used from the Stanford 3D Repository. Each point-cloud pair alignment is shown from two poses to help reveal the total extent of overlap.

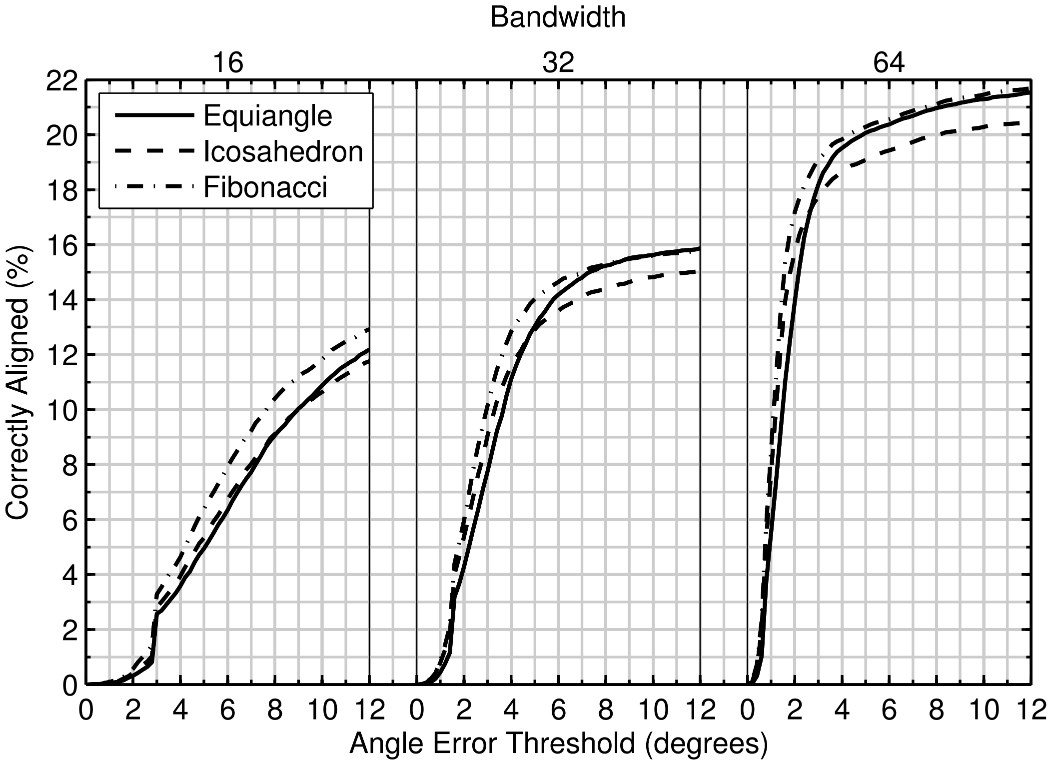


Figure 5.13: The percentage of point-cloud pairs that each of the three binning approaches (equiangle, icosahedron and Fibonacci) correctly aligns. Equiangle binning is performed at the pole and each bin is divided by its surface area.

the rotation error threshold increases, equiangle binning surpasses icosahedron and comes close to matching Fibonacci binning. At a rotation error threshold of 10°, Fibonacci has an approximate 1% improvement on equiangle binning when the bandwidth is 16. Increasing the bandwidth provides minimal advantage between equiangle and Fibonacci. The icosahedron scheme tapers off early, with a 1% decrease in performance.

5.3.3 Efficiency

The goal of using an alternative binning scheme to equiangle binning is to improve registration accuracy, but accuracy is not the only metric which must be considered. To justify using these alternative binning schemes, the cost to efficiency should be taken into account. The cost of binning each normal is a measure of total GCD calculations (see Equation (5.14)) that must be performed. Figure 5.14 shows the GCD cost for binning each normal with relation to the bin count. Equiangle binning does not require any GCD calculations as it bins each normal directly. Equiangle is only shown with bin counts at powers of two, as the bin count is determined by bandwidth; the bandwidth can be any natural number, but the fast spherical-harmonic algorithm (see Section 3.2.2) is most efficient at powers of two.

The cost of the Fibonacci binning scheme is highly dependent upon the location of the normal with respect to the surrounding bins, as this determines the number of bins that must be checked before the closest bin is confirmed. The line showing the Fibonacci cost in Figure 5.14 is the mean number of GCD calculations performed when binning a normal. The fluctuations in this line are due to the changing relationship between the normals and bins, which was produced by creating and binning 10 000 random normals, with the bin count being increased in steps of 200. Fibonacci binning is presented as a line as its bin count can be any odd number.

The cost of binning normals using the icosahedron scheme is a fixed value based on the desired depth, where if d is the depth, the required number of GCD calculations is given as $10 + 4d$. Because the icosahedron bins are formed by successively splitting triangles, the bin count is very restricted, as shown in Figure 5.14 by the six marked icosahedron bin counts. The cost of performing icosahedron binning tapers off as bin count increases, with Fibonacci binning having the same mean performance when there are approximately 20 000 bins. Depending on the individual location of a normal, the GCD cost of Fibonacci can be much greater than the mean; at a bin count of 20 001, the maximum number of GCD calculations for a normal was 129. Overall, equiangle binning is the most efficient scheme as it does not require an iterative algorithm to bin each normal.

5.3.4 Equiangle Noise

When an imaging system captures range data, the produced point-cloud is rarely free from noise. As this noise affects individual points, then their calculated normals will also be noisy. By applying varying levels of noise to the normals, the extent to which spherical-harmonic correlation is capable of achieving correct alignment is determined. Only the equiangle binning scheme is examined with respect to noise. This is due to equiangle binning being able to achieve a greater registration performance than either Fibonacci or icosahedron as it can be executed at bandwidths greater than 64.

The ability of the spherical-harmonic correlation to correctly align point-cloud pairs at six different levels of noise is shown in Figure 5.15. No improvement is gained by increasing the bandwidth when the noise is 5° or 10°, while

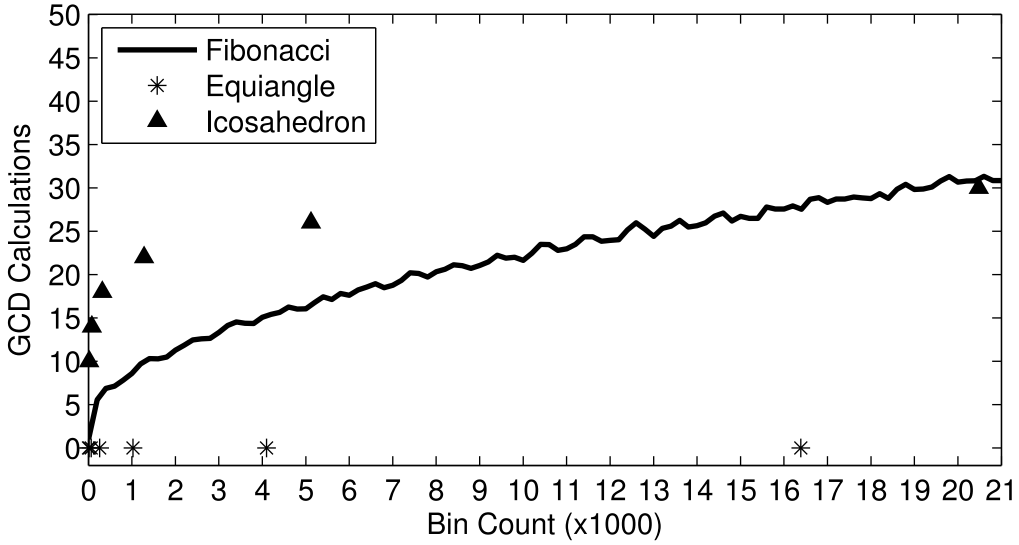


Figure 5.14: The number of great circle distance (GCD) calculations that are performed to bin an individual normal at different bin counts.

at the lower noise levels, increasing the bandwidth to 128 produces a minimum of 2% improvement. By further increasing the transform bandwidth to 256, the 0.5° and 1° noise levels see an approximate 1% improvement.

The greater the noise contamination, the higher the overlap be between point-clouds needs to be for correct alignment. Figure 5.16 shows this relationship for a transform bandwidth of 256 and a correlation bandwidth of 128. It can be seen that as the noise level increases, correct alignments rapidly decrease as overlap decreases. The extent of this decrease is that even adding 1° of noise results in the average overlap needing to be 10% higher for it to match the registration ability when no noise is added. Unexpectedly, having a slight amount of noise improves registration when there is 70% to 90% overlap.

5.4 Discussion

This section discusses in greater detail the individual findings revealed in the previous results section. It is broken up into three sections: accuracy, efficiency and noise.

5.4.1 Accuracy

One of the key aspects of this analysis was quantifying the effect that bin surface area and shape has on registration accuracy. Using an equiangle grid has

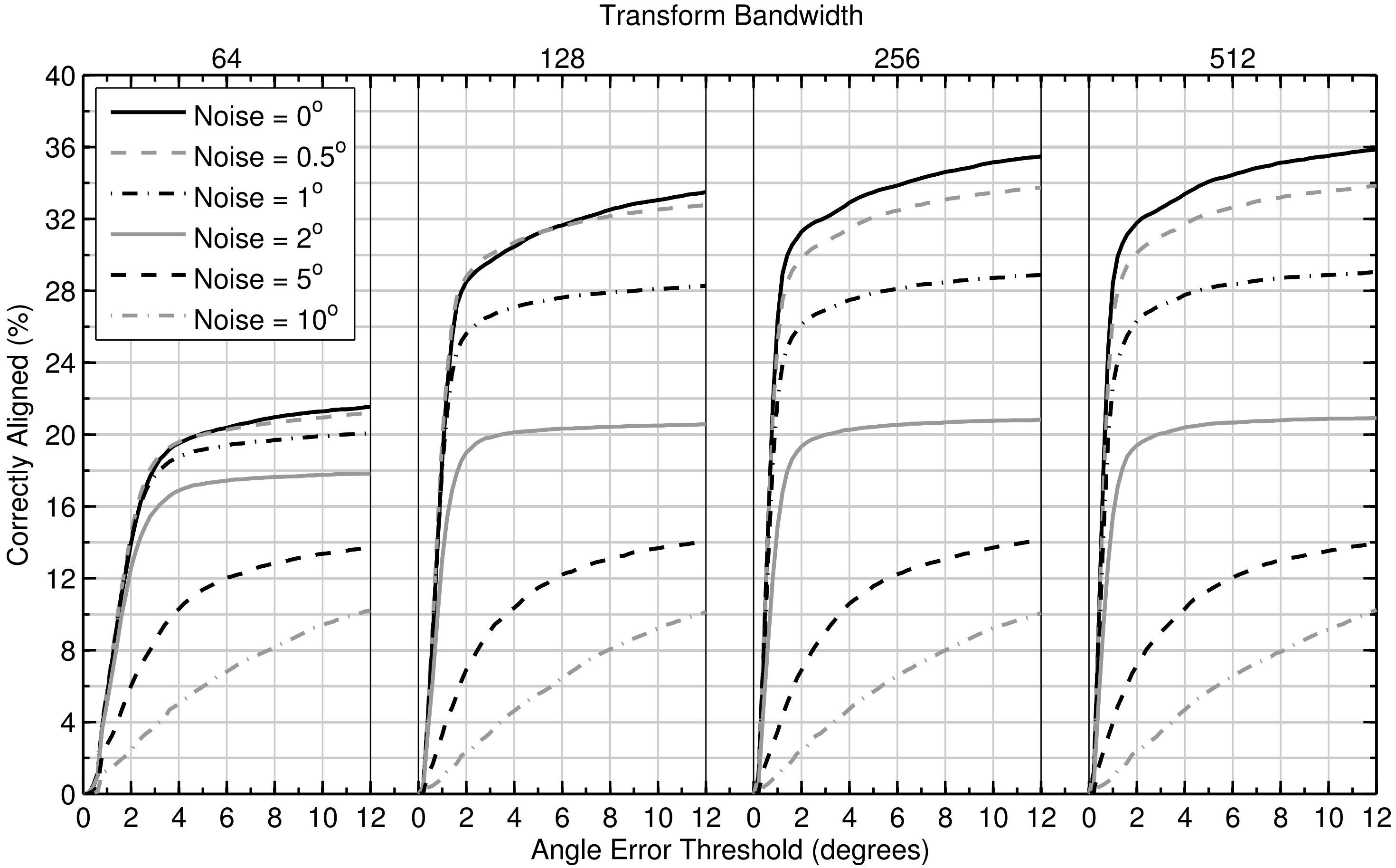


Figure 5.15: The percentage of correctly aligned point-cloud pairs where each normal is contaminated with Gaussian noise. The amount of Gaussian noise applied to each normal is determined by an angle which specifies one standard deviation from the normals initial position.

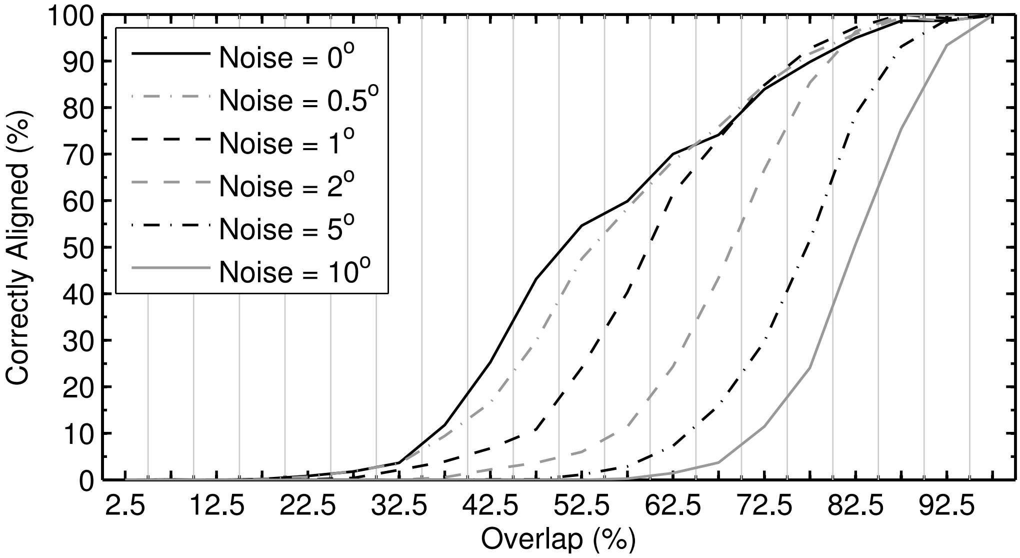


Figure 5.16: The percentage of correctly aligned point-cloud pairs versus their overlap. The graph lines are shown for a variety of noise levels used to contaminate the normals. The overlap is broken into twenty 5% steps, where the correctly aligned percentage is the mean of all point-cloud pairs in a given step. Alignment is deemed correct if its rotation error is 10° or less.

the primary advantage of allowing an efficient conversion to spherical harmonics, but both the size and shape of each bin changes with respect to its polar angle. Dividing the count of normals that each bin collects by the size of its surface area improves performance, as this more accurately reflects the true collecting power of each bin. If surface area is not incorporated, then there is bias towards the larger bins.

The icosahedron and Fibonacci binning schemes provide a more uniform distribution of bins about the unit sphere, producing individual bins which have a similar size and shape. Due to this, the bins were not normalised by their size, as this is a non-trivial task in itself, especially for Fibonacci binning. If these bins were normalised, there may be a marginal improvement, but it is unlikely to be significant. The main benefit that these two binning schemes provided was allowing more point-cloud pairs to achieve correct alignment at a lower rotation error. However, as their bandwidth is limited to 64 (due to computational cost), equiangle has a similar performance if a marginally higher rotation error is permitted. At higher correlation bandwidths, the equiangle rotation error is less than 2° for the initial spike of aligned point-cloud pairs.

Makadia *et al.* (2006) noted that the best binning histogram is one comprised of bins that have the same surface area and shape. Their goal was to create a fast algorithm as opposed to improved accuracy. The results shown in this chapter prove that using a binning scheme with a more uniform distribution of bins does provide a minimal improvement to registration accuracy. However, equiangle binning can be kept computationally more efficient as its bandwidth is increased, thus producing a better registration accuracy. Furthermore, Makadia *et al.* (2006) do not rescale the equiangle bins, meaning their registration accuracy will have suffered. Their algorithm attempts to overcome any incorrect alignments by weighting the normals and adding a verification step; both of which are explored in Chapters 6 and 8, respectively.

One result that must be noted is that the icosahedron binning scheme tapers off at a lower alignment percentage than either Fibonacci or equiangle binning (shown in Figure 5.13). The exact cause for this has not been identified, but may be due to two reasons. The first is that icosahedron bins do differ in size; bins closer to the centre of an icosahedron triangle are larger, as the distance to project them onto the sphere surface is greater. The second potential cause is that the icosahedron, due to the manner in which its bins are created, produces more bins per bandwidth than the other two schemes. By having more bins than necessary for performing the least-squares conversion to spherical harmonics, their representation of normals in the frequency domain may be adversely affected.

The primary goal of performing this analysis was to determine the capability of different binning schemes for registering range images acquired from independent poses. As each range image is from a single fixed pose, the extracted normals only occur on a hemisphere of the total unit sphere. The orientation of the normals with respect to the equiangular grid affects the registration performance. If the pole of the equiangle grid is aligned with the optical axis of the imaging device, then there is a greater density of bins for collecting and representing normals. Additionally, the larger bins that are near the equator will collect normals which are noisier or less useful, especially as the surface they are calculated from trends towards being parallel with the optical axis. By increasing the bandwidth to 128, the effect of orientation diminishes (when bins are divided by their area size), suggesting that as long as the bandwidth is high enough, sampling on the equator is sufficiently represents the distribution of normals in the frequency domain. This improvement is attributed to the overall increase in bin density. Given that a high bandwidth is used, the orientation of normals with respect to the equiangle grid has little influence on registration. Therefore, using imaging devices that capture a larger field-of-view should have an equal level of performance (as long as the imaged scene is static).

When determining the best correlation between two sets of normals, the entire rotational space is searched (albeit in a discrete manner), therefore, the angular difference between a point-cloud pair is irrelevant. What is important is the percentage of overlap between point-cloud pairs, as this increases the similarity between the two sets of spherical harmonics, and in turn the correlation value of the correct rotation, allowing it to be identified. It may be expected that overlap would intrinsically be related to the angular difference between poses, but this relationship does not exist due to two reasons. The first is that a scene structure can have aspects which introduce occlusions when imaged from one pose, obscuring a large portion of the scene that may become visible even with a minimal shift in imaging pose. When a large portion of the scene is revealed, the percentage of overlap can change considerably, but as the imaging pose only had a minimal shift, the angular difference remains minimal. The second reason is that an imaging device can be rotated about its optical axis, which can introduce an angular difference of up to 180°, while still maintaining an overlap percentage that is near 100%.

While increasing the overlap between point-cloud pairs is one method for improving registration accuracy, the key benefit comes from increasing both the transform and correlation bandwidths. The transform bandwidth defines the level of representation of the normals in the frequency domain. The higher the transform bandwidth, the better this representation will be due to two reasons. The first is that as the transform bandwidth goes up, so does the bin density, thus the surface of the unit sphere is sampled in finer increments. This provides more detail of the overall distribution of normals and how they are related to each other once binned. This increased bin density benefits all calculated spherical harmonics. The second is that the transform bandwidth is the maximum spherical-harmonic degree used to express the normals in the frequency domain. The correlation bandwidth stipulates both the size and the maximum degree of spherical harmonics used for forming the correlation matrix; because of this, the transform bandwidth must be equal to or greater than the correlation bandwidth. By having a transform bandwidth greater than the correlation bandwidth, the registration accuracy does improve as the bin density increases and the harmonic aliasing is reduced. This improvement does taper off as the transform bandwidth goes up, which shows that the effect of aliasing reduces, and the sampling density on the sphere is able to adequately represent the normals.

Varying the bandwidths produced two unexpected outcomes. The first was that registration accuracy diminished when the transform bandwidth (1 024) was much greater than the correlation bandwidth (128), though this decrease was insignificant. The second was when the bandwidths were greater than 128, the minimum overlap capable of achieving alignment was inconsistent. This is because when the overlap is low, there are a variety of potential alignments that maximise correlation. Combining this with the fact that the representation of the normals and the available discrete rotations both change with different bandwidths, one particular alignment will produce the maximum correlation. It is just coincidental if a lower bandwidth pair produces the correlation that has a rotation closer to the true rotation between two point-clouds.

The registration accuracy could be further improved if greater correlation bandwidths are computed, as there is no indication that a bandwidth of 256 sufficiently represents the normals. Lower bandwidths can be used to perform the registration if the two sets of normal distributions provide a representation that is distinct enough for identifying the correct rotation. This however comes at a cost, as there less rotations to select from, causing larger rotation errors to occur. Therefore, when performing spherical-harmonic correlation, using the highest feasible combination of transform and correlation bandwidths is recommended as this helps maximise registration accuracy and minimise the resulting rotation error.

The rotation error that remains after rotational registration is not related to the overlap between point-cloud pairs. This is due to there only being a discrete set of rotations, where even if the overlap is large, the most suitable rotation may result in a larger rotation error. Because there is no relationship, the opposite can occur, where minimal overlap may produce a rotation error that is near zero. It is due to this reason, combined with the low number of point-cloud pairs at 92.5% overlap (shown in Figure 4.2), that Figure 5.11 has a dip in correct alignments at 92.5% overlap for the 1° threshold. This assessment is reinforced by the fact that this dip is much less apparent when the threshold is 2°.

Knowing how well the spherical-harmonic correlation performs to within a given rotation error, when a particular set of parameters are used, provides an indication of how well a fine registration algorithm will perform. This is because fine registration algorithms, such as ICP, are limited to a maximum rotational misalignment that they can handle. As ICP is the primary algorithm used, a preliminary study (Larkins *et al.*, 2010) was performed which found

ICP capable of registering up to 10° of rotational misalignment. This provides enough of a buffer for spherical-harmonic correlation, which at a correlation bandwidth of 128 achieves correct rotational alignments within 2°.

5.4.2 Efficiency

The process of binning normals is a step that must be performed, as attempting to represent every individual normal in spherical harmonics by the way of least-squares will be cost prohibitive. In terms of the most efficient binning scheme, the equiangle grid has the best performance as it is able to directly bin each normal without having to perform any form of search. The icosahedron and Fibonacci schemes both require searching to identify the closest bin to a normal. Therefore, they are incapable of achieving the same level of efficiency; unless of course there is an underlying aspect to their formation that permits direct binning. From the results shown, the benefit to accuracy of using icosahedron and Fibonacci is minimal, therefore the equiangle grid gives the best combination of efficiency and accuracy.

The computational cost of identify which bin collects a normal is marginal with respect to the computation required for converting the bins to spherical-harmonics. Using least-squares permits any point-distribution to achieve this conversion, but for a specified bandwidth, there is a preferred number of sample points ($4B_t^2$), otherwise the fitting becomes underdetermined or overdetermined. Singular value decomposition can handle both of these situations, but its ability to do so was not tested. Because Fibonacci can have any odd number of bins, the additional cost of it forming an overdetermined system is minimal, whereas icosahedron does not have these fine increments, further increasing its computational cost.

Using least-squares is an expensive means for converting a distribution of points on a sphere to spherical harmonics. Optimal conversion occurs when the points have constant latitude, as this means that the associated Legendre polynomial only needs to be calculated once for many points. Neither icosahedron nor Fibonacci meet this requirement, and therefore they are not conducive to the fast spherical-harmonic transforms implementations. This is the primary advantage of the equiangle grid when it comes to efficiency, as it allows the conversion to be performed significantly faster. A more detailed overview of the computational costs are covered in Section 3.2.2.

5.4.3 Noise

When noise corrupts a normal, shifting it from what would be its true position, an adjacent bin may collect it. For equiangle binning, the higher the bandwidth, the greater the occurrence of incorrect binning, especially if the normal is near the pole. However, the results presented in Figure 5.15 show that while registration accuracy does decrease with noise, this decrease is not related to the sampling density of the transform bandwidth, as performance remains consistent. Equiangle binning was the only scheme tested with noise, the capability of icosahedron and Fibonacci for handling noise was not determined. These two schemes are unlikely to provide any benefit as the combination of increased sampling density and noise does not decrease registration accuracy. The ability of binning to deal with noise occurs when bin size is larger than the noise discrepancy, which would keep the normal in the bin. Ensuring a bin is large enough would require a lower bandwidth; which provides no benefit as a higher bandwidth maintains or improves accuracy even when noise is present.

An unexpected result that occurred in Figure 5.16 was that the percentage of correctly aligned point-cloud pairs increased between the 70% to 90% overlap range when a marginal level of noise was present. The exact reason why noise boosted the alignment percentage is unknown, though it may be reducing the effect of quantisation error. This boost however does not improve the overall registration ability, especially considering that when the overlap is below 70%, the performance drops more rapidly when noise is present as opposed to no noise.

5.5 Summary

The initial hypothesis was that if normals are placed into bins that are both more uniformly distributed on a sphere and maintain a similar size, then a more accurate representation of the normals will be formed, thus improving registration accuracy. The experiments and analysis performed show that this hypothesis proves to be somewhat true, as icosahedron and Fibonacci binning schemes are able to align more point-cloud pairs at a lower rotation error than equiangle binning. Achieving alignments which minimises rotation error is advantageous for fine registration algorithms that are only capable of handling small rotational misalignments. By further evaluating equiangle binning, it was seen that if a higher rotation error is permitted, the performance of equiangle surpassed icosahedron and was similar to Fibonacci.

The fundamental advantage of equiangle binning over icosahedron and Fibonacci comes from its efficient conversion to spherical harmonics. This allows equiangle to make use of higher bandwidths, increasing the total number of spherical harmonics used. Being able to perform both the sampling on a sphere and the spherical-harmonic inversion at higher bandwidths was shown to provide a significant benefit to registration accuracy. Due to this, equiangle binning provides a sufficient means of achieving rotational alignment when given a pair of overlapping point-clouds.

An additional aspect of this analysis was to determine whether bin size and the orientation of normals are factors that should be taken into account when using equiangle binning. From this, it was found that both of these factors have a large influence on registration ability. The best performing combination was when the normals were binned at the pole of the equiangle grid, and each bin's weight was scaled by dividing its count of normals by the area it encompasses on the sphere surface.

The final area of investigation was quantifying how robust equiangle binning is when normals are contaminated by noise. This was measured by examining the registration performance achieved by spherical-harmonic correlation. The results showed that the noise level adversely affected performance, especially as point-cloud overlap decreases. At high levels of noise, the area of investigation should not be the registration algorithm, but on the capturing and pre-processing steps, so as to improve point-cloud quality.

Chapter 6

Weighting Normals

Treating all normals equally in preparation for performing spherical-harmonic correlation was shown in Chapter 5 to be capable of bringing 35% of the point-cloud pairs into correct rotational alignment. However, each normal does not provide an equal contribution in terms of allowing the correlation to determine the correct alignment; this is due to two primary reasons. Firstly, there is no guarantee that the surfaces common to two overlapping point-clouds have the same sampling density or size, meaning that the number of normals between them will differ. Each surface in a point-cloud is a contiguous section of points that maintains a consistent level of surface curvature; the pose of the imaging system and the nature of the scene will determine the size, shape and number of individual surfaces that constitute the point-cloud. However, identifying surfaces based on contiguous points is not pertinent as only the normals are used in the correlation. Because of this limitation, surfaces are identified by clusters of normals that have a similar orientation and are subsequently set to have an equivalent weighting. Secondly, a normal can be contaminated by noise, in particular, it is unreliable if it is collected from areas with high surface curvature or from surfaces that are nearly parallel to the optical axis of the imaging system. Weighting normals based on their surrounding surface curvature allows noisy normals to be identified and culled, improving the overall quality of the normals used in the correlation.

This chapter investigates how registration accuracy is improved by mitigating the effect of unreliable normals. This is achieved by exploring weighting schemes that are designed to emphasise normals that are beneficial to the registration, and de-emphasise those that are not. The goal of this investigation is to determine if weighting schemes provide an advantage to registration, and if so, which scheme is the most conducive to spherical-harmonic correlation.

Makadia *et al.* (2006) proposed that equiangle bins deemed local maxima should be set to a value of one, with the rest set to zero. This scheme attempts to normalise each surface of a point-cloud so that their size is not taken into account. Because equiangle bins are not equal in size, the idea of setting bins to a single fixed value (as opposed to normalising them based on bin size) is suboptimal. The weighting scheme by Makadia *et al.* is built upon by evaluating how alternative bin values enhance the registration accuracy. Fibonacci binning is also used to determine if its incorporation will alleviate issues that arise with the equiangle approach. Weighting individual normals, as opposed to reweighting each histogram bin, is an additional weighting scheme that is introduced to help minimise the inclusion of detrimental normals. Because these two weighting schemes target different issues, their combination further benefits registration accuracy. This is accomplished by introducing the idea of transforming complex values into spherical harmonics instead of simply using real values.

6.1 Weighting Schemes

Incorporating weighting into the correlation process emphasises preferred properties contained within a point-cloud. If these properties are similar between two point-clouds, then their emphasis means that the correlation is better able to identify the rotation that correctly brings these two point-clouds into rotational alignment. Two types of weighting schemes are investigated, the first gives each normal its own weight, while the second weights all the surfaces of a point-cloud equally, regardless of their size. A means of combining these two weightings is then introduced, providing the benefit of both schemes.

6.1.1 Weighting Normals

Each normal when binned is initially treated as having an equivalent weight, that is, they each have a weight of one. The weight attributed to a given normal can be modified based on a variety of aspects; here, the local surface curvature about a point is used. The weight is normalised to a value between zero and one, allowing each weight to be dealt with in the same manner, where zero is the maximum surface curvature and one is a perfectly flat surface. Because a normal is the first-order derivative of a surface, it only describes the surface orientation at the corresponding point, whereas surface curvature is the second-order derivative. Surface derivatives at a point are calculated based on its neighbouring points; this is inclusive of normals being used to calculate surface curvature, as normals are also calculated from the points. The number of neighbours that the current point has is given by the variable n_c. Two approaches of measuring surface curvature are tested. The first uses the angular difference between a normal and its neighbours; a similar approach has been used previously by Jiang *et al.* (2009) to allow an iterative based registration algorithm to be formed. The second measure, introduced here, calculates the perpendicular distance that points lie with respect to the tangent plane fitted to the current point.

The surface curvature weighting can be used in a variety of ways, but it is simply used here to determine if a normal should be kept or culled. It is hypothesised that points residing near edges or corners in a scene produce normals that *flare* out in unexpected directions, particularly if the imaging system has sparse sampling or if its 3-D acquisition process is prone to noise. As such, these normals are treated as noise and culled. Because the weights are normalised, a threshold, q, termed the *cull-point*, gives a binary divide, namely the weight ω of the normal used is

$$\omega = \begin{cases} 0 & \omega' < q, \\ 1 & \omega' \geq q; \end{cases} \tag{6.1}$$

where ω' is the calculated surface curvature. If a normal receives a weight of zero, it is culled by not binning it. The cull-point can be specified as an angle that limits how great the local curvature can be about a point. If this limiting angle is λ, the cull-point q is given as

$$q = 1 - \lambda/\pi. \tag{6.2}$$

Angular Difference Between Normals

The angular difference, ψ_j, between the normal $\hat{\mathbf{n}} = [n_x, n_y, n_z]$ and one of its neighbours, $\hat{\mathbf{n}}_j$, gives a measure of the surface curvature, calculated by

$$\psi_j = \cos^{-1}(\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_j), \tag{6.3}$$

where $j = 0, \ldots, n_c - 1$ identifies each of the n_c neighbours, and $\hat{\mathbf{n}} \cdot \hat{\mathbf{n}}_j$ is the dot product between $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}_j$. Normals expressed in Cartesian coordinates can be re-expressed in spherical coordinates by

$$\hat{\mathbf{n}}_{\theta\phi} = \left[\cos^{-1}(n_z),\ \tan^{-1}\left(\frac{n_y}{n_x}\right)\right]. \tag{6.4}$$

The angular difference between two normals expressed in spherical coordinates, namely $\hat{\mathbf{n}} = [\theta,\, \phi]$ and $\hat{\mathbf{n}}_j = [\theta_j,\, \phi_j]$, is given by

$$\psi_j = \cos^{-1}\left(\sin\theta \sin\theta_j \cos\left(\phi - \phi_j\right) + \cos\theta \cos\theta_j\right). \tag{6.5}$$

The weight, ω', of the normal $\hat{\mathbf{n}}$ is given by

$$\omega' = 1 - \frac{1}{\pi n_c} \sum_{j=0}^{n_c - 1} \psi_j, \tag{6.6}$$

where the summation is over all the angular differences between $\hat{\mathbf{n}}$ and its neighbours. Note that $\omega' \in [0, 1]$, with $\omega' = 1$ when the normal and all its neighbours point in exactly the same direction, otherwise it is smaller.

Plane Distance

Let P be the plane with normal $\hat{\mathbf{n}}$ and that passes through the point $\mathbf{p}$. A surface curvature weighting for this point is calculated by taking the mean perpendicular distance between P and the points neighbouring $\mathbf{p}$. A neighbouring point, $\mathbf{p}_j$, is a point in the plane if

$$\hat{\mathbf{n}} \cdot (\mathbf{p}_j - \mathbf{p}) = 0. \tag{6.7}$$

If $\mathbf{p}_j$ is not in the plane, then

$$\hat{\mathbf{n}} \cdot (\mathbf{p}_j - \mathbf{p}) \tag{6.8}$$

gives the perpendicular Euclidean distance of $\mathbf{p}_j$ to P, with the sign positive if $\mathbf{p}_j$ is on the same side of P as the direction given by $\hat{\mathbf{n}}$, and negative otherwise. This distance gives a measure of the surface curvature between $\mathbf{p}$ and $\mathbf{p}_j$, though if it is not normalised, the same surface at a different scale will give a different weighting. Therefore, the distance is normalised based on the separation between $\mathbf{p}$ and $\mathbf{p}_j$, namely by

$$\frac{\hat{\mathbf{n}} \cdot (\mathbf{p}_j - \mathbf{p})}{||\mathbf{p}_j - \mathbf{p}||}. \tag{6.9}$$

The calculated weight is the mean of the normalised distances to each neighbour, that is,

$$\omega' = 1 - \left| \frac{1}{n_c} \sum_{j=0}^{n_c-1} \frac{\hat{\mathbf{n}} \cdot (\mathbf{p}_j - \mathbf{p})}{||\mathbf{p}_j - \mathbf{p}||} \right| ; \tag{6.10}$$

taking the absolute value of the mean ensures that surface curvature weighting remains consistent no matter what side of the plane the points are located.

6.1.2 Reweighted Bins

Image acquisition of a scene will typically sample the same surface multiple times. As the overlap between point-clouds is a major factor that influences registration, it is desirable to have a similar number of samples from surfaces common to both point-clouds. This is because the correlation uses the two distributions of normals from the two point-clouds to identify a rotation. If the number of samples, thus the number of normals, differs for the same surface between point-clouds, then the two point-clouds have a different representation of the same surface when the correlation is performed. This occurs for a variety of reasons, but is primarily due to part of the surface in one of the acquisitions being either obscured or outside the field-of-view of the imaging system.

Changes in the number of normals on a surface between acquisitions can be mitigated by naïvely assuming that normals clustered together represent a single surface and should only be counted once; even though in reality a cluster of normals may constitute multiple surfaces. This is because multiple surfaces can have normals with similar orientations, or similarly, a curved surface can produce normals with a variety of different orientations; they are therefore treated as a single surface. Despite these failings, this approach attempts to make clusters of normals equal, allowing the correlation to be performed with a single representation of a surface's orientation, minimising the effect that surface size and changes in sampling density will have on the correlation. Figure 6.1 demonstrates how different camera poses affect sampling distribution, which can lead to an incorrect rotation being identified.

Determining which normals to cluster together can be achieved by binning them and setting the weight of any bin which has collected a normal to a value that appropriately represents the binned normals; four values are evaluated as part of this investigation. The equiangle and Fibonacci binning approaches, shown previously in Figures 5.1a and 5.5a, respectively, are used for this task. However, one aspect that must be taken into account is that curved surfaces

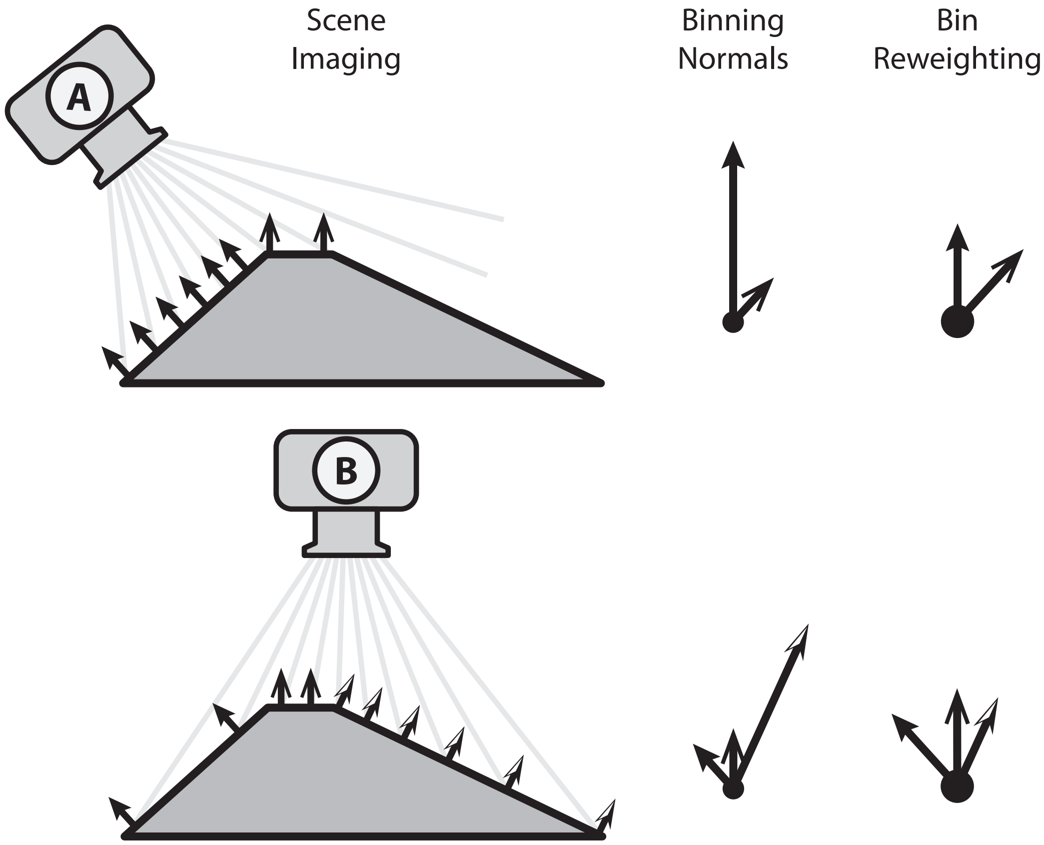


Figure 6.1: Demonstration of how the sampling distribution of three surfaces in a scene change as the camera pose changes. The surfaces which camera A and B sample more densely depends on their pose, thus affecting the number of normals that are binned at each orientation. These changes in sampling distribution can cause the correlation to match two separate surfaces, as they appear similar, where as if each surface is reweighted equally, the correct rotation has a better chance of being distinguished.

or surface edges can distribute normals into a wide array of orientations. The result of this distribution is that individual bins may only collect normals that are effectively noise. As all bins that contain normals are reweighted, incorporating a threshold that stipulates the minimum number of normals required in a bin helps alleviate the effect of this noise, which in turn only allows bins that represent prominent surface areas to contribute to the correlation. While this threshold could be set as a fixed value, this is not suitable as the total number of normals varies with pose, creating different sampling densities. The approach used here to help make the threshold invariant to the total number of normals is to set it as a percentage of this total.

Equiangle

The percentage of normals that each equiangle bin must collect for it to contribute to the SHT is dependent upon its size, but because each bin is normalised based on its surface area, a single threshold can be calculated. This threshold, t, is the percentage of normals that a bin at a particular polar angle will likely collect. By specifying this bin as the smallest bin, the threshold is given by

$$t = \frac{np_{B_t}}{A_{B_t}(0)}, \tag{6.11}$$

where n is the number of normals, p_{B_t} is the percentage of normals to collect at a given transform bandwidth B_t, and $A_{B_t}(0)$ is the area of the smallest equiangle bin (as calculated by equation (5.7)). The bins equal to or greater than t then influence the correlation. If the sampling density increases, there will likely be a corresponding increase in the number of noisy normals from the varying noise sources. Using a dynamic threshold based on the percentage means that the threshold remains applicable if the sampling density changes. Unfortunately using a percentage value requires it to be varied depending on the transform bandwidth. This is because the number of equiangle bins and thus their size are dependent on the transform bandwidth.

Let the sampled value of each equiangle bin be f_{jk}, and the reweighted bin be, f'_{jk}, as determined by the threshold t. Four reweighting schemes for equiangle binning are investigated, with all of them setting any bins that have a value below the threshold to zero. The first scheme does not modify the bins that are equal to or greater than the threshold, and is given by

$$f'_{jk} = \begin{cases} f_{jk} & f_{jk} \geq t, \\ 0 & f_{jk} < t. \end{cases} \tag{6.12}$$

The second scheme simply gives each bin a value of one, making all bins equal if their value is large enough,

$$f'_{jk} = \begin{cases} 1 & f_{jk} \geq t, \\ 0 & f_{jk} < t. \end{cases} \tag{6.13}$$

This scheme is equivalent to that of Makadia *et al.* (2006), thus allowing the performance of their algorithm to be compared against. The remaining two schemes reweight each bin based on their size, as equiangle bin size varies with polar angle; they are given as

$$f'_{jk} = \begin{cases} \dfrac{1}{A_{B_t}(j)} & f_{jk} \geq t, \\ 0 & f_{jk} < t, \end{cases} \tag{6.14}$$

and

$$f'_{jk} = \begin{cases} A_{B_t}(j) & f_{jk} \geq t, \\ 0 & f_{jk} < t, \end{cases} \tag{6.15}$$

which is a weight of one normalised by area, and the bins area, respectively.

Fibonacci

The disadvantage of reweighting equiangle bins is that a suitable percentage value is dependent on the transform bandwidth. A solution that avoids changing the percentage value is to bin the normals using the Fibonacci binning scheme. The Fibonacci bins are instead reweighted, with their bin centres being fed into the equiangle binning scheme. Each equiangle bin then collects the Fibonacci bin centres that fall within its collection area and sums up their reweighted values. This approach allows the number of Fibonacci bins, and thus the percentage value, to remain constant, even if the transform bandwidth of the equiangle binning scheme changes. However, the percentage value is also dependent on the number of Fibonacci bins. The threshold, t, (calculated in a similar fashion to equation (6.11)) determines if a bin or its normals are used, and is given as

$$t = np_b, \tag{6.16}$$

where n is the total number of normals and p_b is the used percentage value for a specific number of Fibonacci bins, b. If the count of normals in a bin is less than t, then this bin is weighted as zero.

Two options are explored once the threshold has been applied to a Fibonacci bin. These are to either use the bin centre in place of the normals, allowing them to be represented by this bin centre, or use only the normals that are in the bins that are not culled. As each Fibonacci bin is approximately equal in size, the expectation is that the benefit to the registration is logarithmic; as the number of bins increases the benefit wanes. The reasoning for this is that the total percentage of surface area a bin encompasses on the unit sphere follows a reciprocal relationship, as shown in Figure 6.2.

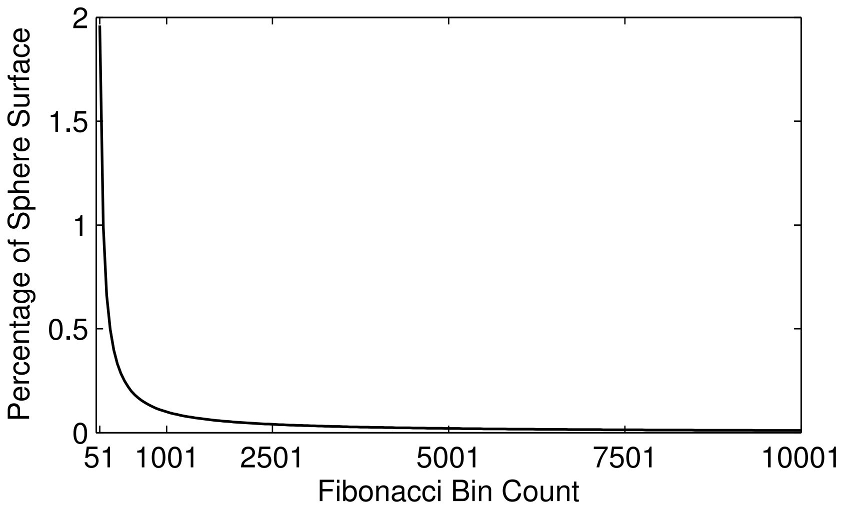


Figure 6.2: The percentage of area that each Fibonacci bin encompasses on the surface of a sphere, with respect to the number of Fibonacci bins. As the number of bins increases, the subsequent difference in the percentage of surface area of each bin decreases.

6.1.3 Complex Value

Transforming the binned normals to the frequency domain only requires a real value, as the value at each sample point is the normalised count of normals collected by the bin. The SHT though, is capable of transforming complex values, thus there is an extra dimension available for encoding additional information about the normals or the point-cloud itself. This supplementary information helps the correlation, increasing its ability in identifying the desired rotational alignment between two point-clouds. However, simply encoding this information using the imaginary component does not work as this modifies the magnitude of the complex value. Instead, the magnitude, r, is taken as the count of binned normals, with the phase, ρ, of the complex value encoded with some additional information. This does however restrict what can be encoded, as the phase value is limited to $[0, 2\pi)$.

The additional aspect investigated here to be encoded in the phase is the surface curvature weighting, ω, attributed to each normal. Because normals are binned, their surface weightings must also be dealt with on a per bin basis. This is achieved by taking the mean weighting of the normals in the bin, $\bar{\omega}$, which is still a value between zero and one. However, if the normals have been culled by their surface curvature weighting using the threshold q, only a small section of the overall phase is used. In this situation, the weighting between q

and one can be *re-spread* between zero and one by

$$\bar{\omega}' = \frac{\bar{\omega} - q}{1 - q}, \tag{6.17}$$

making it the new surface curvature weight for each bin. Converting this weight to a phase value is simply achieved by multiplying it by 2π. One issue with this approach is that as the phase is cyclic, surface curvature weightings of zero and one will appear the same when multiplied by 2π. While this limits the improvement to registration provided by the complex value combination, it is an insignificant issue if registration accuracy increases.

6.2 Methodology

The weighting schemes described in this chapter extend the schemes for binning normals presented in Chapter 5. Due to this, the same Dragon, Buddha statue and Bunny datasets are used as they provide a ground-truth, allowing the exact rotation error to be calculated. Once again, the rotational alignment is deemed correct if the resulting rotation error is less than or equal to a prescribed threshold, otherwise it is incorrect. The rotation error threshold is tested at 0.2° increments from 0° to 12°. In certain situations the rotation error threshold must remain fixed to examine a particular aspect of the registration, if this arises, the threshold is set to 10°. This is chosen as it permits a coarse registration to be classified as correct, while also being within the bounds of what a fine registration algorithm, particularly ICP, is capable of handling (Larkins *et al.*, 2010; Minguez *et al.*, 2006).

The effect that the described weighting schemes have on the registration accuracy is dependent upon the applied parameters. Culling normals based on their surface curvature weighting requires a threshold or cull-point, q, to be specified. As these weightings have a value between zero and one, the threshold is also specified as a value between zero and one; though it can alternatively be given as an angular measure of the surface curvature, with this angle having a value between 0° and 180°. Six cull-points are tested to determine how the registration accuracy varies as greater restrictions on the surface curvature are applied; these cull-points are shown in Table 6.1. Because it is hypothesised that lower surface curvature improves registration accuracy, both the mean angle and plane distance measures are inverted and tested to ensure that this hypothesis is correct.

Table 6.1: The cull-point threshold used to specify which normals are kept or culled based on their surface curvature weighting.

Cull-point	Surface Curvature Angle
0	180°
0.8	36°
0.9	18°
0.95	9°
0.975	4.5°
0.9875	2.25°

Both of these schemes for weighting normals require that the neighbours of a given point be known. If a point is determined to have no neighbours, then its associated normal is also culled. The weighting schemes used in the final comparison have these normals culled, whether it is required or not, so as to ensure that the comparison is fair. Because of this, the results presented in this chapter differ slightly from those presented in the previous chapter.

The manner in which bins are reweighted is dependent upon whether equiangle binning or Fibonacci binning is used, though both binning schemes use a percentage threshold (see Section 6.1.2) to determine whether a bin's value is set to zero or a particular value. Because the number of equiangle bins increases with bandwidth, the size of each bin decreases, therefore three bandwidths, 16, 32, and 64, are tested to identify how the percentage threshold varies, and if this trend is related to bandwidth. Because equiangle bin size changes with the polar angle, four values are tested to identify the optimal value for reweighting the bin. These values are one, bin area, the reciprocal of the bin area, and using only the normals that persist after the percentage threshold has been applied. The Fibonacci bins are near uniform in size, therefore only two reweighting schemes are applied to them. The first feeds the Fibonacci bin centres into equiangle binning if the given bin collects more normals than the percentage threshold. The second approach only feeds the normals into equiangle binning that are collected by the Fibonacci bins with a greater value than the percentage threshold. Both of these approaches are tested using bin counts of 1 001, 5 001 and 10 001, at bandwidths of 16, 32, and 64. Testing at these three bin counts reveals whether registration accuracy improves when increasing bin count, while the selected bandwidths test if the percentage threshold is invariant to bandwidth.

The complex value weighting is a combination of the weighting of normals scheme and the reweighting of bins scheme. The mean weighting value of the normals in a bin is converted into a phase value, while the value attributed to the bin is expressed as the magnitude of the complex value. The experiments using the two schemes separately will identify which set of parameters perform the best; the two schemes with these parameters are used in conjunction with each other to form the complex value. The goal of forming the complex value in this manner is to further boost the registration accuracy, as the complex value provides an orthogonal system for encoding both schemes.

Testing how the best weighting scheme handles noise is performed using the same testing structure established in Chapter 5, in which each normal has Gaussian noise applied to it using the approach outlined in Section 4.2.3. The variability of the noise is specified by a base angle which is set at one standard deviation from the normal: five such base angles are used for introducing noise, with these being 0.5°, 1°, 2°, 5° and 10°. Because the weighting schemes only use the normals, applying noise to the normals in this manner is sufficient for testing their robustness, except for plane-distance. Noise is applied directly to the plane distance weighting, ω, by treating it as an angular value between zero and π (in radians), with the noisy weighting being given as

$$r = \omega + \frac{u\sigma}{\pi}, \tag{6.18}$$

where u is a normally distributed random number with σ specifying the one standard deviation angle. This noisy weighting still needs to be a value between zero and one, therefore, if it is outside these bounds, the value is *reflected* back to maintain this restriction. This reflection is achieved by

$$\omega_r = \begin{cases} 4\pi - r & r > 1, \\ |r| & r < 0, \end{cases} \tag{6.19}$$

with ω_r being the weighting with noise. To maintain continuity, the angle used to introduce noise to the plane distance value will be the same as the angle used to introduce noise to the normals. It must be noted that the noise applied to the normals does differ from that applied to the plane-distance weighting, as both of these noise sources would be symptomatic of sampling noise shifting the points from their true position. The level to which artificial noise is added to the two weighting schemes is more than what a real imaging system should produce, therefore, it provides a suitable evaluation of the correlation robustness.

6.3 Results

The presented results show how different weighting schemes impact the accuracy and efficiency of SHT registration. The robustness of the best performing scheme is tested further by adding varying levels of noise to the normals.

6.3.1 Accuracy

The primary purpose of incorporating weighting schemes into the registration pipeline is to improve the registration accuracy by refining which normals are used or alternatively how they are used. This section investigates how the three described classes of weighting schemes perform when a variety of parameters are applied. These classes are the weighting of normals, reweighting of bins, and combining them together using complex values.

Weighting of Normals

Normals can be weighted using any number of aspects, with the results presented here showing the registration accuracy of two weightings that are produced by measuring surface curvature. These weightings favour flat surfaces, but to ensure that the nature of this weighting is appropriate, their inverse is also tested. Figure 6.3 shows how these four weighting variants perform when one of six thresholds are applied to specify which normals are kept and which are culled. The threshold of zero, deemed the baseline, permits all normals to be kept and binned, with its purpose being purely to demonstrate whether culling provides any benefit to registration accuracy.

The mean angle weighting shows that as the threshold increases, the percentage of correctly aligned point-cloud pairs also increases, with a threshold of 0.975 aligning approximately 10% more point-cloud pairs as opposed to no weighting being applied. However, if the threshold is too restrictive, then there is a detrimental effect on registration, as shown when the threshold is 0.9875. The thresholds perform differently when applied to the plane distance weighting, with the percentage of correctly aligned point-cloud pairs continually increasing with threshold. The highest tested threshold, 0.9875, correctly aligned a further 18% of the point-cloud pairs compared with the baseline; this is the best performing scheme for weighting the normals.

The inverse mean angle results were interesting in that when the threshold was decreased, the registration accuracy increased. Because the inverse mean angle favours surfaces with high curvature, its results were based purely

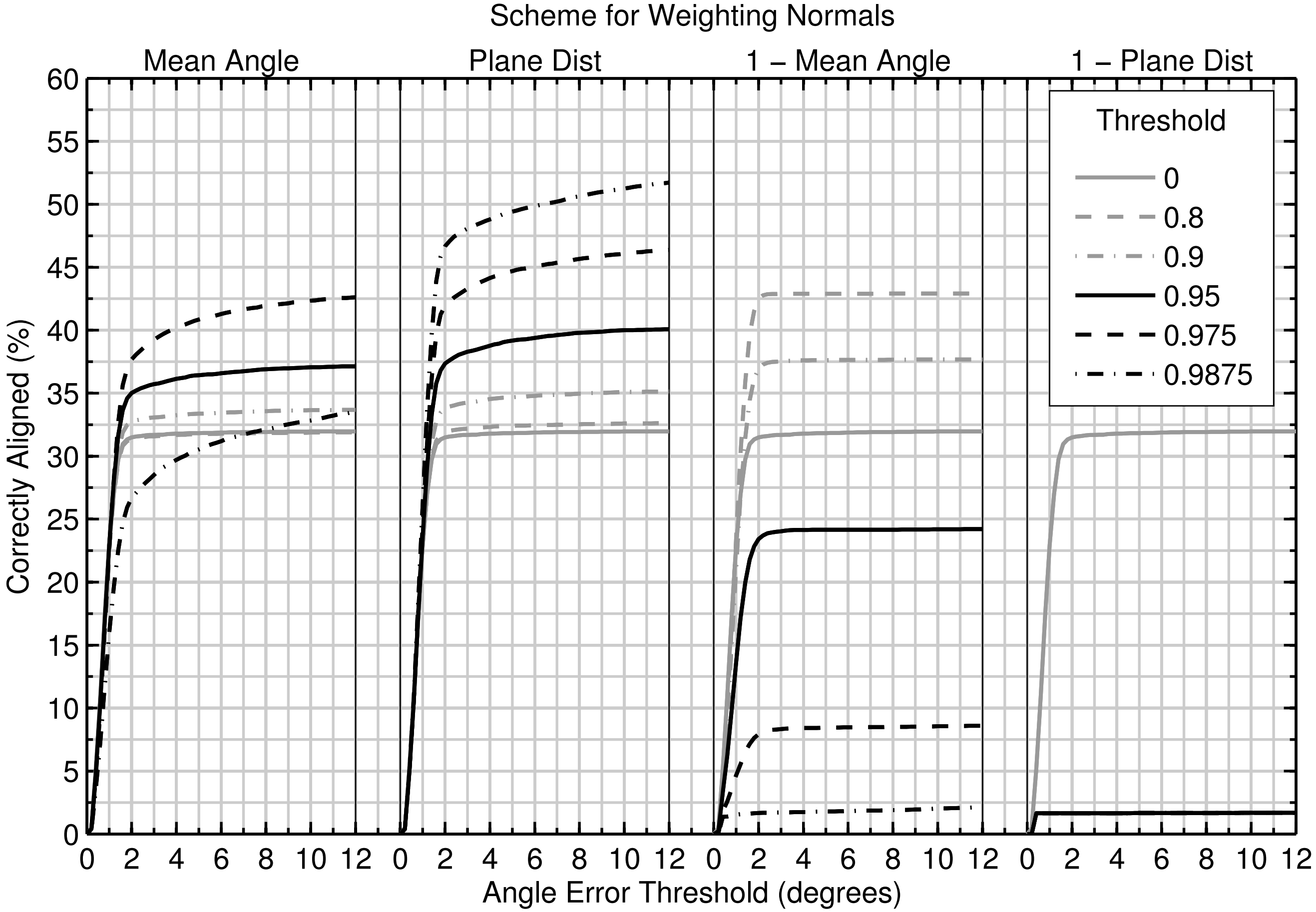


Figure 6.3: Graph comparing four different schemes for weighting and culling normals. The threshold value is the cull-point, any normals that have a weight less than this are culled. These results are shown for a bandwidth of 128.

on the Buddha statue and dragon model, as all the normals in the bunny model were culled. The effect of the inverse mean angle weighting scheme is described with greater detail in the discussion (see Section 6.4). The inverse plane distance weighting is unsuitable as at the tested thresholds all normals were culled. The correct alignments produced by the inverse plane distance are purely coincidental as the first element in the correlation matrix is selected in this situation; this element corresponds to the smallest available rotation, which works for point-cloud pairs that are already aligned.

Reweighted Bins

Binning normals and then reweighting the bins based on the number of normals collected is an alternative approach presented by Makadia *et al.* (2006), who reweighted equiangle bins. The implementation by Makadia *et al.* differs from that presented here, with these differences being compared and analysed in the discussion (see Section 6.4). The bin reweighting is determined based on a percentage threshold, with the appropriate threshold varying with bin count. Because bandwidth dictates the bin count for equiangle binning, three bandwidths, 16, 32 and 64, are used to determine if there is a relationship between the threshold value and bin count.

Figure 6.4 shows a trend between the three bandwidths and the percentage threshold, with each doubling of the bandwidth dropping the best performing threshold by approximately a seventh. Because the threshold was determined numerically (as opposed to analytically) for the three bandwidths, this trend will be extrapolated for higher bandwidths. The primary revelation of Figure 6.4 is how the four bin reweighting schemes compare with each other. Using a percentage threshold of zero with a reweighting scheme that only sets bins to zero is equivalent to performing the registration without any weighting, and as such is the baseline. With respect to this baseline, setting a bin's weight to one and normalising by its size has the worst performance, this is followed by the scheme which only reweights bins to zero based on the percentage threshold. Reweighting certain bins to zero does improve registration accuracy, which at a bandwidth of 16 is on par with the best performing reweighting scheme. The reweighting scheme which performs the best is dependent upon bandwidth, with reweighting bins to one being the better scheme at bandwidths 16 and 32, while at a bandwidth of 64, it is surpassed by reweighting bins to their area size. Both of these schemes perform better than the baseline as long as the percentage threshold has been correctly selected.

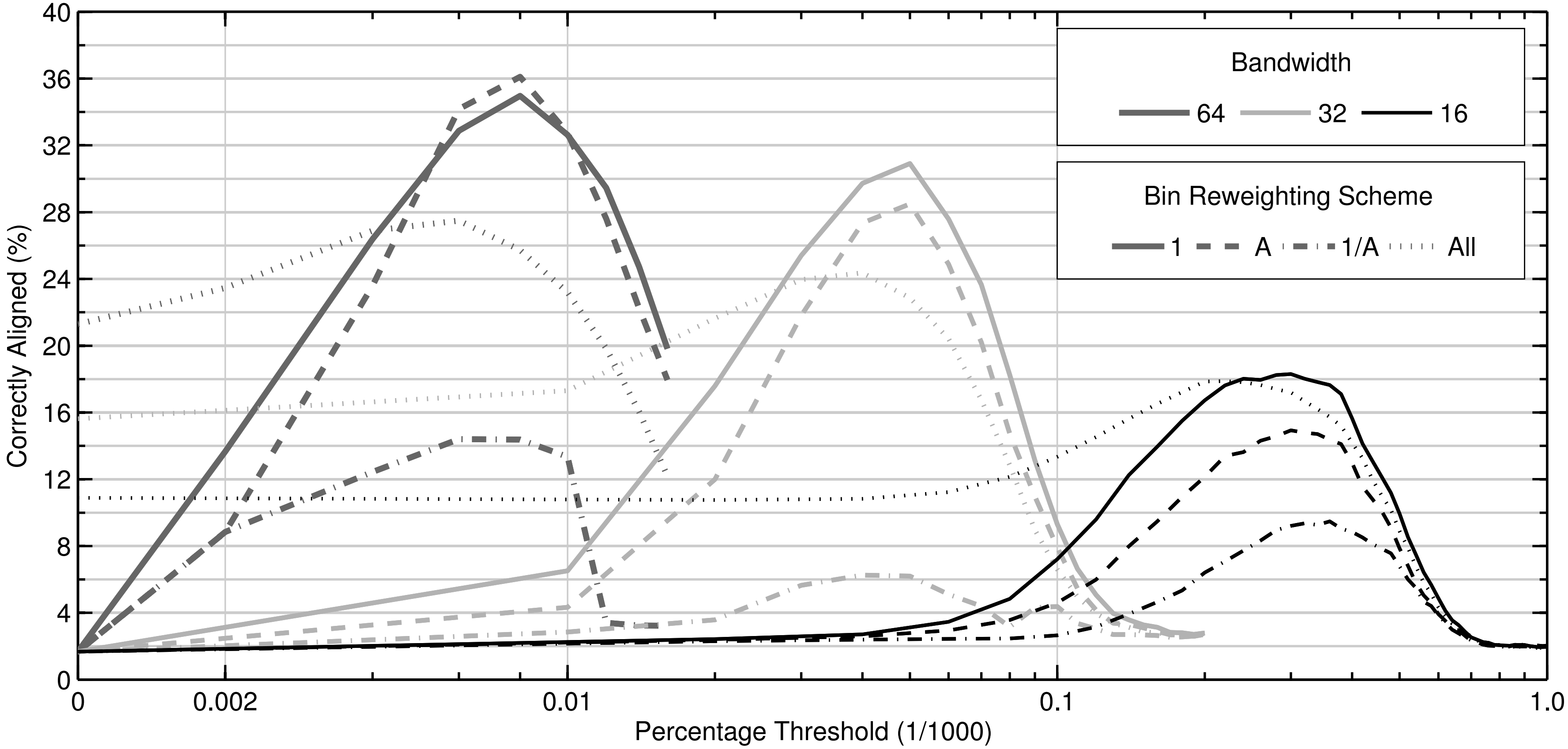


Figure 6.4: Graph comparing four different schemes for reweighting equiangle bins when they are thresholded by a given percentage value. This reweighting is tested with bandwidths 16, 32 and 64; each bandwidth is used for both the transform and correlation. An alignment is deemed correct if its rotation error is less than 10°.

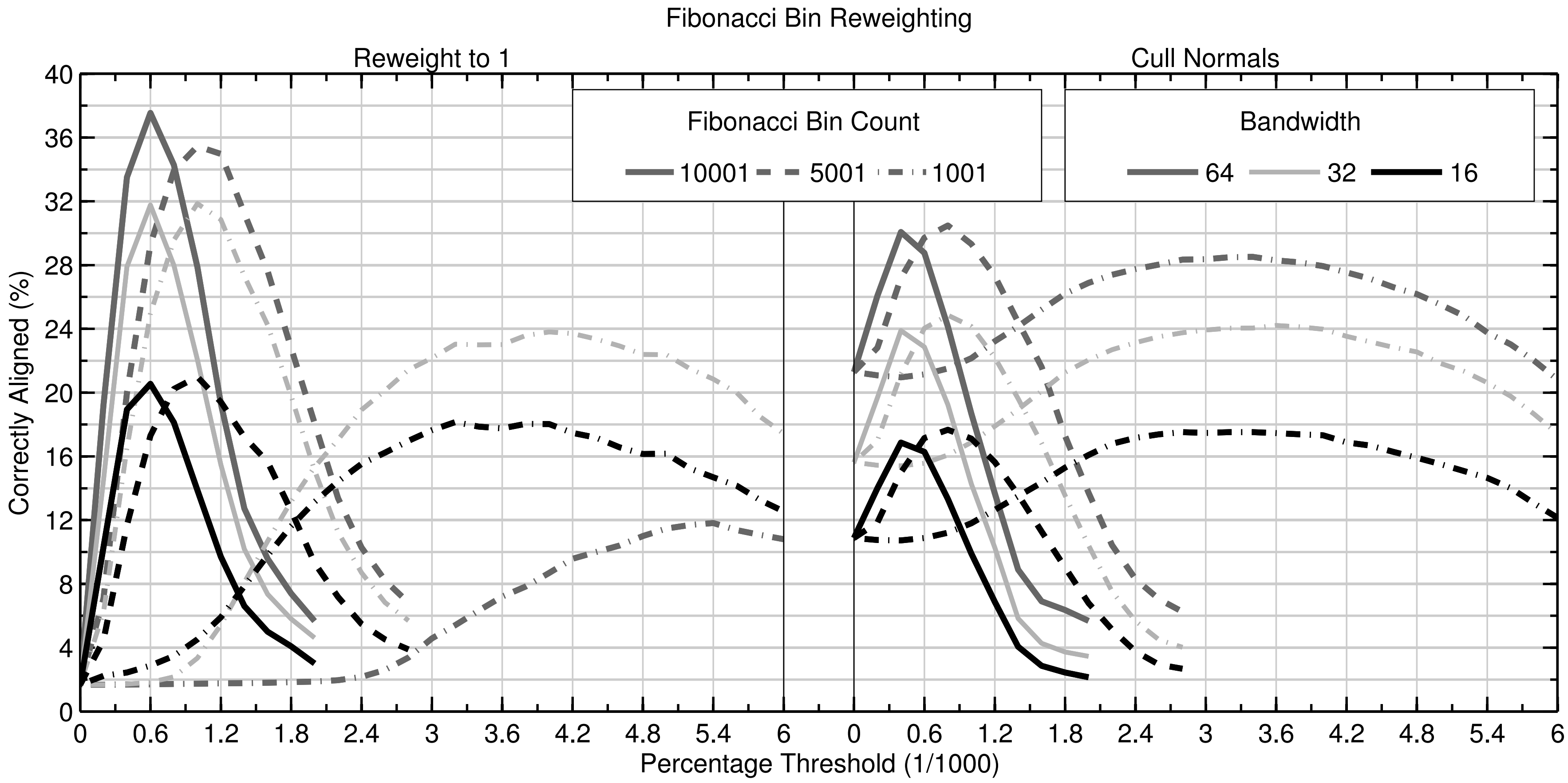


Figure 6.5: Graph comparing four different schemes for reweighting Fibonacci bins when they are thresholded by a given percentage value. This reweighting is tested with bandwidths 16, 32 and 64; each bandwidth is used for both the transform and correlation. An alignment is deemed correct if its rotation error is less than 10°.

The idea of reweighting bins is not exclusive to equiangle binning, as it can be applied to other binning schemes such as Fibonacci binning. The registration is still performed using equiangle binning, with the additional binning scheme providing an abstraction which allows the bin count, and hence the percentage threshold, to remain constant if the bandwidth changes. This consistency is demonstrated in Figure 6.5 with the best percentage threshold remaining the same with bandwidth, though it does vary with bin count, as there is a respective change to the number of normals that a bin collects. Simply using Fibonacci binning to cull normals provides a benefit to registration accuracy, though if the Fibonacci bin centres are treated as individual normals which are fed into equiangle binning, the registration accuracy is further benefited by an approximate 7% at a bandwidth of 64. The Fibonacci bin count does have an effect on registration accuracy, though this benefit plateaus as bin count increases. This is evident by the distinctly larger jump in performance between 1 001 and 5 001 bins, than that between 5 001 and 10 001 bins, when the Fibonacci bin centres are used with a bandwidth of 64. This plateauing is attributed to the decreased difference in bin size as shown by Figure 6.2.

The equiangle and Fibonacci results shown in Figures 6.4 and 6.5, respectively, identify which bin reweighting schemes have the best performance. Figure 6.6 increases the bandwidth to 128 to further evaluate the performance of equiangle binning when the bins are reweighted to one or the bin area, and when the centres of Fibonacci bins are used to represent the normals. Using a percentage threshold of 1.5×10^{-6} shows that reweighting the equiangle bins to their area has a better performance than reweighting the bins to one. This threshold was chosen as it follows with the trend identified in Figure 6.4. The performance of Fibonacci binning when 10 001 bins, at a threshold of 0.6×10^{-3}, are used only has a 2% lower performance than when the equiangle bins are reweighted to one, making them comparatively similar.

Complex Value

Combining the highest performing reweighting of bins scheme and the weighting of normals scheme into complex values produces the best registration accuracy, allowing the SHT to correctly align 85% of the point-cloud pairs. Using these two schemes together provides a representation of the point-clouds for correlation in the frequency domain that is more suitable than using these two schemes separately. The results in Figure 6.6 show how different weighting schemes compare at a bandwidth of 128, with the reweighted complex value scheme comprising of the plane-distance surface curvature at a threshold of 0.9875 and reweighting equiangle bins to their area using a percentage threshold of 1.5×10^{-6}. Applying these thresholds to the two weighting schemes is still a necessary step as simply combining all the normals and their plane-distance weightings to form the complex values is only marginally better than not using any weightings at all.

Figure 6.7 further analyses registration by examining how the three best weighting schemes compare to no weighting with respect to the overlap between point-cloud pairs. The equiangle scheme maintains the SHT registration capability at higher overlap percentages, while improving it as the overlap decreases. This is significantly improved upon by the complex values scheme, which only drops to a 50% correct alignments at an approximate 11% overlap. The aggressive nature of the plane-distance culling at a threshold of 0.9875 has an unusual effect on the relationship between the initial overlap percentage and the achieved alignment percentage. The result is that while more point-cloud pairs with lower overlap achieve correct alignment, increasing overlap does not improve registration accuracy as rapidly as the other schemes; potential reasons for this distinct growth rate are described below in the discussion (see Section 6.4). The minimum overlap required for the complex value weighting scheme to correctly align the Buddha statue, the Stanford bunny and the dragon model were 4.9%, 4.8% and 5.7%, respectively.

6.3.2 Efficiency

The computational cost of incorporating a weighting scheme is a minimal expense when contrasted with the achieved results. The cost of applying a surface curvature weighting to each normal is based on two factors: the total number of neighbouring points that are used and the equation which calculates the weight. If the imaging system samples the scene in a uniform or grid like manner, then it is expected that each point has eight neighbours. This means that the total number of times that either equation (6.6) or (6.10) are invoked will be $8n$, where n is the total number of points.

Fortunately, calculating the weight for each normal only occurs once, as it can be stored; this is useful when registering multiple point-clouds if SHT is applied to every point-cloud pair. Reweighting the equiangle bins is effectively a free weighting scheme as each bin's weight can be modified at the same time it is normalised by its size. However, the Fibonacci bin reweighting scheme has a binning cost (see Section 5.3.3). The computational cost of the complex

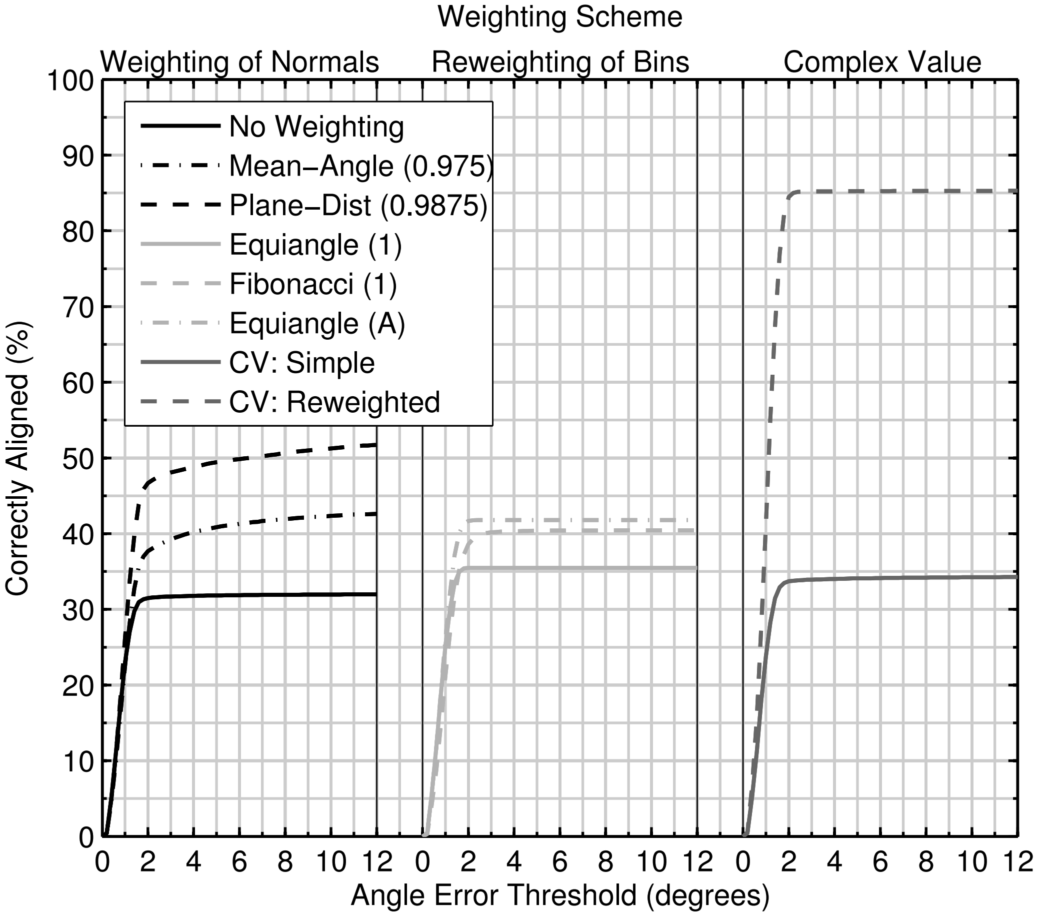


Figure 6.6: Comparison of the three weighting schemes types when the best performing parameters are used with a bandwidth of 128.

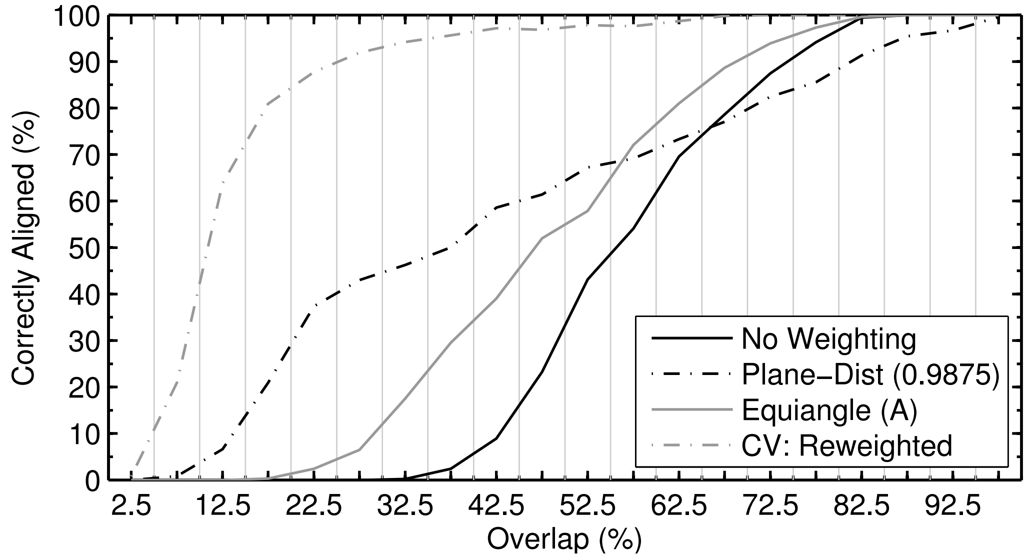


Figure 6.7: The percentage of correctly aligned point-cloud pairs versus their overlap for four weighting schemes. An alignment is deemed correct if its rotation error is less than 10°.

values scheme is based solely upon the two weighting schemes which are used in conjunction to form the complex value. Because the fast SHT is inherently capable of handling complex values, there is no additional cost attributed to using them with the fast SHT. In saying this though, additional symmetry does occur in the Fourier coefficients when the sampled values are strictly real. This symmetry can be exploited to provide a computational advantage, though there is only a marginal benefit (Kostelec and Rockmore, 2008).

6.3.3 Noise

The complex value weighting scheme has exceptional performance with respect to other tested weighting schemes, though as ground-truth data was used, the results it produced are unable to reflect how it handles noise. Keeping both the plane-distance culling and equiangle percentage thresholds at 0.9875 and 1.5×10^{-6}, respectively, reveals how well the complex value weighting scheme performs when noise is independently introduced to both the normals and the plane-distance weighting. These results are shown in Figure 6.8 at six different noise levels. The addition of noise is shown to have an immediate detrimental effect on registration accuracy, with the percentage of correctly aligned point-cloud pairs quickly decreasing as the noise angle increases. The decrease in performance is a reflection of the two used thresholds being highly sensitive to noise. It is expected that the ability of the correlation to handle noise will improve if the thresholds are less restrictive, though this will reduce the achieved performance of the ground-truth data. The implications of these noise results are described in greater detail in the discussion below.

6.4 Discussion

Currently, the extent of literature identified which uses weighting schemes in conjunction with spherical-harmonic correlation is limited to that presented by Makadia *et al.* (2006) and Larkins *et al.* (2013). The weighting scheme proposed by Makadia *et al.* was only described at a high level, therefore, comparing results against their implementation can only occur by making educated guesses. It was determined that the *constellation images* formed used a value of one to represent the equiangle bins deemed to be local maxima, with the rest being set to zero. The work presented here has revealed that this reweighting value has a lower performance than reweighting to the bin size when the bandwidth is 64 or greater.

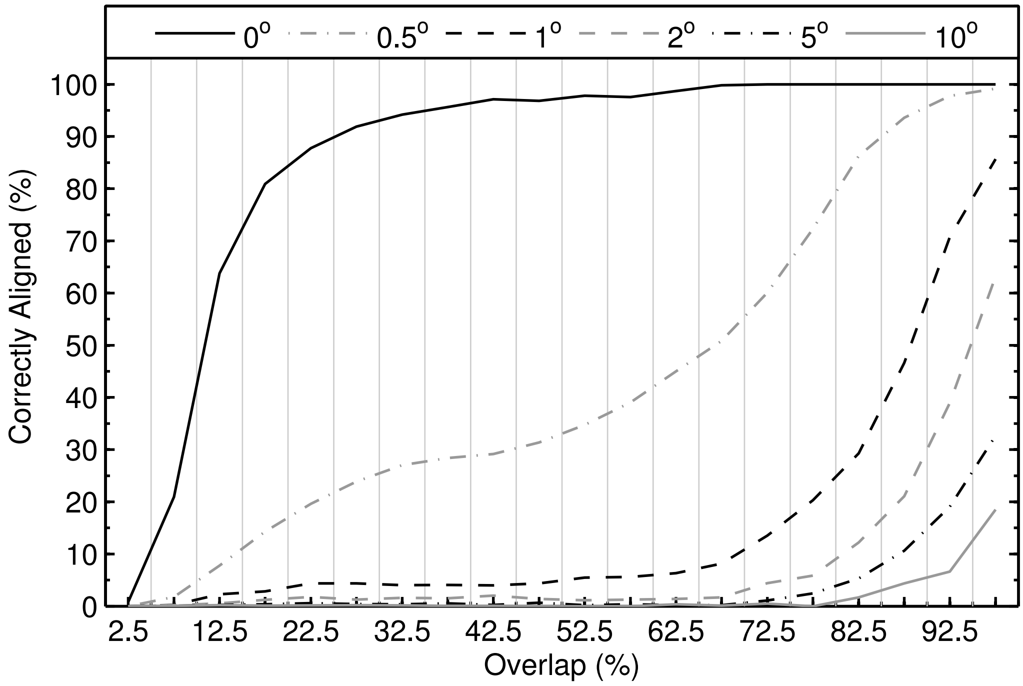


Figure 6.8: Performance of the complex values weighting approach at six different noise levels with respect to overlap. A bandwidth of 128 was used.

Two differences in implementation that may also be present are the use of extended Gaussian images (EGI) and selecting of local maxima. EGIs (Horn, 1984) are Gaussian images where each normal has been extended to include a weighting equivalent to the surface area that the normal represents. However, if Makadia *et al.* incorporated true EGIs into their pipeline, the surface area weighting becomes irrelevant when bin values are reweighted. This is because the goal of reweighting is to remove any bias that exists due to the same surface having different sampling densities when imaged from separate poses. The only situation where surface weight may have a role is when selecting which bins to keep. While the inclusion of surface area weighting may benefit registration accuracy, this chapter shows that exceptional results are achieved without it.

Makadia *et al.* do not disclose how bins are determined to be local maxima, though their approach may simply select a predetermined number of bins with the largest values. This approach is not expected to have any benefit over using a percentage threshold as they both have the same limitation: the best percentage threshold or number of bins to keep will differ with both the bandwidth and the distribution of normals in the scene. Makadia *et al.* also stated that a point-cloud pair with overlap as low as 45% was successfully aligned;

while their results might be a limitation of the testing data, this chapter has shown that using the complex value weighting scheme can align point-cloud pairs with overlap as low as 5%.

6.4.1 Surface Curvature Weighting

The results presented in Figure 6.3 show that reverse mean-angle weighting has a similar performance as mean-angle weighting, however, these results are only provided by the Buddha and dragon models. The reason why the reverse mean-angle weighting performs well for these two models is that they both contain surfaces with high curvature. This means that after culling has occurred, only the normals of these high curvature surfaces remain. Because a minimal number of these normals remain, with a subset of them common to both point-clouds, the correlation is able to calculate the desired rotation with greater accuracy. This is likely due to the percentage of normals common to both point-clouds decreasing as the total number of used normals increases. These extraneous normals still influence the correlation, and as such, the selected rotation can give an alignment with decreased accuracy. The mean-angle and plane-distance weightings are more susceptible to this issue as they use surfaces with low curvature. However, because fewer normals are culled, the correlation has a very low risk of being performed with no normals. As it is possible for the reverse mean-angle weighting to cull all the normals in a point-cloud, it is not recommended as a weighting scheme, with this recommendation also applying to the reverse plane-distance weighting due to its very poor performance.

The result of the plane-distance weighting presented in Figure 6.7 shows that it does not provide the best registration at higher overlap percentages. This is likely due to the aggressive culling threshold reducing the overlap between the two point-clouds; if different normals are culled from the two point-clouds (when there is high overlap), the overlap will decrease. The opposite effect occurs when the initial overlap is lower, as the normals in the non-overlapping regions are more likely to be culled, increasing the overlap between the remaining normals. Any decrease in the overlap between normals also decreases the ability of the spherical-harmonic correlation to identify the correct rotation. This is because the superfluous normals negatively influence the correlation, obscuring the preferred rotation. Lowering the culling threshold will improve the registration of point-cloud pairs with high overlap, but will in turn decrease the registration accuracy at lower overlaps.

Table 6.2: The percentage of correctly aligned point-cloud pairs for each of the three tested models at different weighting cull-points.

Threshold	Mean Angle			Plane Distance		
	Buddha	Bunny	Dragon	Buddha	Bunny	Dragon
0	26	30	40	26	30	40
0.8	26	30	40	27	30	41
0.9	28	30	42	30	32	44
0.95	31	33	47	32	35	52
0.975	33	37	57	31	43	64
0.9875	27	37	35	28	60	66

6.4.2 Thresholding Bins and Weights

Using a percentage threshold to restrict which equiangle bins are kept allows the bin reweighting to be invariant to the number of normals; though as mentioned earlier, the distribution of normals affects which threshold provides the best benefit to registration. This is because the distribution of normals is based on the shape of the scene, changing the number of normals that each bin collects. The effect is that the optimal threshold differs between a scene with high surface curvature and one with low surface curvature. These changes in surface curvature also impact the weighting of normals and the optimal culling threshold applied to them. The results presented in Figure 6.3 do not make this readily apparent as all point-cloud pairs are grouped together. Table 6.2 shows the percentage of correctly aligned point-cloud pairs for each model at different culling thresholds. As a whole, the performance across the three models follows a similar trend, except for the Buddha statue when using the plane-distance weighting, which has a drop in performance when the threshold increases from 0.975 to 0.9875. This drop is due to the Buddha statue having a higher level of surface curvature. While the point-clouds taken from the same model have a similar surface curvature, it must be noted that the optimal threshold will differ for each point-cloud pair.

A relationship between bandwidth and the best performing equiangle percentage threshold was revealed in Figure 6.4. To confirm that a linear trend is maintained between bandwidth and threshold, additional bandwidth values need to be examined. It is for this reason that the 1.5×10^{-6} percentage threshold may not be optimal, as it was determined by extrapolating this trend from a bandwidth of 64 to 128. Once again, the optimal value for this threshold will fluctuate with the aforementioned factors that affect the acquired data.

If a robust trend can be determined, the registration ability may be further improved over the current 85% by increasing the transform bandwidth to 512 and the correlation bandwidth to 256.

Even though surface curvature affects which threshold is optimal for culling normals, noisy normals will be still be removed, improving registration accuracy (if the remaining normals match). However, if the applied threshold culls all normals, then the correlation cannot identify the correct rotation. The exception to this occurs when the two point-clouds are already aligned, allowing the correct alignment to be inherently found; this is due to the first element in the correlation matrix being selected, which corresponds to the smallest available rotation. This issue with culling normals is especially true when using the reverse mean-angle weighting, as only the normals on high curvature surfaces are kept, and these surfaces are typically less common than low curvature surfaces. Additionally, noise and low sample density make these normals unreliable, as their calculated direction will not maintain consistency between poses.

The plane-distance culling threshold and the equiangle percentage threshold identified as providing the optimal registration accuracy both run the vulnerability of being biased towards the ground-truth data due to over-fitting. The risk of using over-fitted thresholds is that they may not accommodate data that has been affected by a range of factors. These factors include the shape of the scenes, the overall distribution of normals, the manner in which the imaging system samples the scene along with any inherent limitations of the hardware, and any noise that is introduced to the data. Therefore, the optimal threshold combination will vary with the point-cloud pair being registered. Alternatively, many of these factors can be mitigated by reducing the two thresholds; less aggressive thresholds are more lenient to changes in the data. The related disadvantage of using lenient thresholds is that for certain point-cloud pairs the correct rotation will not be identified, where as it would be if the thresholds are stricter. Appropriate thresholds need to be determined with respect to the point-cloud pairs being registered.

6.4.3 Noise

The two sources of artificial noise differ from noise that occurs in a real imaging system, especially given that they were independently added, and together will have compounded the overall noise level. Simulating all the different sources of noise that corrupts the data is a difficult task, as noise can occur at all stages of image acquisition. Therefore, the noise applied to the ground-truth data is purely to give an indication of the algorithm's robustness. The only viable means of truly evaluating how well the complex values and spherical-harmonic correlation handles data acquired from a particular imaging system is to use ground-truth data acquired by that imaging system. This is because the sampling distribution and noise sources are unique to each imaging system. If the ground-truth rotation between the two point-clouds is unknown, a visual evaluation will be needed. Chapter 9 tests the coarse registration pipeline with real data.

An additional approach for culling normals that was not implemented here is to simply set all bins that are within a predefined angle of the equator to zero. This is suitable for range imaging systems that are only able to image a scene from a single pose at any given time. Therefore, the normals that are near perpendicular to the optical axis are collected from surfaces that are nearly parallel to the optical axis. The angle of these surfaces with respect to the imaging system diminishes their ability to provide reliable sample points. The carry on effect is that the quality of the normals calculated for these points will be low. It was unnecessary to implement this for the used ground-truth data, as each segment was extracted from a fixed model.

6.5 Summary

Applying a weighting scheme to modify either the normals or the binning has an effect on the registration ability of spherical-harmonic correlation, as it changes the representation of the two point-clouds being registered. A variety of weighting schemes were proposed and tested to determine which are beneficial for maximising registration accuracy. The two primary schemes were weighting normals using their local surface curvature and reweighting bins. The surface curvature was determined using two measures: mean-angle and plane-distance, both of which improved registration accuracy when an appropriate threshold was applied. Once the optimal threshold for each measure was identified, the results showed that the plane-distance weighting was able to correctly align 10% more of the point-cloud pairs than mean-angle, for a total of 50%. The reweighting of bins was performed using two binning schemes: equiangle and Fibonacci. At a bandwidth of 32 or lower, reweighting bins that exceeded the threshold to a value of one (with the rest being weighted to zero) achieved the best results, while a bandwidth of 64 or greater revealed that reweighting the kept bins to the size of their surface was more suitable. Fibonacci binning overcomes the limitation of equiangle binning needing the percentage threshold to change with bandwidth. Using the optimal threshold for both equiangle and Fibonacci found that reweighting the equiangle bins had a slight advantage over Fibonacci binning as it correctly aligned approximately 2% more of the point-cloud pairs. Comparing the weighting of normals scheme with the reweighting of bins scheme shows that weighting normals with the plane-distance measure of surface curvature had better accuracy.

The most valuable contribution that this chapter makes is that it introduces the concept of combining the two aforementioned weighting schemes into complex values. The advantage of complex values is that they cohesively encode two sets of information into an orthogonal system, which are then transformed into the frequency domain. The spherical-harmonic correlation then uses the information from both weighting schemes. The result of using complex values was a significant improvement in registration ability, with 85% of the point-cloud pairs being coarsely brought into correct rotational alignment. This increases the percentage of correctly aligned point-cloud pairs by 35% over the plane-distance weighting scheme, which aligned 50% of the point-cloud pairs, making it the next best weighting scheme. The added benefit was that the complex values weighting scheme was capable of aligning point-cloud pairs with overlap as low as 5%. Adding noise to the data did have a detrimental impact on the registration accuracy of complex values, but this is primarily due to the chosen thresholds being too restrictive with respect to noise.

Chapter 7

Translational Alignment

The found rotational alignment is only half of the coarse registration as the translational alignment also needs to be determined. The translational registration is carried out in a manner similar to rotational registration, with the points being histogrammed and transformed into the frequency domain instead of the normals. Operating in the frequency domain allows the translation correlation to be performed with relative efficiency, as it simultaneously determines the correlation value for all translational shifts that the histogram permits. Identifying the translation by way of phase correlation is an established technique, particularly for 2-D registration (Reddy and Chatterji, 1996). Its use for determining the translation alignment for 3-D point-clouds also appears in literature (Makadia *et al.*, 2006; Bülow and Birk, 2013), but is less common. Makadia *et al.* only described the fundamental mathematics behind how the translation correlation was performed, but did not investigate the algorithmic robustness in any form.

The analysis undertaken in this chapter attempts to identify the accuracy with which the translation correlation is able to bring point-clouds into coarse translational alignment, with respect to a variety of factors. These factors include the initial rotational misalignment of the two point-clouds, the number of voxels used in the histogramming process and the amount of noise contaminating the points.

7.1 Parameters

The translational alignment process is described with detail in Section 3.3. The parameters outlined here are provided as a means of measuring the error remaining after translational alignment, allowing its accuracy to be gauged.

7.1.1 Voxel Size

The side length, ℓ, specifies the size of the cube that encompasses all points of both point-clouds. The cube is then divided up into a number of voxels, v, in each of the x, y and z directions, for a total of v^3 voxels. Because the voxelisation process histograms the points, discretisation occurs, thus translation registration by straightforward discrete correlation of this histogram is only capable of coarse alignment. It is a coarse alignment because the translational shift in each direction occur in steps of ℓ/v, the size of each voxel. While sub-voxel registration is possible (Balci and Foroosh, 2006; Tzimiropoulos *et al.*, 2011), it is not investigated here as the goal is coarse registration. The voxel size also determines the alignment precision, as a margin of error can occur due to equivalent points from both point-clouds being collected by the same bins in both cubes. The maximum distance that these points can be from each other while remaining in the same equivalent bins is

$$\sqrt{3\left(\frac{\ell}{v}\right)^2}, \tag{7.1}$$

which is the diagonal distance through the voxel.

7.1.2 Translation Error

Measuring the exact error in translation after translational alignment of two point-clouds can only occur when the ground-truth translation is known. The true translational shift between the two point-clouds is determined by placing the centroid of the first point-cloud at the origin of the coordinate system and taking the vector to the centroid of the second point-cloud. In preparation for performing the translation correlation, the second point-cloud is shifted so that its centroid is also placed at the origin. The best translational shift given by the maximum impulse response in the correlation matrix will be close to the true translation, but as this is a coarse alignment, some translational misalignment will occur. This is because the translation error is caused by a combination of rotational misalignment, noise, and the discrete steps of the translational shift. The translation error is the distance between the found location and the true location of the second point-cloud's centroid. Figure 7.1 provides an illustration of how the translational misalignment is measured, with $\mathbf{f}$ and $\mathbf{t}$ being the found translation vector and the true translation vector, respectively.

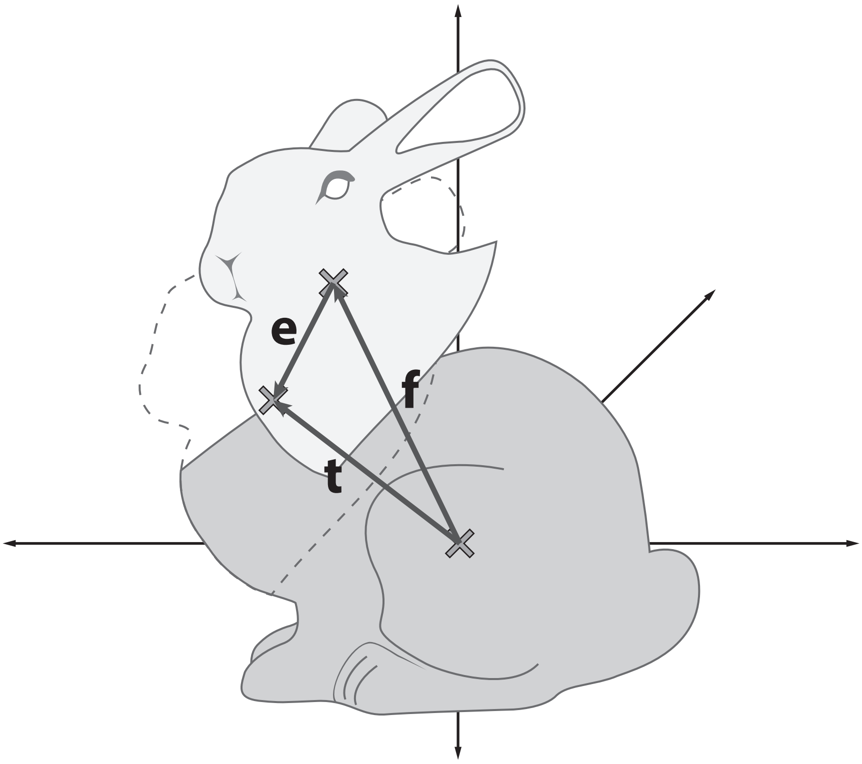


Figure 7.1: The centroid of both point-clouds (the body and head) are initially positioned at the origin of the coordinate system, the second point-cloud (head) is shifted using the found translation, **f**. Because the ground-truth is known, the true translation, **t**, is also known, allowing the translation error to be calculated as the vector **e**, the difference between **f** and **t**.

The translational misalignment is calculated as

$$\mathbf{e} = \mathbf{f} - \mathbf{t}, \tag{7.2}$$

with the *translation error* defined as

$$\epsilon = ||\mathbf{e}|| \,. \tag{7.3}$$

The translation error provides a useful means of measuring the accuracy of the correlation when bringing two point-clouds into translational alignment. However, it is not invariant to point-cloud size; the larger the point-cloud, the larger the resultant translational error will be as voxel size also increases. Therefore, the translation error needs to be normalised to enable comparison between various objects. Attempting to normalise by the maximum translation error that the correlation generates is one possible approach, however, it does not provide useful information with respect to the point-cloud pair being registered. Because of this, the translation error found for a given point-cloud pair is normalised by the mean distance between neighbouring points of the unsegmented model from which the point-cloud segments were extracted. The normalised translation error is then a value related to the mean separation between the points; for example, if the normalised translation error has a value of three, then the optimal translational alignment is off by three times the mean separation.

7.2 Methodology

Following on from the previous two chapters, the translation registration is once again examined using the Dragon, Buddha statue and Bunny datasets (see Section 4.1) allowing consistency to be maintained as well as providing the ground-truth rotation and translation. The process followed for testing how well translation correlation performs for bringing point-cloud pairs into translational alignment is achieved by first shifting both point-clouds so that their centroids are at the origin of the coordinate system. Because no rotational misalignment currently exists between the two point-clouds, their rotational alignment does not affect translational alignment. However, as the rotation correlation is only a coarse registration algorithm, rotational misalignment does occur.

Varying angles of rotational misalignment are tested on each point-cloud pair by rotating the second point-cloud by the prescribed angle around a randomly placed rotation axis. The tested angles of rotational misalignment are 0°, 1°, 2°, 5°, 10° and 15°. A random rotation axis is formed for the second point-cloud by selecting a random point on a unit sphere, with this point being given (in spherical coordinates) as

$$\left[2\pi u_1,\ \cos^{-1}(2u_2 - 1)\right], \tag{7.4}$$

where $u_1, u_2 \in [0, 1)$ are uniform deviates. As there are three models, each with 120 segments, 21 780 tests are generated for each rotation misalignment. While an alternative random axis will give a different result, the large number of tests sufficiently evaluates the ability of the translation correlation for handling varying levels of rotational misalignment.

The points can be located anywhere in the coordinate system, the histogramming process provides estimate values of the points at the sample locations needed for the discrete Fourier transform to the frequency domain. The total number of bins contained by the histogram is determined by the number of voxels v along each side of the histogram cube. The greater the value of v, the lower the quantisation error, thus the translational alignment is determined with greater accuracy. The two downsides of increasing v is that there are fewer points per histogram bin, thus the sample point at a bin has a poorer measurement of the surface, and secondly, the computational cost is increased. A suitable value for v is one low enough to minimise the computational cost while also providing adequate registration accuracy for fine registration. The four values of v tested are 16, 32, 64 and 128.

The point-cloud pairs that are registered together are extracted from the same base model, therefore, the overlapping sections of the two point-clouds contain the exact same points. This situation does not occur when imaging a real scene as the sampling distribution changes with the pose of the imaging system, and in addition, varying levels of noise are inherently introduced into the acquisition based on the ability of the imaging system. These two aspects are mimicked by introducing artificial noise to the points. The level of random Gaussian noise to be added to each dimension of all points is specified by a standard deviation value. This value is the mean separation between neighbouring points of the unsegmented model from which the two point-clouds have been extracted. Increasing this standard deviation value increases the level of noise contamination.

The normalised translation error only provides a measure of translational misalignment with respect to the mean distance between points for a given model. Any threshold value can classify a translation as correct if the normalised translation error is less than or equal to the threshold; the larger the threshold, the larger the allowable translation error. Here a translation is classified as correct if its normalised translation error is 15 mean distances or less. A threshold of 15 is chosen as a fine registration algorithm, such as ICP, will be capable of handling this amount of translational misalignment.

The ability of the translation correlation to handle noise is tested by adding Gaussian noise independently to each of the three dimensions of every point. The level of Gaussian noise is specified by a standard deviation value and a noise multiplier. In order to relate the noise to each model, the standard deviation value is the mean distance between points for the given model. The noise multiplier is simply a value that is multiplied with the standard deviation value to increase the noise level; three noise multipliers are used, 1, 3 and 5.

Table 7.1: The approximate percentage of point-cloud pairs that are correctly aligned within a normalised translation error of 15 for a specified voxel count.

Rotational Misalignment	Voxel Count 16	32	64	128
0°	79%	98%	100%	100%
1°	79%	97%	100%	100%
2°	79%	97%	99%	99%
5°	75%	92%	95%	96%
10°	57%	72%	79%	81%
15°	36%	48%	56%	59%

7.3 Results

The following results quantify the ability of the translation correlation to correctly align point-clouds. This is examined with respect to varying levels of rotational misalignment, and with noise added to the points.

7.3.1 Accuracy

The accuracy of the translation correlation is primarily determined by the voxel count prescribed to the histogramming process. Figure 7.2 shows that the greater the voxel count, the greater the translation accuracy. Table 7.1 shows the percentage of point-cloud pairs that are classified as correct when examining these results with respect to the classification threshold of 15. There is very little difference in performance when the angle of rotational misalignment is 0°, 1° and 2°. As the rotational misalignment increases from 5° up to 15°, there is a decrease in the percentage of point-cloud pairs that are classified as correct. This decrease is less apparent for a voxel count of 16, which is due to larger voxels having a greater resiliency to increased rotational misalignments, as each voxel still captures the same points as when there is a smaller rotational misalignment.

The classification percentage for a given normalised translation error presented in Figure 7.2 is the mean of the three tested models. Figure 7.3 shows that the results for each model have minor fluctuations with respect to the mean, with these fluctuations occurring due to a variety of factors, the primary of which is point density with respect to model size. This is because translation error is normalised based on point density, while voxel size is determined

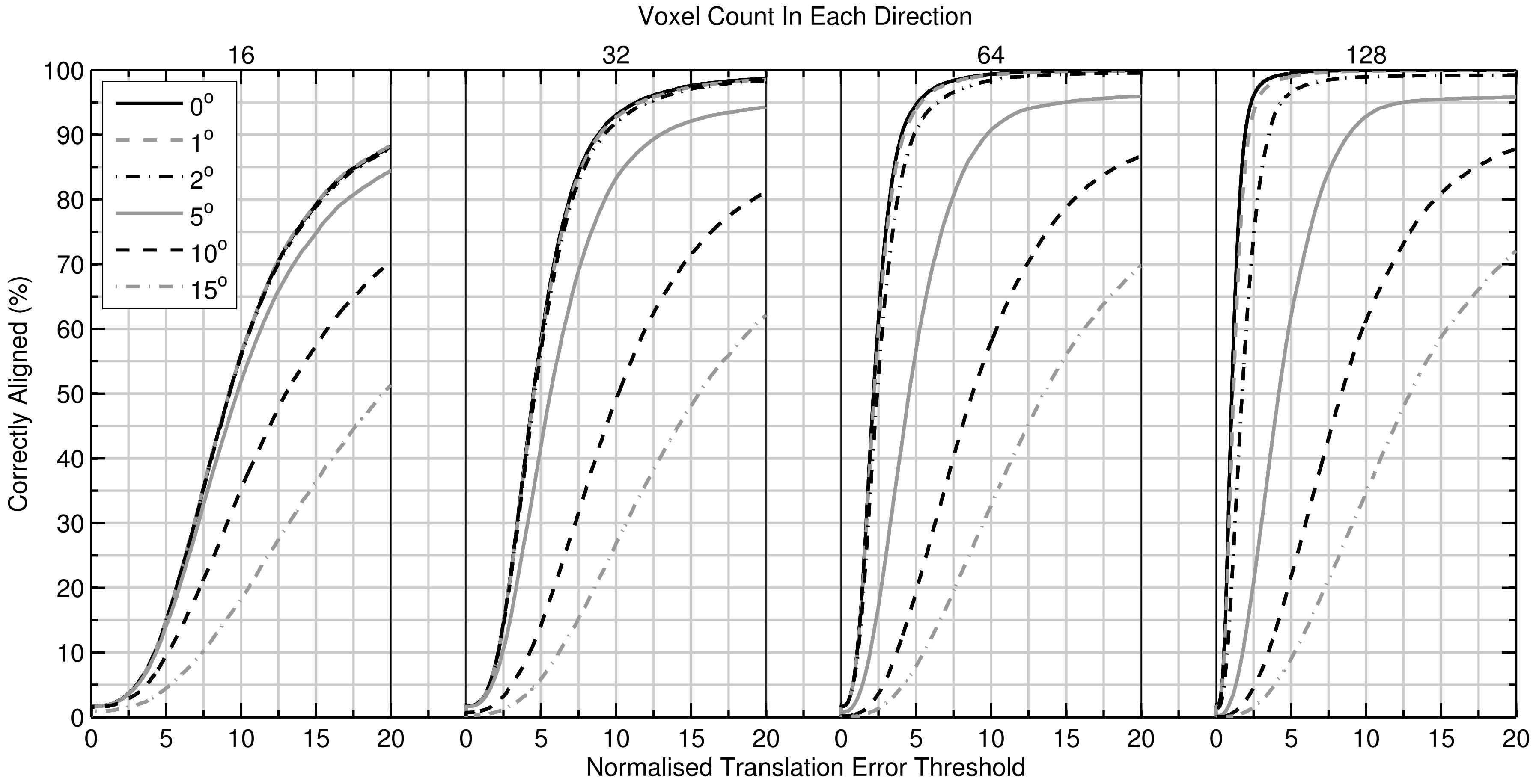


Figure 7.2: Comparison of the translation registration using four voxel counts and six rotational misalignments. The translation error is normalised based on the mean distance between neighbouring points from the respective unsegmented model.

by model size. The result is that a model with greater density with respect to its overall size will increase the normalised translation error compared to one with lower density when using the same voxel count. These fluctuations are also compounded by any factor that influences the translational accuracy. It is these factors that cause the Buddha point-clouds to achieve lower translation errors than the bunny when no rotational misalignment is present, and the reverse when the rotational misalignment is 5°. On the whole, the results from each model follow the same relative trend when using the same voxel count and rotational misalignment.

The overlap between two point-clouds influences whether the translation correlation will determine the correct translational alignment. Figure 7.4 shows the percentage of correctly classified point-cloud pairs for a given overlap; most point-cloud pairs are correctly aligned when overlap is greater than 12.5% when rotational misalignment is 2° or less. As the rotational misalignment between the two point-clouds increases to 5°, overlap needs to be approximately 32.5% before all point-cloud pairs are correctly aligned. The results show that no matter the overlap, higher rotational misalignments do not correctly align all point-cloud pairs. Given this shortcoming, approximately 85% of point-cloud pairs still achieve correct alignment at a rotational misalignment of 10°, when the overlap is greater than 32.5%. However, at 15°, the maximum percentage of correctly aligned point-cloud pairs is approximately 65%, which requires the overlap to be greater than 50%.

7.3.2 Efficiency

The computational cost of performing the translation correlation is determined by two primary factors: the number of points in both point-clouds and the number of voxels used for histogramming. The cost of binning the n points is $\mathrm{O}\big(n\big)$ as the bin that collects a point is determined directly. The two constructed histograms each have a size of v^3, meaning that converting both of them into the Fourier domain by the most efficient fast Fourier transform has a cost of $\mathrm{O}\big(v^3 \log v\big)$. Multiplying the two histograms together in the Fourier domain to form the phase correlation matrix has a cost of $\mathrm{O}\big(v^3\big)$. Inverting the phase correlation matrix into the impulse domain has the same cost as the conversion to the Fourier domain. Once in the impulse domain, finding the maximum impulse has a cost of $\mathrm{O}\big(v^3\big)$ as each element is searched. Identifying the location of the maximum impulse and converting it into a translation completes the translation correlation.

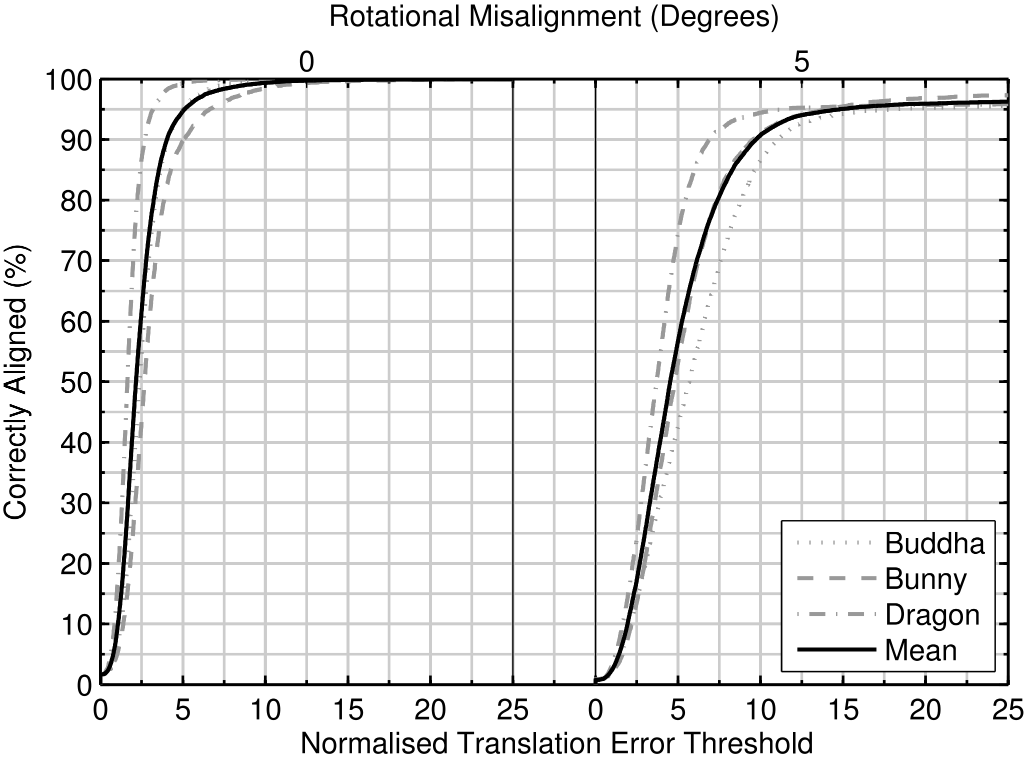


Figure 7.3: The percentage of correctly aligned point-cloud pairs as the normalised translation error threshold is increased. The results are shown for each of the three models and their overall mean for two rotational misalignments. A voxel count of 64 was used.

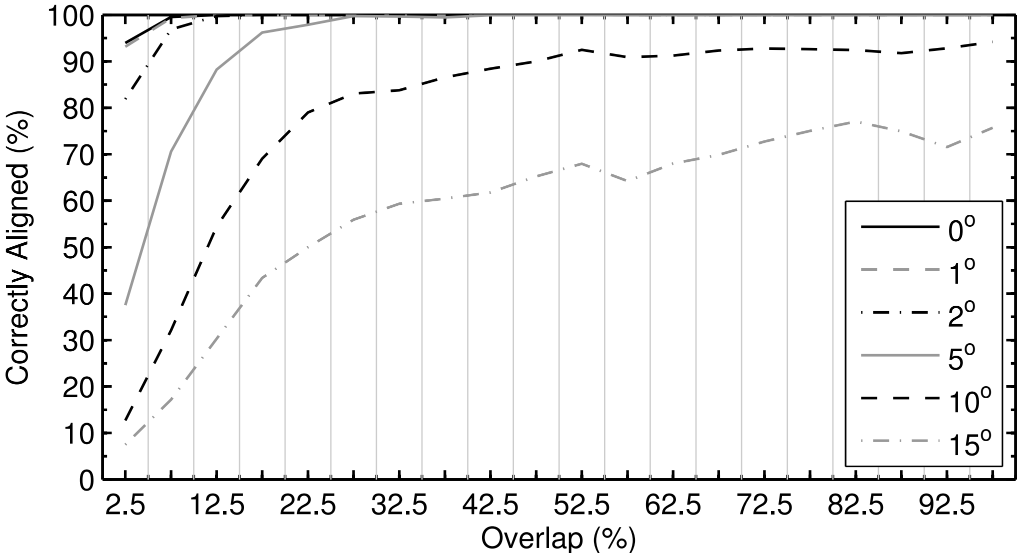


Figure 7.4: The percentage of point-cloud pairs correctly aligned by the translation registration at six different angles of rotational misalignment with respect to overlap. Each dimension was broken into 64 voxels, with the translation being deemed correct when the normalised translation error was 15 or less.

7.3.3 Noise

The accuracy of the translation correlation for bringing a point-cloud pair into correct alignment has been shown to be dependent upon the number of voxels used. Using 64 voxels in each direction to form the histogram for binning points allows the correlation to achieve an adequate level of accuracy while restricting the computational cost. Figure 7.5 compares how different noise multipliers and angles of rotational misalignment affect the percentage of correctly aligned point-cloud pairs. When the noise multiplier has a value of 1, there is a minuscule drop off in performance with respect to no noise being present, though this drop off is more prominent as the rotational misalignment increases. When the noise multiplier is 3, one standard deviation of Gaussian noise is three times the mean distance between points. This results in an approximate 10% drop in correctly aligned point-cloud pairs when the rotational misalignment is minimal, but this drop increases to 20% with 5° of rotational misalignment.

Because a voxel margin-of-error, for a voxel count of 64, is approximately four mean distances in length, this level of noise can shift points up to three voxels away from the one in which they would initially reside. Further increasing the noise level again using a multiplier of 5 limits the ability of the translation correlation to achieve alignment within an acceptable accuracy threshold. Rotational misalignment on the whole still has the biggest impact on translational accuracy, though noise does become more of an issue as it increases, as shown when the noise multiplier is 5. Overall, these results show that if noise is contained within a voxel or its immediate neighbours, noise has minimal impact on the accuracy of the translation correlation in bringing two point-clouds into an acceptable translational alignment.

7.4 Discussion

The translation correlation is a coarse registration algorithm, therefore, the translational shift occurs in discrete steps as established in Section 7.1.1, with these steps being determined by voxel size. Because ground-truth is known, the exact translation error can be measured. For a point-cloud pair where the ground-truth is not known, and the optimal discrete translational from the correlation is found, the translation error can still be up to the margin-of-error of the voxel. The normalised size of this margin of error across all point-cloud pairs for all three models is 3.7, with a standard deviation of 0.8. This

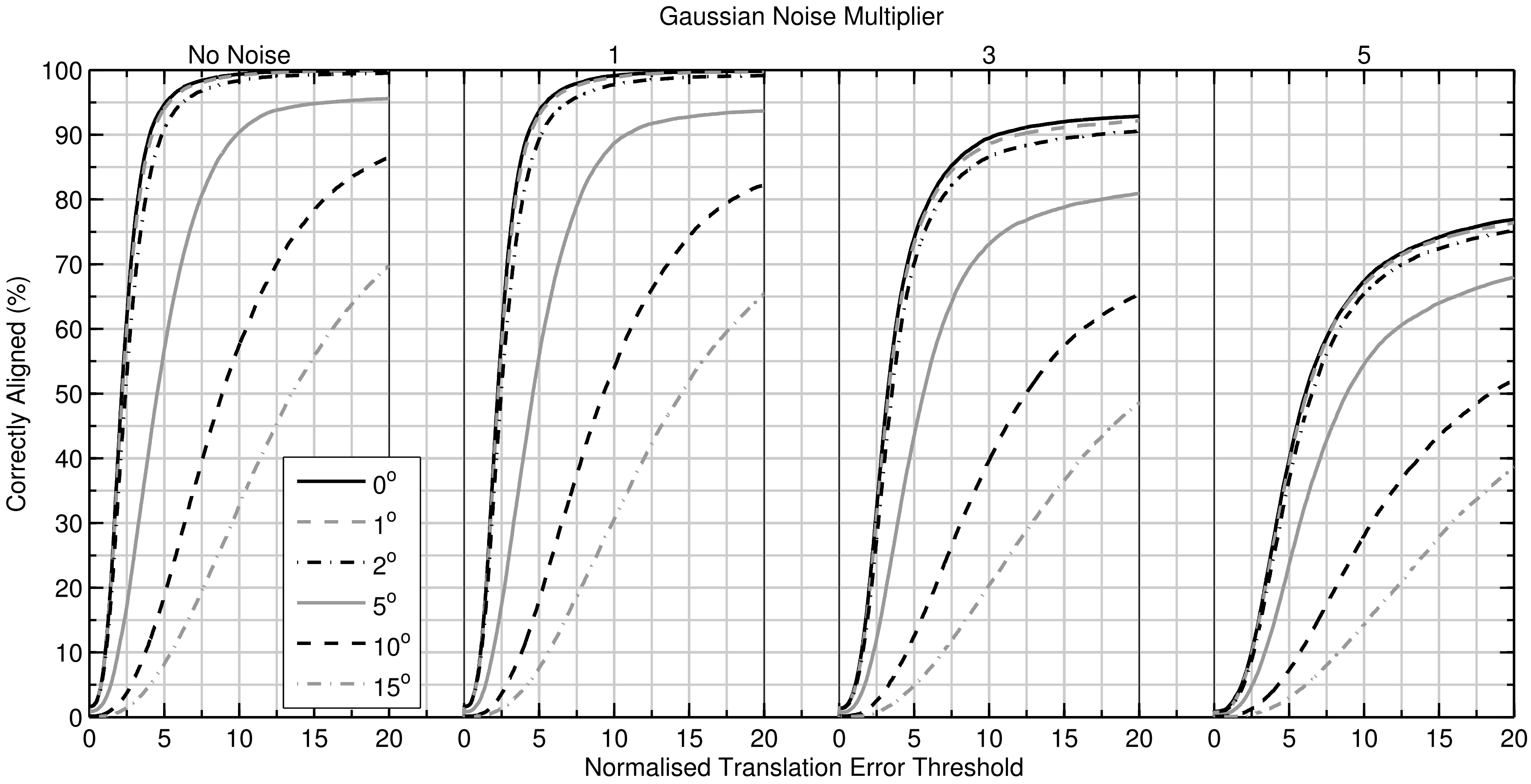


Figure 7.5: Comparison of the different noise levels corrupting the points when each dimension is broken into 64 voxels before histogramming.

means that for a voxel count of 64, even if the optimal discrete translation is found, the translational alignment can still be off by up to four times the mean distance between points. While this ambiguity exists and is expected due to the translation correlation being a coarse algorithm, it is minimal and within the misalignment that a fine registration algorithm is capable of handling.

Translation correlation has been shown to bring the majority of point-cloud pairs into correct translational alignment, even when they have minimal overlap. This is because each point-cloud only represents surfaces within the scene, meaning no points occur behind others when viewed from the pose of the imaging system. The result is that many of the histogram bins or voxels remain empty, causing voxels that contain points to be brought together. Because the translation correlation simply attempts to find the translation which best fits the two point-clouds together, its ability is dependent upon the preceding rotation correlation. The closer the two point-clouds are to being rotationally aligned, the greater the likelihood that the translation correlation identifies the appropriate translation. This improved performance is due to more voxels from the first histogram being able to match their counterparts in the second histogram when the correct translation is applied, making the corresponding correlation value more distinct.

The voxel count used for performing the translation correlation affects both its capability to handle noise and the achieved translation accuracy. Because the size of a point-cloud does not vary, the area encompassed by each individual voxel increases as voxel count decreases. The larger a voxels size, the better its ability to contain points that have been contaminated by noise or have rotational misalignment. The resulting histogram is then a closer match to what it would be if points had no contamination. This allows the translation correlation to be more resilient to noise, permitting a translation to be identified as if there was minimal noise present. While using a larger voxel size reduces the influence of noise, the achievable translation accuracy is also reduced due to the increased discretisation error. This is because the size of the translation step in each direction is determined by the length of a voxel's edge, and it is for this reason that accuracy improves as voxel count increases. The best voxel count is dependent upon the quality of the point-clouds being registered, the ability of the subsequent fine registration algorithm and any constraints related to computational cost. Currently, the appropriate voxel count is identified via testing; a dynamic solution based on the structure of a point-cloud may be feasible, though not investigated here.

The benefit that translation correlation provides for the entire registration pipeline is dependent upon the fine registration algorithm. This is because the used fine registration algorithm may be capable of achieving the correct alignment without needing the translation correlation stage. In the case of ICP, translation correlation will still reduce the total number of iterations that ICP performs, improving computational efficiency. The applicability of translation correlation in the overall registration pipeline is outside the scope of this investigation, as its incorporation should be determined based on the capability of the fine registration algorithm.

The amount of noise that contaminates each point is based on the imaging systems accuracy when sampling a scene. The artificial noise used here for examining the robustness of the translation correlation is Gaussian noise that has been applied to the points in all three directions. This noise is not meant to be representative of any particular imaging system, as the noise introduced by each system will differ due to their own inherent limitations. Instead, it is provided as a means of measuring how well the translation correlation handles points that have been randomly shifted at increasing levels from their original position. It is for this reason that the noise level was generated using multiples of the mean distance between points, as it allows points to be contaminated with noise that is related to the point-clouds themselves. The level of noise that would occur in practice should be low enough that a point has a high probability of still being captured by the same voxel as it would have if no noise was present. The robustness of the translation correlation means that even if noise shifts a point to an adjacent voxel, there is minimal effect on registration accuracy.

7.5 Summary

The analysis of the translation correlation found it to be a highly effective algorithm for performing translation registration. The correlation is able to bring the majority of point-cloud pairs into translational alignment with a high level of accuracy, though the accuracy level is dependent upon the number of used voxels. When the point-cloud is broken into 64 voxels in each direction and no rotational misalignment or noise is present, all point-cloud pairs achieve registration accuracy within the 15 mean distance threshold. The ability of the correlation is affected by the rotational misalignment between the point-clouds, with it decreasing as rotational misalignment increases. The introduction of noise that contaminates the points was shown to have little influence on the accuracy as long as the points are collected by the same voxel or those adjacent to it. Point-clouds with greater noise levels run the risk of having a much lower level of translation accuracy. However, the translation correlation can be optimised to mitigate both the rotational misalignment and any present noise by varying the voxel count. A lower voxel count will decrease the achievable accuracy, but as this is a coarse registration algorithm, the amount of accuracy required is dependent upon the subsequent fine registration algorithm.

Chapter 8

Verification of Correct Alignment

Once the rotational and translational registrations between point-clouds have been performed, it is unknown whether the found alignment is correct. Automatically verifying correct alignment can be accomplished by determining that the point-cloud alignment conforms to a predetermined criterion; this allows an alternative alignment to be sought when the transform is deemed misaligned. The nature of each criterion is dependent upon the verification strategy formed for each alignment or point-cloud aspect. Verification strategies such as surface orientation (Makadia *et al.*, 2006), visibility (Huber and Hebert, 2003; King *et al.*, 2005; Makadia *et al.*, 2006) and transitivity (Pooja and Govindu, 2010) have been used previously in literature to automatically verify correct alignment.

This chapter introduces a new verification strategy based on the size of the maximum impulse response from the translation correlation and expands on the idea of transitivity (see Section 2.5). Surface orientation consistency along with these two new strategies are examined to ascertain their capability to correctly classify a found alignment. If their performance is satisfactory, then they will provide a benefit to the coarse registration pipeline as they will limit the amount of manual verification and intervention required.

8.1 Alignment Verification Approaches

This section describes the three investigated verification approaches. The first uses the size of the impulse response from the translation correlation; the impulse response from the rotation correlation is not applicable as it is not normalised. The other two approaches, surface orientation and transform transitivity, are more generic as they evaluate the consistency between surfaces and the difference between the transforms themselves, respectively.

8.1.1 Translation Correlation Value

Performing the translation correlation produces a three-dimensional correlation matrix whose elements are normalised between one and negative one; a value of one occurs when both histograms are identical. The element with the maximum impulse response is selected as the translation, with the value of this response, termed the translation correlation value (TCV), providing a measure of the overall alignment quality. This is due to the TCV increasing as aspects such as point-cloud overlap, rotational misalignment and noise improve. Therefore, applying an appropriate threshold to the TCV is one approach for classifying an alignment. Because the aforementioned aspects affect the TCV in different ways, a threshold will need to be adjusted to accommodate all alignment situations.

8.1.2 Surface Orientation Consistency

The overlapping surfaces of two registered point-clouds will have the same orientations if correct alignment has been achieved. Because individual surfaces are not inherently given by a point-cloud, their orientations are expressed by the normals. Surface orientation consistency is evaluated by binning the two point-clouds in a manner similar to the translation correlation, except the point-clouds remain in their found alignment position. Binning is performed using the points, with the normals associated to the points in a given bin being collected together; the orientation of the surfaces within this bin is calculated as the mean of the normals across their three dimensions.

The mean orientation is given as a new normal, therefore, each relevant bin will contain two of these new normals, one for each point-cloud; a relevant bin is a bin that contains at least one normal from both point-clouds. The difference in orientation for each relevant bin is calculated as the angular difference between the two normals. The contribution that a bin makes to the overall angular difference (based on the number of normals it collects) is also incorporated. To provide a fair weighting for each bin, a bin's angular difference is multiplied by the number of normals provided by the point-cloud with the smallest presence in the bin. The overall difference in surface orientation,

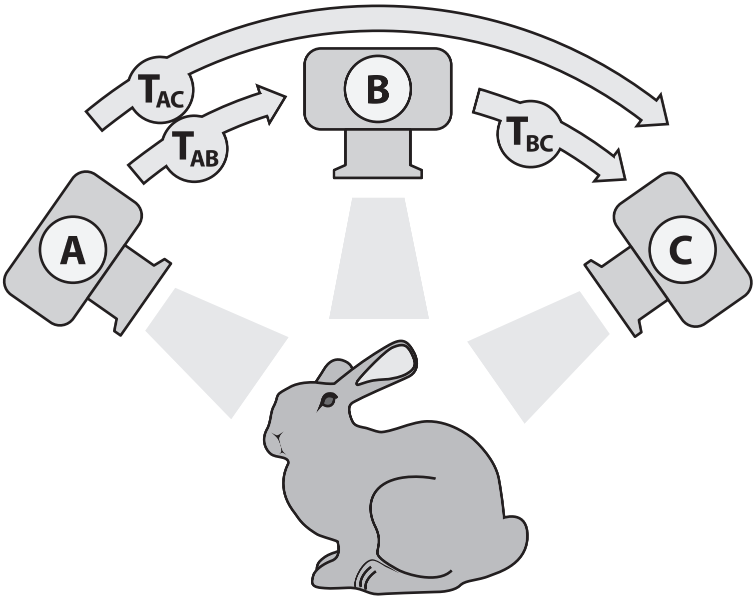


Figure 8.1: Three poses, A, B and C, of an imaging system are related to each other by the three transforms $\mathbf{T}_{AB}$, $\mathbf{T}_{BC}$ and $\mathbf{T}_{AC}$. If the transforms found from registration are correct, then $\mathbf{T}_{AC}$ is equivalent to $\mathbf{T}_{AB}$ followed by $\mathbf{T}_{BC}$, as they both rotate the imaging system from A to C.

ϵ, between the two point-clouds is then the mean of the individual angular differences, which is a value between 0° and 180°. If

$$\epsilon \leq \tau_\epsilon, \tag{8.1}$$

for some margin-of-error τ_ϵ, then the point-cloud alignment is classified as correct, otherwise it is deemed misaligned.

8.1.3 Transform Transitivity

The transforms that separate the poses of an imaging system (or the acquired point-clouds) are deemed correct if they are consistent with each other. Consistency occurs when the combination of two transforms matches a third; this transitive relationship between transforms is illustrated in Figure 8.1. Transitivity will not occur if one or more of the transforms are incorrect, as this makes them inconsistent. While transitive verification requires a minimum of two transforms to be combined together and compared against a third, it can also be performed when three or more transforms are combined together. The verification approached described herein only examines the case of transform *triplets*, in which one transform is written as the combination of the other two.

The three transforms that comprise a triplet each contain a rotation and a translation. Because the rotation and translation are calculated independently, their transitivity is also calculated independently; while transitive verification can use either the rotations or translations between poses, using both in conjunction makes the classification more robust. Therefore, both the rotation and translation transitivity must be consistent for the transform triplet to be classified consistent. When transitivity is inconsistent, at least one of the three transforms misaligns its two corresponding point-clouds. The converse is not necessarily true, as consistent transitivity may coincidently be formed from two or more incorrect transforms, though this possibility is minute as both the rotation and translation transitivity tests must agree.

Rotation transitivity is determined for a triplet of point-clouds, A, B and C, by first identifying the rotations $\mathbf{R}_{AB}$, $\mathbf{R}_{AC}$ and $\mathbf{R}_{BC}$, which connects them together, where $\mathbf{R}_{AB}$ would rotate point-cloud A into rotational alignment with point-cloud B. These three rotations are then said to be consistent if

$$\mathbf{R}_{AC} = \mathbf{R}_{AB}\mathbf{R}_{BC}. \tag{8.2}$$

The disadvantage of this definition is that the rotations found by spherical-harmonic correlation come from a discrete set, resulting in quantisation error. Therefore, even if the coarse rotational alignment is deemed correct, the above definition will produce an inconsistent triplet if $\mathbf{R}_{AC}$ and $\mathbf{R}_{AB}\mathbf{R}_{BC}$ do not precisely agree. In this situation, there is no guarantee that inconsistency is proof of misalignment.

To arrive at a better definition of rotation consistency, let

$$\mathbf{R}_{S} = \mathbf{R}_{AC}^{\intercal}\mathbf{R}_{AB}\mathbf{R}_{BC}, \tag{8.3}$$

which is the rotational difference between $\mathbf{R}_{AC}^{\intercal}$ and $\mathbf{R}_{AB}\mathbf{R}_{BC}$. This works because $\mathbf{R}_{AC}^{\intercal} = \mathbf{R}_{AC}^{-1} = \mathbf{R}_{CA}$. The angular portion, θ, of this rotational difference is extracted from $\mathbf{R}_{S}$ by

$$\theta = \cos^{-1}\left(\frac{\mathrm{Tr}\left(\mathbf{R}_{S}\right) - 1}{2}\right). \tag{8.4}$$

If

$$\theta \leq \tau_{\mathbf{R}} \tag{8.5}$$

for some small margin-of-error $\tau_{\mathbf{R}}$, then the triplet comprising rotations $\mathbf{R}_{AB}$, $\mathbf{R}_{AC}$ and $\mathbf{R}_{BC}$ is deemed rotationally consistent. If the rotations of a triplet are consistent, then the allowable error $\tau_{\mathbf{R}}$ has an inverse relationship with the maximum rotation correlation bandwidth, that is, as the correlation bandwidth increases, $\tau_{\mathbf{R}}$ can be decreased, limiting false positives.

The process for translation transitivity is similar to that of rotation transitivity. If, once again, the triplet is comprised of the point-clouds, A, B and C, the translational shifts connecting them together are given as $\mathbf{t}_{AB}$, $\mathbf{t}_{AC}$ and $\mathbf{t}_{BC}$, where $\mathbf{t}_{BC}$ would be the translation which shifts point-cloud B into translational alignment with C. These three translations are said to be consistent if

$$\mathbf{t}_{AC} = \mathbf{t}_{AB}\mathbf{R}_{BC} + \mathbf{t}_{BC}. \tag{8.6}$$

The order in which translations link the point-clouds must be taken into consideration as translations cannot be simply added together. This is because a translation between two point-clouds is performed with respect to the point-cloud that remains stationary; joining two translations together requires that they both be performed with respect to the same stationary point-cloud. A visual illustration of this is presented in Figure 8.2, in which translations $\mathbf{t}_{AC}$ and $\mathbf{t}_{BC}$ are performed with respect to point-cloud C, whereas $\mathbf{t}_{AB}$ is performed with respect to point-cloud B. The rotation, $\mathbf{R}_{BC}$, applied to $\mathbf{t}_{AB}$ ensures that it is also performed with respect to point-cloud C. It is this rotation that permits the two translations to be correctly added together.

Like rotation transitivity, translations found via translation correlation come from a discrete set, and contain quantisation error, therefore, translational consistency is better defined by letting

$$\mathbf{t}_S = \mathbf{t}_{AC} - (\mathbf{t}_{AB}\mathbf{R}_{BC} + \mathbf{t}_{BC}) \tag{8.7}$$

be the translation separation between $\mathbf{t}_{AC}$ and $\mathbf{t}_{AB}\mathbf{R}_{BC} + \mathbf{t}_{BC}$. The separation error is then given as

$$||\mathbf{t}_S||, \tag{8.8}$$

where if

$$||\mathbf{t}_S|| \leq \tau_{\mathbf{t}}, \tag{8.9}$$

for some margin-of-error $\tau_{\mathbf{t}}$, then the triplet comprising of $\mathbf{t}_{AB}$, $\mathbf{t}_{AC}$ and $\mathbf{t}_{BC}$ is deemed translationally consistent.

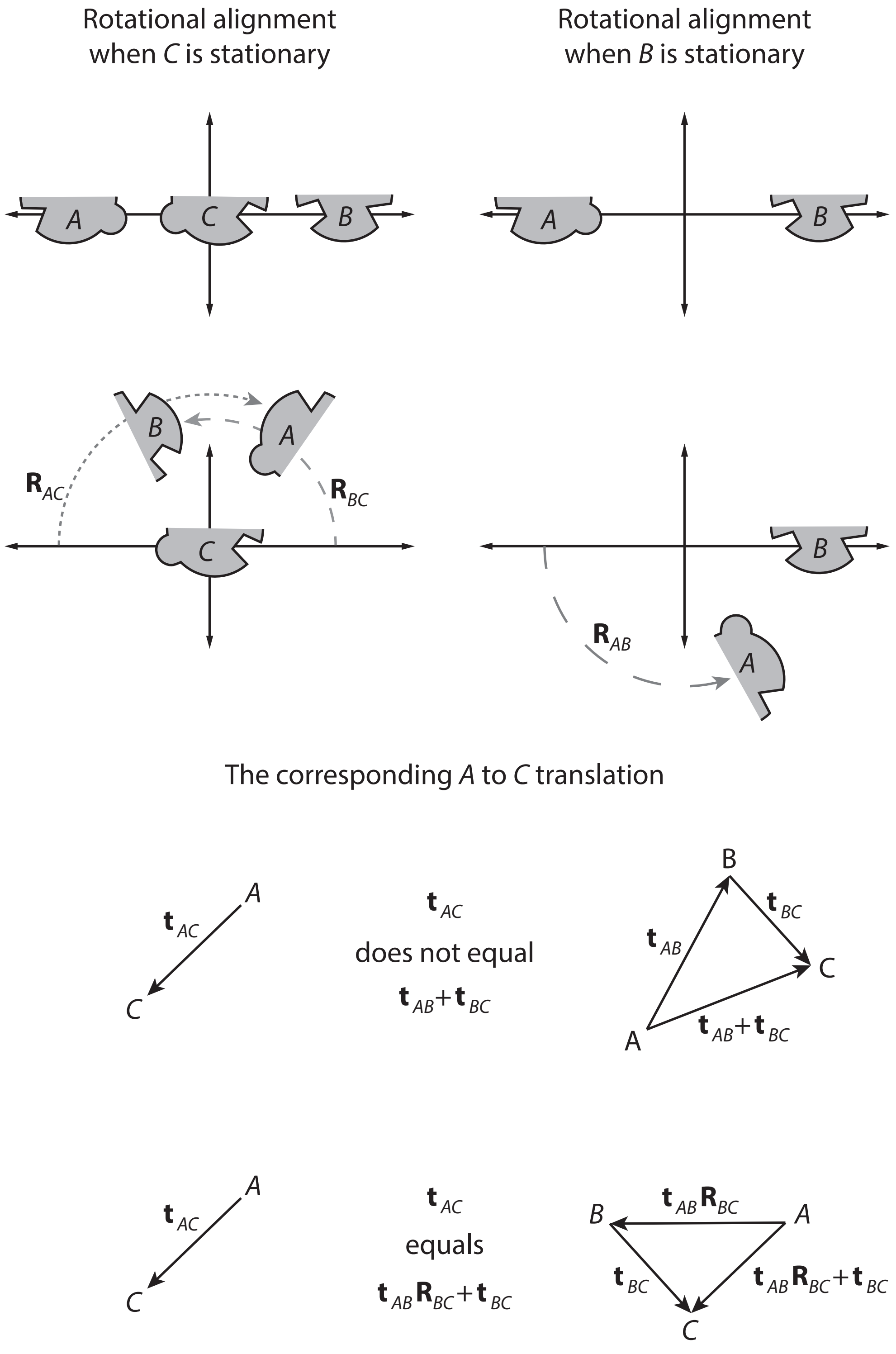


Figure 8.2: Visual illustration of why the translation between point-clouds A and B needs to be rotationally aligned with respect to point-cloud C for transform transitivity to work.

Classifying Individual Transforms

The three transforms that comprise a triplet all receive the same classification as the triplet. If the triplet is consistent, then it is expected that all three transforms are correct, whereas, if the triplet is inconsistent, then up to two of the transforms may still be correct. A total of

$$n_T = \frac{n(n-1)}{2} \tag{8.10}$$

transforms are formed from n acquired point-clouds, with these transforms producing

$$\frac{n\,(n-1)\,(n-2)}{6} \tag{8.11}$$

triplets. The classification of an individual transform is inferred from the $n-2$ triplets to which it contributes. If all $n-2$ triplets are classified as consistent, then it is highly likely the transform has been correctly deduced, however, if the triplets are all inconsistent, there is no guarantee that the transform is incorrect as triplet inconsistency may result from other incorrect transforms. But as the transform only contributes to inconsistent triplets, it cannot be determined correct, thus is treated as incorrect. For all other situations a threshold is specified with the transform being deemed correct if the number of consistent triplets it contributes to are equal to or greater than the threshold. Three thresholds of differing strictness are investigated herein, namely,

$$\tau_a = 1, \tag{8.12}$$

$$\tau_b = \left\lceil \frac{n-1}{2} \right\rceil, \tag{8.13}$$

and

$$\tau_c = n - 2. \tag{8.14}$$

Verification Worst Case

Verification by way of transform transitivity fails when every triplet contains one or more incorrect transforms. If a transform is incorrect, then all $n-2$ triplets that it contributes to are inconsistent. The minimum number of incorrect transforms that render all triplets inconsistent in the worst case scenario is

$$q = \left\lfloor \frac{(n-1)^2}{4} \right\rfloor. \tag{8.15}$$

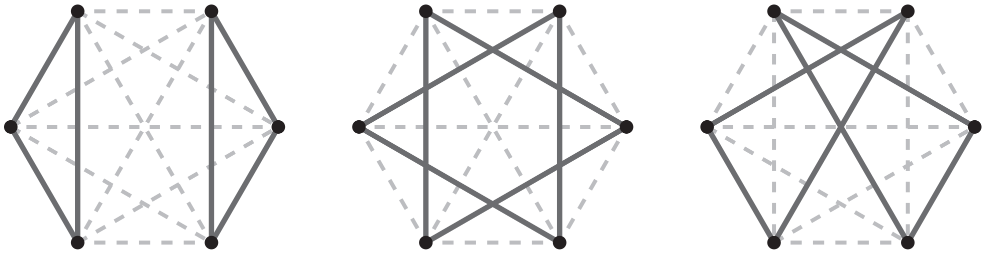


Figure 8.3: In the case of six acquisitions, fifteen transforms are formed. The shown combinations are three of the ten worst case scenarios that occur when six transforms (solid lines) are incorrect. For six acquisitions, six incorrect transforms are the minimum required to cause complete triplet verification failure.

The total number of unique transform combinations comprising of q transforms is given by the binomial coefficient,

$$\binom{n_T}{q}. \tag{8.16}$$

Fortunately, complete inconsistency only occurs for a subset of size

$$\binom{2h-1}{h} \tag{8.17}$$

of these combinations, where

$$h = \left\lfloor \frac{(n+1)}{2} \right\rfloor. \tag{8.18}$$

Figure 8.3 provides an example of the incorrect transform combinations for $n = 6$ that cause all triplets to be inconsistent, in which three of the ten possible arrangements of six incorrect transforms are shown.

8.2 Alternative Rotation Selection

If a particular point-cloud pair has been deemed misaligned by the aforementioned verification approaches, then an alternative transform is required. Typically the translation found between two point-clouds will be incorrect when the rotation is incorrect, therefore, an alternative rotation is needed. The alternative rotation is selected by finding the next largest impulse response in the rotation correlation matrix. Simply selecting the next largest impulse response may identify a rotation that is similar to a previously tested rotation, therefore to mitigate this outcome, the newly selected rotation should be sufficiently different.

Expressing the two rotations in matrix form allows them both to be treated as point sets containing three points each; this is possible as they will remain the same when multiplied with a point set that is identical to the identity matrix. The angle of the optimal rotation that aligns one of these point sets with the other is the measure of difference between the two rotations; this can be achieved by using the Kabsch algorithm (Kabsch, 1976). If this found angle is less than a given threshold, the newly selected rotation does not differ enough from the list of previously tested rotations and is therefore discarded and the selection process repeated. Otherwise, the rotation is added to the tested list followed by applying it to the second point-cloud and re-performing the translation correlation and verification. If the resulting alignment passes verification, then the process of selecting an alternative rotation is finished, otherwise the selection process is repeated. The selection process is stopped and the alignment classified as misaligned if a suitable alignment cannot be found within a predefined number of tests.

8.3 Multiple Point-Cloud Registration

If a suitable alignment between two point-clouds cannot be found, they can still be aligned together using intermediary point-clouds. Such situations occur when imaging around an object, as an acquisition on one side of an object will have minimal overlap with an acquisition on the opposite side. The transforms classified as aligned can form a network of transforms that link point-clouds together. By keeping one point-cloud fixed, all other point-clouds in the network can be brought into alignment with it, even if they are not directly linked. This is achieved for each point-cloud by traversing the network to find the shortest path between it and the fixed point-cloud. Once the path for each point-cloud is found, global alignment of all point-clouds in the network is achieved. Because alignments are coarse, linking multiple transforms together may have the side effect of increasing alignment error between point-clouds. This is due to the error from each transform in the path potentially compounding. A secondary issue occurs when no transforms are classified as correct for aligning a point-cloud to the network. In this situation, the best *misaligned* transform linking this point-cloud to the network could be retested with less restrictive criteria. Alternatively manual verification or registration may be applicable.

8.4 Methodology

The alignment verification approaches described in this chapter are tested to determine their accuracy when classifying individual point-cloud pair alignments as being correct or misaligned. Each approach is independently tested using point-cloud pairs that have been registered using an appropriate set of parameters, thus establishing their respective capability and limitations. Because the ground-truth classification of each point-cloud pair alignment is known a priori, the verification classification is compared against the ground-truth allowing the four possible outcomes to be counted. These four outcomes are true positive (TP), true negative (TN), false positive (FP) and false negative (FN), with their relationship being presented in Table 8.1. The true-positive fraction (TPF) and false-positive fraction (FPF) are then calculated as

$$\mathrm{TPF} = \mathrm{TP}\ /\ (\mathrm{TP} + \mathrm{FN}), \tag{8.19}$$

and

$$\mathrm{FPF} = \mathrm{FP}\ /\ (\mathrm{FP} + \mathrm{TN}), \tag{8.20}$$

respectively.

Each verification approach is analysed by way of receiver operating characteristic (ROC) curves. ROC curves illustrate how the binary classification varies with respect to a particular threshold. Varying the threshold between its lower and upper bounds reveals how the classification responds. The threshold deemed optimal is the one that maximises

$$\frac{\mathrm{TPF} + 1 - \mathrm{FPF}}{2}, \tag{8.21}$$

which is the mean of the true positive and true negative fractions. This measure is chosen as it equally weights the two fractions, making no assumptions whether one should have more influence.

Applying a threshold to the TCV classifies a point-cloud pair alignment as correct if the TCV is equal to or greater than the threshold, otherwise it is classified as misaligned. Because the TCV varies with the structure of the two point-clouds and how well they match, the optimal threshold will fluctuate between point-cloud pair alignments. However, as there is currently no means of automatically adjusting the threshold, a fixed threshold will be used to test the classification performance across all point-cloud pairs. The TCV verification approach is examined using a selection of the datasets previously used to

Table 8.1: The four possible outcomes when verification classification is compared with the ground-truth classification.

Verification Classification	**Ground-Truth Classification**	
	Aligned	Misaligned
Aligned	True Positive	False Positive
Misaligned	False Negative	True Negative

test the translation correlation; Section 7.2 provides the details about the formation of these datasets. The tested angles of rotational misalignment are 0°, 2°, 5°, 10° and 15°; a rotational misalignment of 1° is not included as it does not sufficiently differ from 0° and 2°. Voxelisation is performed by segmenting the point-cloud pair into 64 voxels along each side of the histogram cube, allowing an adequate level of accuracy for fine registration while constraining the computational cost. A point-cloud pair is deemed correctly aligned if the resulting translation error is within 15 mean distances (see Section 7.2). A single *mean distance* is the mean distance between neighbouring points of the model from which the point-clouds are extracted. The impact of noise on the TCV is also examined by using translation correlation datasets with Gaussian noise with a standard deviation of three mean distances added; this is also tested at the five angles of rotational misalignment.

Surface orientation consistency is examined in a similar manner to the translation correlation value, by using the same datasets. The five tested angles of rotational misalignment are used to simulate the effect that discretisation error has on the rotation correlation. Artificial Gaussian noise is not tested due to the difficulty of adding it in a manner that relates it to both the points and the normals, as they are inherently connected. A threshold is applied to the found difference in surface orientation, where if it is equal to or less than the threshold, the point-cloud pair alignment is classified as correct, otherwise it is classified as misaligned. The voxelisation of each aligned point-cloud pair is performed using 64 voxels along each side of the histogram cube. The ground-truth alignment remains the same as the TCV: a point-cloud pair alignment is deemed correct if its translation error is within 15 mean distances.

The testing of the transform transitivity verification only examines the number of consistent triplets that are formed from the transforms that link all point-clouds in a set together. Due to only the transforms being relevant, testing with multiple datasets and parameters is unnecessary as the only factor that changes is the percentage of correctly aligned point-cloud pairs drawn from the overall set. The four tested datasets, each contain 120 point-clouds, were constructed by registering the dragon dataset using different parameters. The overall percentage of correctly aligned point-cloud pairs for these four datasets are 23%, 37%, 60% and 79%. The transform linking a point-cloud pair together is deemed correct if its rotational error is equal to or less than 5°, and its translation error is within 15 mean distances when the point-clouds are voxelised with 64 voxels along each side of the histogramming cube.

The accuracy of the transform transitivity is further tested by varying the number of available point-clouds, for which three counts of point-clouds are used: 6, 12 and 20. To provide a sufficient assessment, the 12 combinations of point-cloud count and percentage of aligned point-cloud pairs are each evaluated 10 000 times. It must be noted that the percentage of correct point-cloud pairs is dependent upon the randomly selected point-clouds from each dataset, therefore, the mean and standard deviation of correct alignment percentages are taken across the 10 000 evaluations and are presented in the results section for each combination. An individual evaluation is constructed by randomly selecting the predefined number of point-clouds from the set of 120 point-clouds, and then performing the transitivity verification. The counts of true positive, true negative, false positive and false negative results for each evaluation are individually summed together for each threshold. The threshold determines the minimum number of consistent triplets that a transform needs to belong for it to be classified as correct. The TPR and FPR for each threshold of the 12 combinations are calculated to form ROC curves.

The benefit that alternative rotation selection provides to the registration pipeline is examined by counting the number of additional point-cloud pairs that it brings into correct alignment. An alignment is correct if both its rotation and translation error are within 5° and 15 mean distances, respectively. The rotation correlation is performed using all normals from each point-cloud, with each bin being normalised by its size. The initial rotation aligns approximately 31% of the point-cloud pairs, providing alternative rotation selection the opportunity to correctly align the remaining 69%. If correct alignment is found by the tenth rotation check, excluding the initial rotation, then alternative rotation selection aids the registration pipeline. A rotation is discarded and not included as a rotation check if it is within 5° of a previously tested rotation.

8.5 Results

The three verification approaches are evaluated by measuring their accuracy when classifying alignments. The verification approaches are the translation correlation value, surface orientation consistency and transform transitivity. The benefit provided by alternative rotation is also examined.

8.5.1 Translation Correlation Value

Thresholding the TCV is suitable for classifying alignments, especially if noise and rotational misalignment are minimised. The ROC curve presented in Figure 8.4 shows the TCV verification performance when point-clouds are uncontaminated by noise. When no rotational misalignment is present, the optimal threshold correctly classifies the alignment 96% of the time. As the rotational misalignment increases, the performance decreases as the separation between the aligned and misaligned TCVs becomes less pronounced. This is shown by the classification dropping to 90% and then 74% when the rotational misalignment is 5° and 10°, respectively. A rotational misalignment of 15° would be misaligned, but as the ground-truth classification is based on translational error only, correct alignments can still be achieved. At this error, a point-cloud pair alignment is correctly classified 66% of the time when using the optimal threshold. The optimal threshold value does differ with rotational misalignment, though the difference is minimal; when the rotational misalignment is 0° the threshold is 0.059, which lowers to 0.032 as the rotational misalignment increases to 15°. This lowering of threshold value is related to the overall decrease in alignment quality as rotational misalignment increases.

Contaminating the points with Gaussian noise ($\sigma = 3$ mean distances), as shown in Figure 8.5, reduces the TCV separation between aligned and misaligned point-cloud pairs, diminishing the achieved correct classification of the TCV threshold. The TCV threshold correctly classifies 88% of the point-cloud pairs when the noise is combined with 0° of rotational misalignment. The percentage of correct classifications, when the rotational misalignment is increased to 5°, 10° and 15°, is 83%, 74% and 69%, respectively. The added noise combined with a rotational misalignment of 0° has an optimal threshold value of 0.025, this threshold also lowers as rotational misalignment increases, which at 15° is 0.022. Once again, this difference in threshold value is minimal, thus allowing a single fixed value to adequately classify point-cloud pairs with any amount of rotational misalignment.

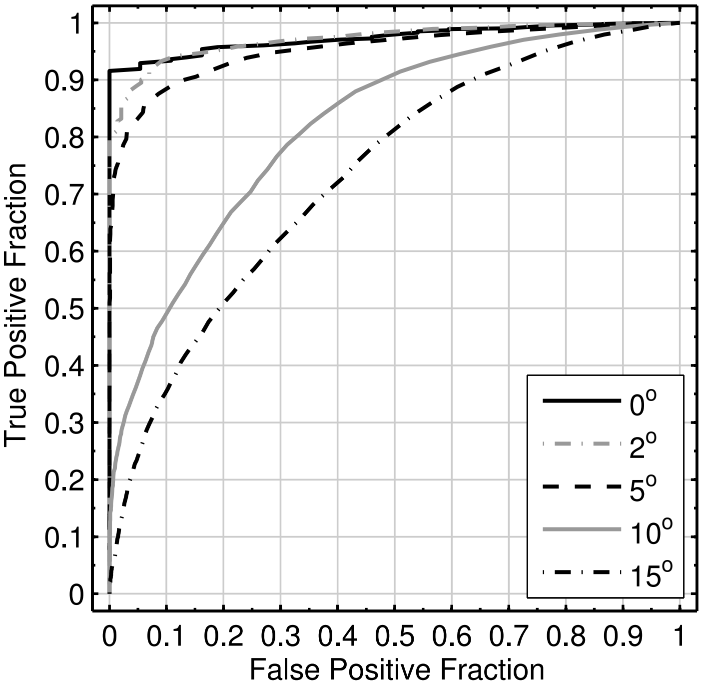


Figure 8.4: ROC curve showing how well different TCV thresholds work for classifying point-cloud pairs at varying rotational misalignments.

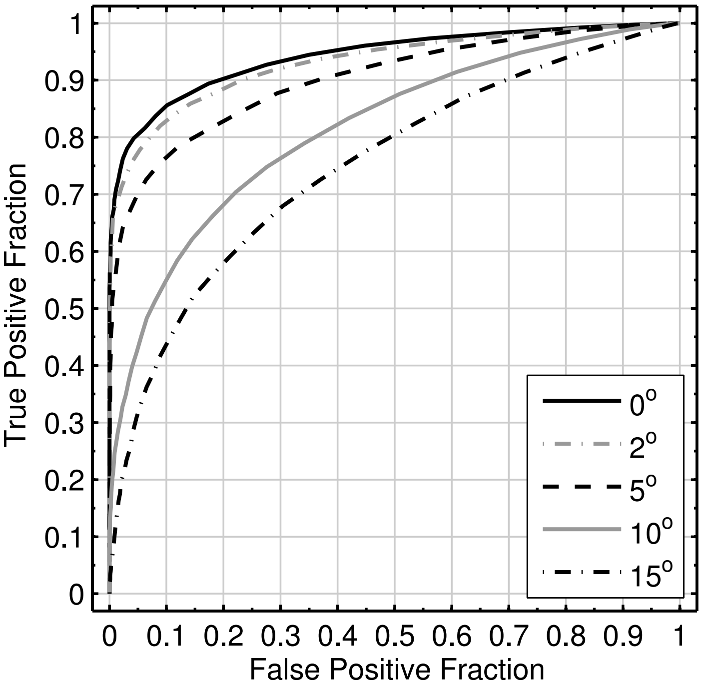


Figure 8.5: ROC curve showing how well different TCV thresholds work for classifying point-cloud pairs at varying rotational misalignments when Gaussian noise has been introduced to the points.

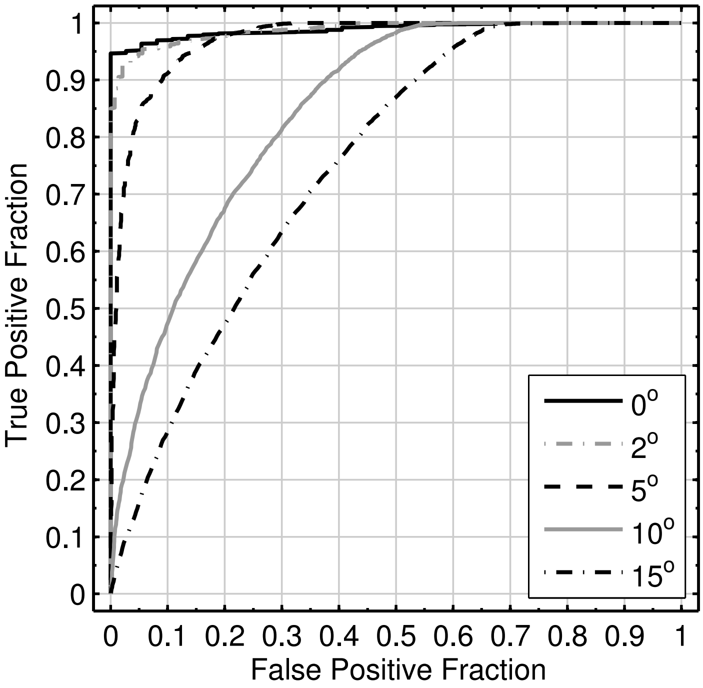


Figure 8.6: ROC curve showing the performance of surface orientation consistency at varying rotational misalignments.

8.5.2 Surface Orientation Consistency

The level of classification accuracy that surface orientation consistency provides is dependent upon the amount of rotational misalignment between two point-clouds. Figure 8.6 shows that when there is no rotational misalignment, a point-cloud pair has a 97% chance of being correctly classified when applying the optimal threshold. The classification accuracy at a rotational misalignment of 2° drops slightly to 95%, with the ROC curve following a similar trend as that of 0°. Increasing the rotational misalignment further causes the classification trend to change, with 5° reaching a true positive fraction of one sooner than 0° or 2°. The classification accuracy of 5° does decrease though, with 90% of the point-cloud pairs being correctly classified; at 10°, the accuracy reaches 76%. While the classification accuracy can perform well, surface orientation consistency is prone to false positive classifications as rotational misalignment increases. The discussion below outlines why false positives are detrimental to multiple point-cloud registration, and describes aspects of the surface orientation consistency that should be taken into consideration.

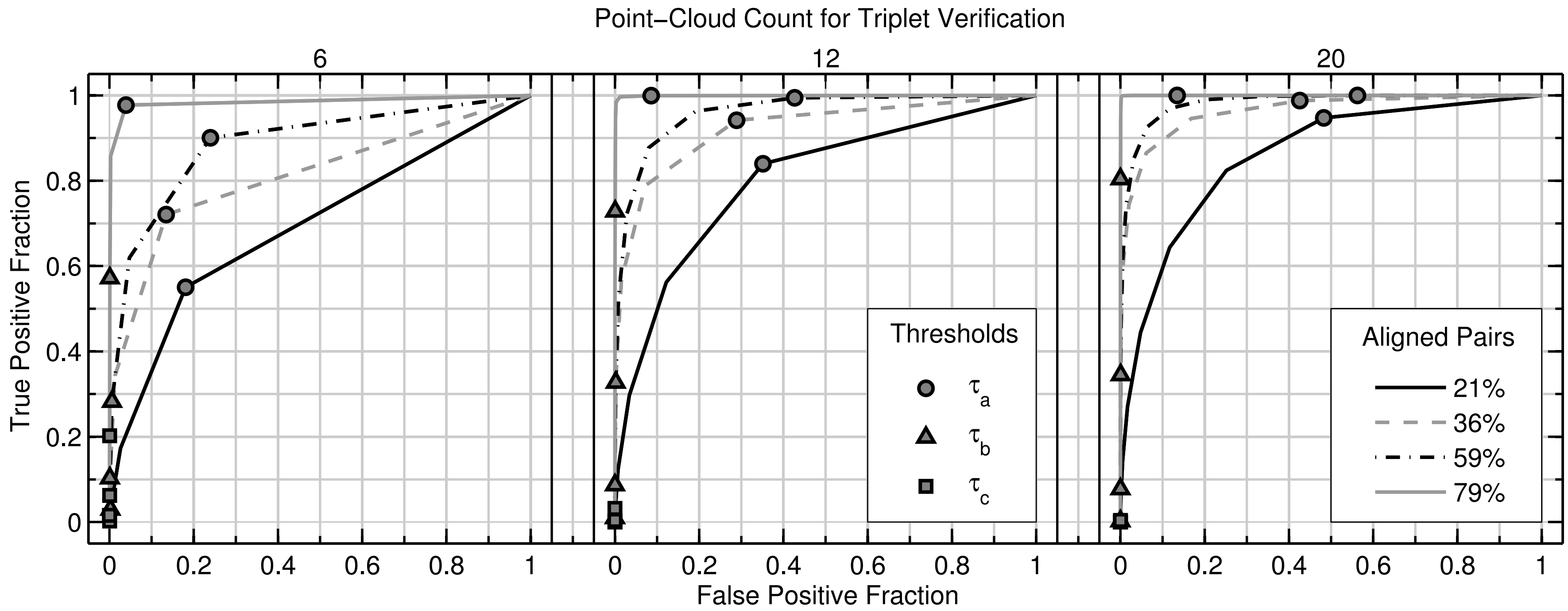


Figure 8.7: ROC curves presenting the classification accuracy of transform transitivity for three individual point-cloud counts. Each curve shows the performance when using a particular percentage of correct point-cloud pairs, along with three possible classification thresholds.

8.5.3 Transform Transitivity

Transform transitivity differs from the other verification approaches as it needs a minimum of three point-clouds for classification. The greater the point-cloud count, the more triplets formed for testing transitivity, thus benefiting verification as there is more cross-checking. This benefit does diminish as the number of point-clouds increases as the greater level of cross-checking becomes redundant. Figure 8.7 shows the performance of three point-cloud counts, 6, 12 and 20, used for triplet verification; there is a greater classification improvement between a point-cloud count of 6 and 12, but this improvement is minimal between 12 and 20.

The aspect with the greatest influence on performance is the percentage of correctly aligned point-cloud pairs used to perform the transitivity. When the percentage is lower, a triplet has a greater chance of containing an incorrect transform, which makes it inconsistent; the higher the percentage, the better the classification accuracy. The percentage of correct point-cloud pair alignments for each of the four tested datasets are presented in Figure 8.7, which are determined as the mean of correctly aligned point-cloud pairs from each of the 10 000 evaluations. The associated standard deviation primarily relates to the point-cloud count instead of the dataset, and is approximately 13%, 7% and 4% for the point-cloud counts of 6, 12 and 20, respectively.

Classifying the transform between a point-cloud pair using transform transitivity requires a specified threshold, where if the number of consistent triplets that the transform contributes to is equal to or greater than this threshold, then it is classified as correct. The ROC curves presented in Figure 8.7 are formed using every available threshold, though three thresholds, τ_a, τ_b and τ_c (see Section 8.1.3), are specified to determine if any one threshold is suitable. On the whole, only τ_a is suitable, though as point-cloud count increases so does its false positive fraction; thresholds τ_b and τ_c are far too restrictive to be viable. The results show that as the point-cloud count increases, the threshold should also be increased accordingly. As the number of point-clouds to register is known, a suitable threshold can be determined and applied that is more appropriate than the three specified.

8.5.4 Alternative Rotation Selection

The ability of alternative rotation selection to identify the appropriate rotation is dependent upon the formed correlation matrix. The impulse response for each rotation in the matrix is based on the representation quality of the normals used when performing the correlation. Therefore, alternative rotation selection will only find a suitable rotation when its impulse response is similar in size to the global maximum impulse. Testing the alternative rotation selection found that there are instances where checking through the sorted impulse responses will identify a rotation that brings the point-cloud pair into alignment.

Using only the global maximum rotation, 31% of the point-cloud pairs in the tested dataset were correctly aligned. By allowing five or ten alternative rotations to be checked, the number of correctly aligned point-cloud pairs increased to 36% and 39%, respectively. Further increasing the allowable number of checks may correctly align more point-cloud pairs, but the total number of rotations that are evaluated and discarded will grow. When five rotation checks are performed, 16 rotations on average are discarded because they are too similar to previously tested rotations. Increasing the number of checks to 10 further increases the number of discarded rotations to 46. The number of discarded rotations grows at a faster rate than the number of checked rotations as similar rotations will also have similar sized impulse response. While alternative rotation selection is beneficial, its computational cost is its primary limitation; this aspect was not revealed by Makadia *et al.* (2006) as they did not analyse alternative rotation selection in any depth.

8.5.5 Efficiency

The three verification approaches have their own computational costs, which are dependent upon what needs to be performed to evaluate whether correct alignment has been achieved. The cost of each verification approach is outlined, followed by the cost of performing alternative rotation selection.

The TCV is a cost-free verification approach as it is a by-product of the translation correlation; all that it needs for classification is a comparison with a threshold. Its only cost occurs if an initial evaluation needs to be performed to ascertain an appropriate threshold value. The cost of performing surface orientation consistency is dependent upon two primary factors: the number of voxel divisions and the total number of points across both point-clouds. The individual steps consist of binning the points, independently taking the mean of the normals in each voxel for both point-clouds along with their angular difference, then taking the mean of these angular differences. In the worst case scenario, this will have a cost of $\mathrm{O}\big(n + v\big)$, where n is the total number of normals and v is the total number of voxels. Transform transitivity relies on triplets of point-clouds being tested for consistency, with each triplet comprising of three transforms, two of which get combined together and compared with the third. The combination and comparison requires minimal processing to perform, therefore, the number of triplets determines cost. Given m point-clouds, the cost based on the number of triplets formed is $\mathrm{O}\big(m^3\big)$.

Alternative rotation selection can be an expensive extension to the registration pipeline, especially given that it requires both the translation correlation and verification stages to be recalculated for every additional rotation check. This cost does not include assessing the difference between the new rotation and those that have been previously checked. While this cost can escalate, the results show that alternative rotation selection can be a benefit to the registration pipeline, as long as the number of alternative rotational checks is limited to the required computational performance of the registration process.

8.6 Discussion

Correctly aligning every point-cloud pair, while advantageous, is not crucial if the transforms classified as correct can be combined to bring all point-clouds into cohesive alignment with the point-cloud that remains fixed. This is because many of the transforms linking point-clouds together are redundant as global alignment can be achieved without them. However, global alignment can only be achieved if there is a path of correct transforms between a given point-cloud and the fixed point-cloud. Attempting to register all point-cloud pairs increases the likelihood of global alignment occurring as the additional transforms provide alternative paths that allow incorrect transforms to be circumvented. To ensure that the selected paths result in the desired alignment, it is better if the verification approaches limit false positive classifications at the cost of producing a higher false negative rate.

Verifying that point-clouds have been correctly aligned using the TCV requires a suitable threshold to perform the classification. The TCV is the size of the maximum impulse response once the translation correlation has been performed, and as such, impulse size is governed by all aspects that influence the translation. These include overlap, rotational misalignment, noise and voxel count. Therefore, a fixed threshold will not be optimal for all point-cloud pairs, as it will need to be adapted to point-cloud quality and the chosen registration parameters. The threshold cannot be easily adapted to the overlap between point-cloud pairs as the TCV fluctuates with the other aforementioned aspects, meaning that overlap can vary while still producing the same TCV. Currently there is no automatic means of identifying the optimal TCV threshold for a given set of point-clouds, thus, a suitable threshold needs to be determined manually. This is achieved by *calibrating* the TCV by visually classifying both correct and misaligned point-cloud pair alignments.

Surface orientation consistency was shown to be a viable verification strategy for classifying alignments, though it does need further evaluation to better define how it performs in all cases. This is because the results produced by the evaluation revealed an unusual trend with 5° rotational misalignment reaching a true positive fraction of one sooner than either 0° or 2°. At present the cause of this trend is undetermined as there are many factors influencing the found difference in surface orientation; these include overlap, voxel count, both rotational and translational misalignment, and noise. Because the consistency takes the mean angular difference between normals from both point-clouds, only voxels containing normals from both point-clouds can be used. This means that if the overlap is low, the verification will not be as robust as when overlap is higher. While increasing voxel count will provide a better measure of consistency, it is limited by rotational and translational misalignment, and noise. This is because misalignment and noise can change which normals occur in a given voxel, causing unrelated normals to be compared or limiting the number of normals that can be compared. Surface orientation consistency may provide a better classification once fine registration has been performed, as points common to both point-clouds are more likely to occur in the same voxel and their normals will have closer matching orientations.

The transform transitivity between a set of point-clouds has been shown capable of providing very accurate verification, with some instances being able to correctly classify 100% of the transforms. The advantage of this approach is that it only requires the found transforms between point-clouds to perform the verification and is therefore very efficient. Although transform transitivity has these advantages, it does have its own inherent limitations that separate it from the other verification strategies. The primary limitation is that it requires a minimum of three point-clouds before transitivity can be evaluated. Because transitivity uses multiple point-clouds to perform verification, its reliability is intrinsically linked to the number of point-clouds in the set, as this dictates the amount of cross-checking that can be performed on an individual transform. The benefit provided by cross-checking does taper off as point-cloud count increases, but the number of point-clouds does influence the overall reliability of each transform classification. Another limitation is that the accuracy of the transitivity is influenced by the percentage of correct alignments. This is because it only takes one incorrect transform to cause a triplet to be inconsistent, therefore, more triplets in a set will be found inconsistent as the number of incorrect transforms increases. The result is that more false negative classifications occur as there is a greater chance of a correct transform being a member of inconsistent triplets. Transform transitivity is not limited to using only point-cloud triplets, as any number of transforms can be combined and compared with a single transform. This capability allows transform consistency to be checked at a greater level, which may be necessary when triplets are unable to be formed due to the overall overlap between acquisitions.

Alternative rotation selection can be a useful addition to the registration pipeline, as it automatically selects and tests an alternative rotation if the current one is incorrect. Because alternative selection starts at the largest impulse response and works its way down, this approach provides the most benefit when the impulse response of the correct rotation is nearly the same size as the largest impulse. Therefore, alternative selection is dependent upon the representation of the normals used for the rotation correlation, as this determines the impulse size of each rotation. Due to this, it is still better to improve the representation of the point-cloud surfaces and normals that are correlated, even though improving the correlation will limit the benefit of alternative rotation selection. In theory, alternative rotation selection will eventually find the correct coarse rotation regardless of its impulse size, though achieving this would be a computationally expensive endeavour as at a correlation bandwidth of 128 there are over 16 million selectable rotations.

The artificial Gaussian noise applied to the points for testing the verification based on the TCV threshold provided a minor improvement to the ROC curve when the rotational misalignment was 10° or greater. It was expected that the combination of noise and rotational misalignment would further decrease the performance. Instead, what may have happened is that the points placed into the suboptimal voxel by rotational misalignment have been shifted by the noise, causing some to be shifted back into their optimal voxel. Those that are shifted further away from their optimal voxel have less effect on the translation correlation, and in turn the TCV, as their initial voxel is already suboptimal. The applied noise does not provide a true representation of the noise that would occur in a real imaging system for two reasons. The first is that at three times the mean distance between points, the applied noise is assumed to be greater than what an imaging system would introduce. Secondly, the noise introduced by an imaging system will not be Gaussian in all three dimensions, as lens calibration will only allow point error caused by noise to occur along a straight line between it and the imaging sensor. Chapter 9 examines the overall ability of the coarse registration pipeline, including the proposed verification strategies, using real data. These tests will provide an additional insight into how well the verification strategies classify point-cloud alignments that contain noise produced by an imaging system.

8.7 Summary

The three evaluated verification strategies were each capable of correctly classifying the majority of point-cloud pair alignments. Overall, transform transitivity has the best performance, with its ability improving as point-cloud count increases, though this performance is dependent upon the percentage of correctly aligned point-cloud pairs. Both the TCV threshold and surface orientation have a similar performance of approximately 90% classification accuracy when rotational misalignment is within 5°. The advantage that these verification strategies provide the coarse registration pipeline as a whole is that they limit the need to manually verify point-cloud alignments. By being able to automatically check the found alignment permits an alternative registration to be found if the first is incorrect. When applied to a set of point-clouds, these verification strategies combined with multiple point-cloud registration has a greater likelihood of bringing all point-clouds into the correct coarse alignment as incorrect registrations can be circumvented. The addition of registration verification completes the coarse registration pipeline, as the set of point-clouds to be registered together should now be in a position that allows a fine registration algorithm to perform optimally.

Chapter 9

Performance Evaluation Using Real Datasets

The ground-truth datasets used in the previous chapters allowed the performance of different algorithmic aspects and parameters of the coarse registration pipeline to be ascertained without the influence of noise or changes in sampling density. This chapter will instead concisely evaluate the performance of the pipeline using real-world data. Testing with real-world data further characterises the the pipeline, revealing where it works well and where further examination is required to mitigate its limitations. The chapter is broken into two sections: the first describes the datasets, while the second provides an evaluation of the registration performance for each dataset.

9.1 Real World Datasets

This section describes the five real-world datasets collected from different sources and used to evaluate the performance of the coarse registration pipeline. These datasets each have aspects that limit the performance of the registration pipeline. The goal of using potentially difficult datasets is to identify situations where the pipeline needs additional attention and to also show where the pipeline works well.

9.1.1 Angels

The angels dataset, rendered in Figure 9.1, is a high density dataset used by Bonarrigo *et al.* (2011) for testing their algorithm. This dataset was supplied by Bonarrigo *et al.* as they have copyright permission for it. Each of the eight scans that comprise this dataset were stated as being formed by a commercial high-resolution structured-light scanner, which used a 1280×1024 pixel charge-coupled device (CCD) sensor (Bonarrigo *et al.*, 2011). Not all sample points within a given scan were provided, with those occurring outside the object boundary or within shadowed areas having been discarded. The angels in their entirety are approximately 320 mm high and 400 mm wide, which at these dimensions produces very dense sampling.

Figure 9.1: A rendering of the angel model.

9.1.2 Gargoyle

The gargoyle dataset, rendered in Figure 9.2, is comparatively large with respect to the other four datasets as it contains 27 individual point-clouds. This dataset was supplied by Fantoni *et al.*, and was used by them to test their algorithm (Fantoni *et al.*, 2012). It is stated that the sampling distribution changes with each point-cloud, and that sampling noise is present. The number of points in each point-cloud varies, with the smallest having 19 433 points and the largest having 86 286.

Figure 9.2: A rendering of the gargoyle model.

9.1.3 Helicopter Gearbox

The helicopter gearbox dataset was provided for this research by Nova Mechanical Design Ltd (Novamech), a New Zealand based company that specialises in surveying, reverse-engineering and rapid prototyping related services. The six overlapping point-clouds that compose this dataset were said to be scanned from the tail-rotor gearbox-casing of a 1950s helicopter. This acquisition was performed using a ROMER Absolute Arm with an integrated scanner (Hexagon Metrology, 2014). While this system samples the scene with respect to the arm, the manoeuvrability of the arm is limited, therefore, the gearbox needed to be shifted and rotated to different orientations to scan it in its entirety.

Novamech reconstructed the gearbox from these six overlapping segments, giving a full representative model of its original form. This process was performed in two stages via the commercial software package SolidWorks[1]. The first stage required manually selecting at least three common points between pairs of scans, allowing Solidworks to identify the set of transforms which bring the segments into coarse alignment. In the second stage this alignment was refined by using a built-in ICP algorithm. Following these two registration

---

[1] SolidWorks 3D CAD (Computer Aided Drafting) is a commercial software package (http://www.solidworks.com) used to construct 3-D models.

stages, the individual segments were merged together, giving the final model shown in Figure 9.3; a photo of the original gearbox is also presented. The acquisition and registration process used by Novamech is a standard process within the surveying and reverse-engineering industry. Incorporating an automated coarse registration step into this process would minimise the manual input required, thus improving the overall efficiency of the process.

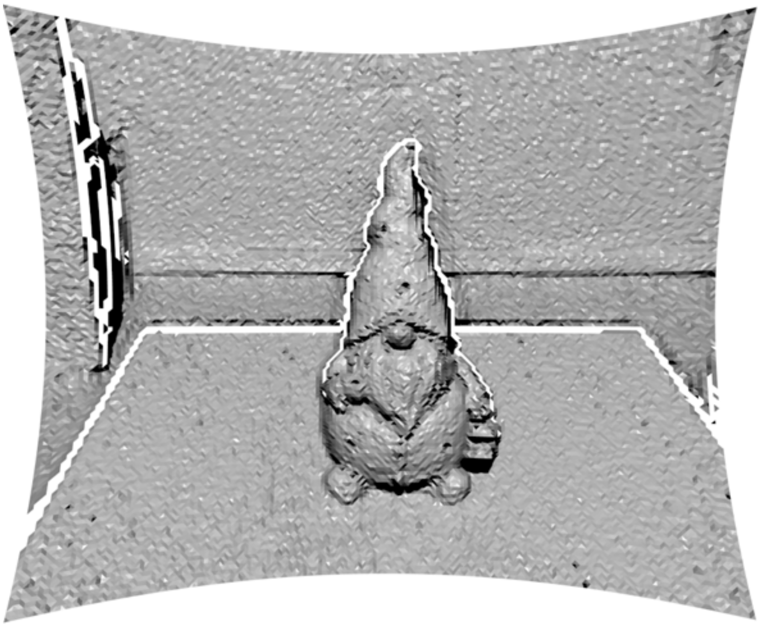
(a) Model of gearbox. (b) Photo of gearbox.

Figure 9.3: The reconstructed model and photo of the helicopter gearbox.

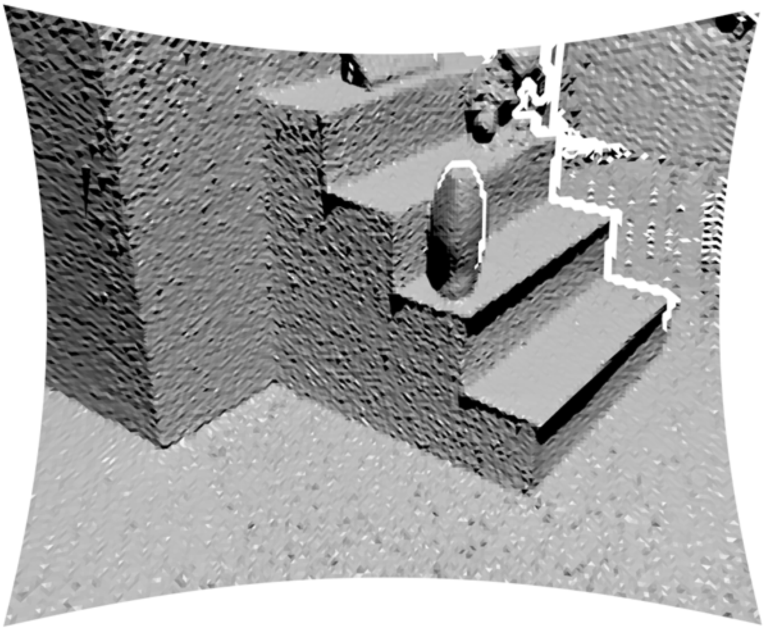
9.1.4 Gnome and Stairs

The gnome and stairs datasets were captured using the Mesa SwissRanger 4000 (SR4K), which is a full-field AMCW range imaging camera. The SR4K has a resolution of 176×144 pixels, and at the set modulation frequency of 30MHz it has an ambiguity distance of 5m. Each image was acquired by taking the mean of each sample point across 100 frames to improve the precision of each sample point. Due to the inherent issues of this system (see Section 1.2), the samples have limited accuracy due to noise. The impact of this noise is that the normals produced by MeshLab do not provide an accurate representation of a surface's orientation. This is because MeshLab produces a normal at a point by taking the average of the normals collected from the faces that include the point; a face is the triangular region formed between three adjacent points. To mitigate this issue, the normals were instead calculated by placing a plane on each point and fitting it to its neighbours contained within a 7×7 window. The normal used for each point is the normal of the fitted plane.

The gnome dataset was acquired by placing the gnome on a square table in an enclosed room. A total of twelve point-clouds were acquired by placing



(a) Gnome point-cloud



(b) Stair point-cloud

Figure 9.4: Rendering of a single point-cloud from each of the Gnome and Stair datasets.

the imaging system at roughly even spacings around the table. Figure 9.4a presents a rendered version of one of these point-clouds. The stairs dataset consists of four steps positioned against an office divider with three objects placed on the steps. The steps were imaged from six poses, with Figure 9.4b showing a rendering of one of the acquired point-clouds.

9.2 Evaluation Results

This section evaluates the performance of the coarse registration pipeline with respect to each of the five real-world datasets. Each dataset tests the pipeline, identifying how well it works for the given data quality and the relationship between the individual point-clouds. The benefit of this additional testing is that it further characterises the reliability of the pipeline in different situations. As part of this evaluation, the strengths and weaknesses of the pipeline with respect to each dataset are reviewed. This includes discussing the steps of the pipeline that require future investigation to reduce these weaknesses and further enhance its reliability.

The pipeline is constructed using the enhancements that were found to provide the most benefit to registration. With respect to this, both the spherical-harmonic transform and correlation bandwidths are set to 128, and the translation correlation bandwidth is also set to 128. Each bin is normalised by the size of its collection area, and a value of $2{\times}10^{-8}$ is used to calculate the percentage of normals required for a bin to receive a value of one instead of zero. The normals are weighted using the local surface curvature as determined by the plane distance weighting. The weighting threshold used to cull the normals is 0.9875, except for the two range imaging datasets, which use 0.95 due to the lower quality sampling.

The triplet transitivity and translation correlation strategies are used for the verification step to better characterise their classification accuracy. The classification accuracy is confirmed by visually inspecting the alignment of each point-cloud pair. In order to produce a concise evaluation of the introduced components of the pipeline, and to limit computational cost, the surface orientation verification and alternative rotation selection are not included.

9.2.1 Angels

The eight point-clouds of the angel dataset produced a total of 28 point-cloud pair combinations, of which ten were correctly aligned in the initial registration

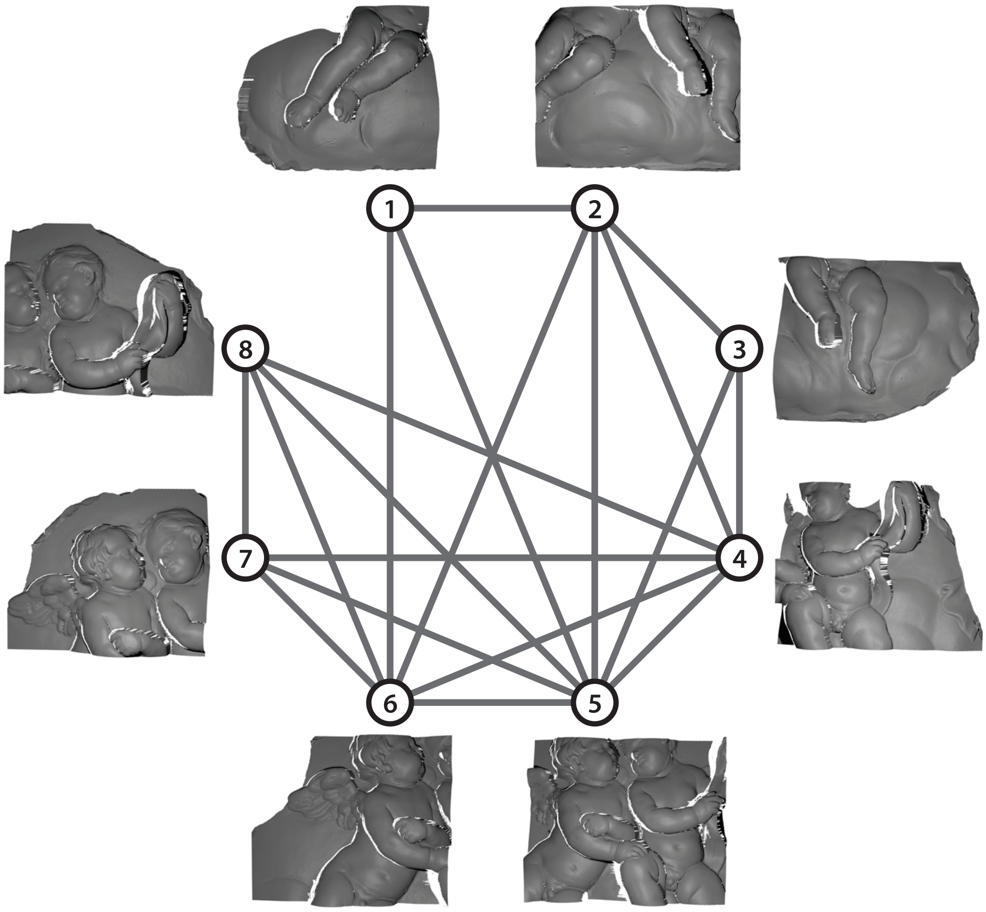


Figure 9.5: Graph showing which angel point-cloud pairs achieved correct alignment.

attempt. All combinations are checked as the registration pipeline has no prior knowledge of acquisition order. The algorithm by Bonarrigo *et al.* (2011) was constrained to only registering consecutive point-cloud pairs that required a minimum of 20% overlap; this order is the same as that shown in Figure 9.5. The initial registration had a similar performance as Bonarrigo *et al.* (2011) for consecutive point-clouds. Visually inspecting each alignment revealed that all point-cloud pairs achieved correct rotational alignment, but the translational alignment was often misaligned. The size of the bounding cube placed around each point-cloud was identified as the cause of the misalignment.

The translational shift that occurs between two point-clouds is limited by the size of the bounding cube. This is because the correlation is only able to select a translation that is up to half a side length of the cube. Setting the size of the bounding cubes to encompass both point-clouds as tightly as possible (see Section 3.3) reduces the coarseness of the translational alignment. However, using this sized bounding box is not conducive for registering all point-cloud pairs, as identified for the angel dataset. The issues with translation arise due to a combination of minimal overlap and the shape of the angel point-clouds. To guarantee that the translation correlation is able to select a translation that is appropriate for all situations, the found side length of the bounding cube was doubled.

Doubling the side lengths of the bounding cube restored the capability of the translation correlation, permitting the coarse registration pipeline to correctly align 19 of the 28 point-cloud pairs. Figure 9.5 presents a graph showing which point-clouds were correctly aligned together. The additional nine correct alignments occur between point-clouds that have minimal overlap, such as that between point-clouds one and five. The nine point-cloud pairs that do not achieve correct alignment are due to their overlap being either minuscule, such as between point-clouds three and six, or non-existent.

Verifying correct alignment was performed using both the translation correlation value (TCV) and triplet transitivity. The TCV can provide an accurate means of verification if an appropriate threshold is selected, which for the angel dataset is 0.03. The mean and standard deviation of the TCV for the correct and incorrect alignments are 0.115 ± 0.065 and 0.021 ± 0.002, respectively. While the TCVs of the correct and incorrect alignments for this dataset can be neatly split, this is not guaranteed for other datasets as the TCV varies with point-cloud shape, overlap and the found alignment. Additionally, as there is currently no method of automatically calculating the threshold, it needs to be manually chosen. Due to these aspects, verifying the alignment via the TCV can produce false positive and negative classifications. Triplet transitivity was able to correctly classify all point-cloud alignments of the angel dataset. The advantage of using transitivity is that it avoids the need for fine-tuning as it is dataset invariant.

9.2.2 Gargoyle

The gargoyle dataset is a failure case for the registration pipeline, with none of the total 351 point-cloud pairs being confirmed as correctly aligned; visually inspecting all point-cloud pairs was not feasible, and as such, there may be correct but unconfirmed alignments. This performance is in contrast with the algorithm by Fantoni *et al.* (2012), which was able to correctly bring the point-clouds into alignment. It is suspected that a variety of factors contribute to the gargoyle being a failure case, with the primary cause being the reduced quality of the normals. MeshLab calculates the normal for a point by averaging the normals from the faces around the point. Combining this small surface area for calculating normals with the rough texture of the gargoyle and lower sample point quality, the normals can be scattered quite extensively, giving a poor representation of surface orientation. Using a high weighting threshold of 0.9875 is a secondary cause that may have compounded the difficulty of point-cloud registration. The result is that the produced representation of the normals is predominantly noise, effectively making the pipeline incapable of identifying the correct rotation.

The gargoyle dataset identifies the need to examine and tailor the calculation of the normals in a manner that provides an accurate representation of surface orientation. Additionally, the pipeline parameters need to be set appropriately for each dataset, especially considering that an automatic means of achieving this has yet to be devised. These issues principally pertain to the rotation steps of the pipeline, with the translation and verification steps being more robust, though the translation is dependent upon the rotation being correct. The triplet transitivity correctly classified all transforms as being misaligned, though this occurs if there is a minimal number of correct transforms available. Depending on the chosen threshold, the TCV classification is a relatively accurate verification strategy, as the overall mean and standard deviation of the TCVs are 0.014 ± 0.006. Using a threshold of 0.03 would still produce nine false positives, with the greatest TCV being 0.056.

9.2.3 Helicopter Gearbox

Registering together the six point-clouds of the helicopter gearbox correctly aligned ten of the fifteen point-cloud pairs. The alignment relationship between each point-cloud is shown in Figure 9.6. The ten correct alignments are easily sufficient for bringing all six point-clouds into the same coordinate system, which is the desired goal. The five incorrect alignments were primarily due to two factors, the first is insufficient overlap, which typically occurs when point-cloud overlap is 10% or less, and is caused by large difference in the capturing poses with respect to the gearbox. The second factor is the structure of the gearbox which produces a distribution of normals which are near symmetric, resulting in an incorrect rotation appearing to be the appropriate solution. Figure 9.7 provides examples of incorrect rotations due to symmetry within the normals between point-clouds one and four, and two and three. In this situation, the incorporation of alternative rotation selection may be beneficial.

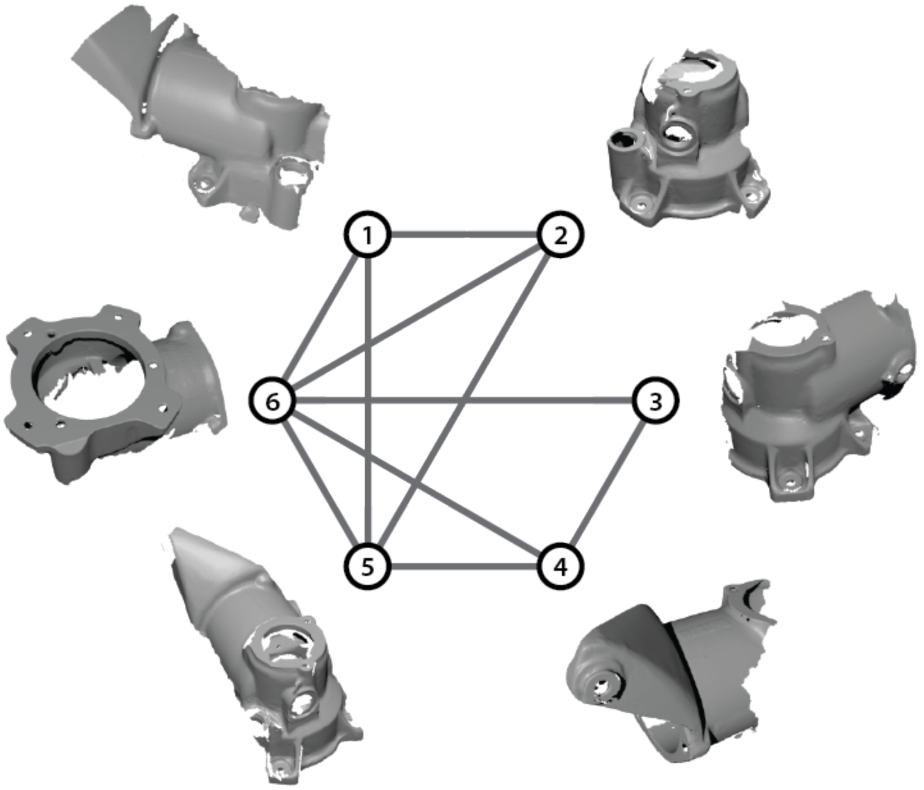
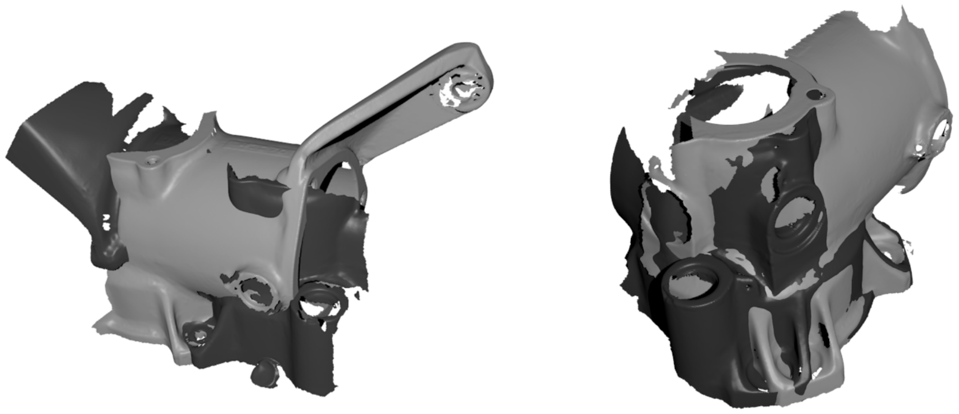


Figure 9.6: Registration connections that correctly aligned each gearbox acquisition.

Triplet transitivity is the best verification strategy for this dataset due to both a sufficient number of available point-clouds, and an adequate number achieving correct alignment allowing triplets to be formed. The transitivity correctly classified all 15 transforms using a threshold of one; Chapter 8 showed that a threshold of one is the most applicable when registering six point-clouds. The TCV based verification is able to classify this dataset if an appropriate threshold is selected. The mean and standard deviation of the TCV for the correct and incorrect alignments is 0.099±0.085 and 0.024±0.012, respectively. This distribution shows that a single threshold cannot guarantee perfect classification as correct and incorrect alignments can produce similar TCVs. Using a threshold of 0.05 ensures no false positive classifications, but it does result in four false negative classifications. These false negatives are due to the pipeline correctly aligning point-clouds with very low overlap, which stunts the size of the TCV; an example of this is the alignment between point-clouds one and six. While this classification is not perfect, there is still enough correctly classified transforms to bring all the point-clouds into global alignment.



(a) Point-clouds one and four. (b) Point-clouds two and three.

Figure 9.7: Two misalignments between helicopter gearbox point-clouds.

9.2.4 Gnome

The manner in which the gnome dataset was acquired did not facilitate registering all point-cloud pairs together. This is due to the SR4K imaging system being shifted in steps around the gnome, restricting registration to the two immediate neighbours of each acquisition. As such, only twelve point-cloud pairs were registered, of which nine were correctly aligned. Figure 9.8 presents an alignment example, which shows that adjacent acquisitions (those that are immediate neighbours) still have large sections with minimal overlap, especially on the back wall. The three misaligned point-cloud pairs were a result of this being a particularly difficult dataset to register due to the combination of acquisition poses, scene structure, and the method for calculating normals.

The scene structure resulted in each point-cloud producing a very similar distribution of normals across each acquisition. This is because the back wall and the table have the two most distinctive groupings of normals, and their positions with respect to each other maintain a relatively consistent relationship. The surfaces of the gnome itself produce a distribution of normals that remains the same from different poses, as it does not provide a distinctive representation of the surface. This is because each normal is calculated from across a relatively large patch of points, which is akin to smoothing the surface of the gnome. The normals were calculated in this fashion to improve their quality, otherwise they would be ineffective for calculating the rotation due to noise from both the lower quality and sparseness of the sample points provided by the SR4K. Despite these limitations, the majority of adjacent point-clouds achieved correct alignment.

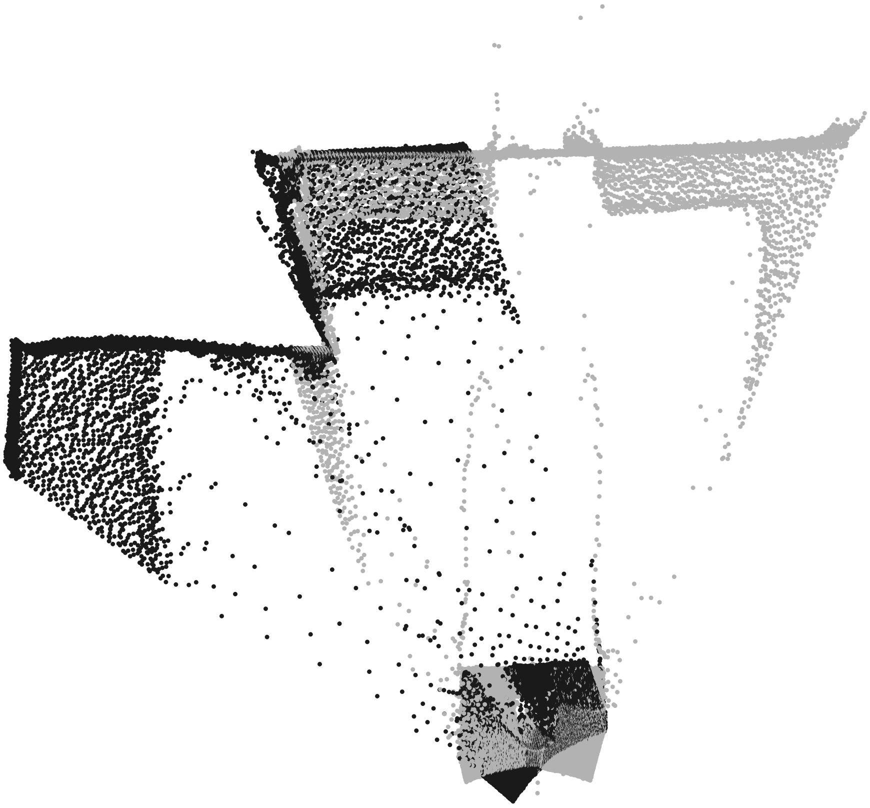


Figure 9.8: Top-down view of a correct alignment between two adjacent point-clouds from the gnome dataset. The sample points in what should be empty space are the mixed pixels.

Automatically verifying whether correct alignment has been achieved is also difficult with the gnome dataset. The triplet transitivity strategy was unable to be applied as no correct triplets could be formed due to the acquisition steps. If additional acquisitions from different poses were included, then triplet transitivity should improve as more point-clouds would link together. Using the TCV showed promise, with the mean and standard deviation of the correct alignments being 0.3 ± 0.063, compared to 0.17 ± 0.064 for the incorrect adjacent alignments. Using a threshold of 0.21 would adequately classify the alignments, though this would only be applicable to adjacent point-clouds, if all point-cloud pairs were registered this TCV threshold is unlikely to be suitable. In this situation it is expected that manual verification would be required.

9.2.5 Stairs

The stairs dataset proved to be a challenge for the pipeline as the majority of its surfaces are perpendicular to each other, which carried through to the

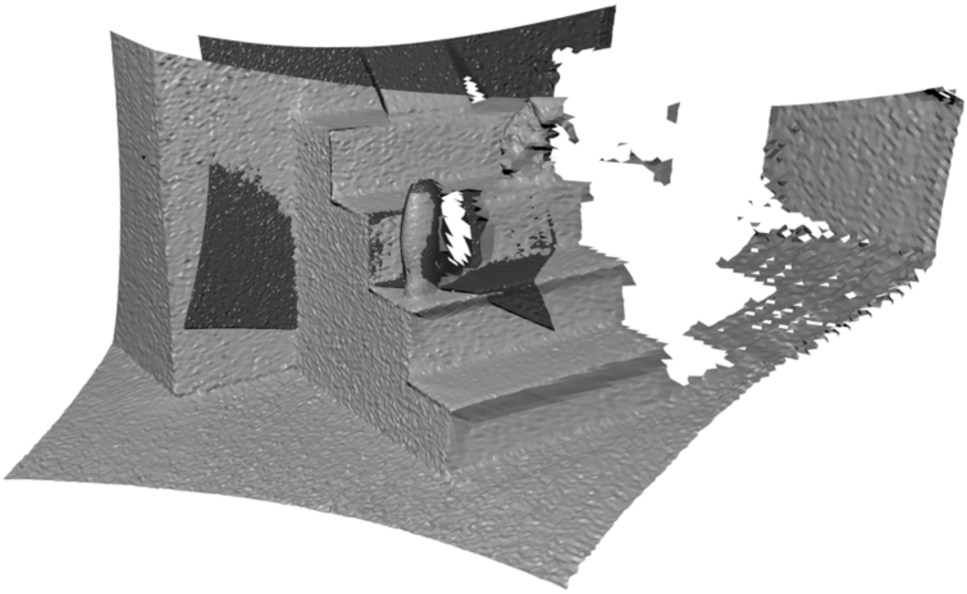


Figure 9.9: Rendering of two correctly aligned point-clouds from the stairs dataset. The resolution of the two point-clouds are the same, but their sampling density differs due to their size difference.

representation of the normals. The symmetry within the normals only allowed five of the fifteen point-cloud pairs to achieve correct alignment, with the incorrect rotations corresponding to this symmetry. While more correct alignments are preferred, those found brought five of the six point-clouds into global alignment with each other. The manner in which the normals were calculated resulted in two of the objects placed on the stairs only adding noise to the scene, instead of providing distinguishable surfaces. The pipeline was able to handle this noise along with the mixed pixels that occur within each acquisition. It was also confirmed that smaller point-clouds can be aligned with larger point-clouds as shown in Figure 9.9. It must be noted that the resolution of these two point-clouds are the same, therefore their sampling densities are different. This shows that the pipeline is able to register point-clouds acquired at different distances from the object, even though this distance changes their respective sampling density and size.

The low number of correct alignments restricted the performance of triplet transitivity, as only one triplet was produced. This triplet does however correctly classify three of the five correct alignments. The incorporation of alternative rotation selection, while computationally expensive, is expected to alleviate the issues of symmetry, thus improving the results. This would in turn increase the classification performance of triplet transitivity. Using TCV with a threshold of 0.1 correctly classified three of the five point-cloud pairs, the other two are incorrectly classified as their TCV is indistinguishable from the point-cloud pairs with incorrect alignments. The stairs dataset shows that symmetry within the normals is still an issue that can be detrimental to registration performance.

9.3 Summary

Concisely evaluating the performance of the coarse registration pipeline using real-world data helped to further characterise its strengths and weaknesses. The pipeline was able to correctly align all the point-clouds of both the angel and gearbox datasets, as well as correctly classifying each point-cloud pair alignment. This allowed point-clouds with minimal overlap to be correctly aligned, outperforming alternative registration algorithms. However, the pipeline is not infallible, as the other three datasets showed that rotational alignment needs further refinement. Its weakness was attributed to the quality and distribution of the normals. If the normals do not accurately reflect the true orientation of the point-cloud surfaces from which they are extracted, then the pipeline is susceptible to identifying an incorrect rotation. Additionally, if the representations of the normals contain a form of symmetry, then a rotation linked to this symmetry may be selected. Both of these issues have a higher priority than being able to handle low overlap, as their effect on registration is invariant to overlap.

The performed evaluation was also able to identify a limitation in the translation correlation with respect to the size of the bounding cube. Correcting this limitation allowed the translation correlation to select translational shifts of up to twice their previous size, which depending on the dataset, can be necessary. Triplet transitivity continued to perform exceptionally well when enough point-cloud pairs were correctly aligned; if this was not the case then it had a tendency of classifying all transforms as misaligned. The TCV based verification also achieved a reasonable level of accuracy, though it is more sensitive to changes in the structure of the two point-clouds it was classifying. The classification threshold was manually tuned to each dataset as a means of automatically calculating it has yet to be established; if an individual threshold could be produced to classify each alignment, then its true positive and negative classifications would increase.

Chapter 10

Conclusion

The task of accurately and efficiently bringing point-clouds into coarse alignment has been achieved by the coarse registration pipeline analysed in this thesis. The pipeline was constructed from distinct steps which, due to the mathematical techniques used, calculated the rotation and translation independently from each other, before combining them together as a single transform. The final step then automatically verified whether the found transforms successfully aligned the point-clouds.

The initial motivation for undertaking this research was to register sparse range images acquired by AMCW based imaging systems. The coarse registration pipeline proposed by Makadia *et al.* (2006) was identified as a promising solution, but it had minimal investigation since its conception. Because of this lack of investigation, the objective of this thesis became analysing and characterising the pipeline to determine its strengths and weaknesses, an examination that has not been performed previously. Based on this analysis, enhancements have been developed and incorporated to significantly improve the pipeline's performance.

The remainder of this conclusion is divided into three parts, a summary of the analysis performed at each step and the associated findings. This is followed by an overview of the contributions made to the registration field by the beneficial enhancements identified throughout the performed research. The thesis is then concluded by exploring the future directions in which this coarse registration pipeline can be taken to further improve its performance.

10.1 Summary of Analysis and Findings

The coarse registration pipeline was able to efficiently calculate the rotation and translation by independently correlating the normals and vertices in the frequency domain, respectively. This approach permitted point-clouds with any rotational and translational misalignment to be registered as long as they overlap. The introduced enhancements correctly aligned point-clouds with overlap as low as 5%, which is a significant improvement in comparison to the 45% attained by the initial incarnation of the pipeline proposed by Makadia *et al.* (2006). These low levels of overlap are achieved by the enhanced pipeline as it is able to provide a superior representation of the normals, thus making the correlation more effective. The coarse registration pipeline's capabilities can be further refined by performing additional analysis, especially with respect to using a variety of real data from different sources.

The primary weakness of the pipeline is calculating the rotational alignment, which is affected by the normals. Binning the normals on the unit sphere is unable to include the position from which they were extracted, therefore, normals from disparate surfaces on a point-cloud can be collected by the same bin. However, this is a minor issue, as the main problem is producing normals that adequately represent surface orientation. The quality of this representation is affected if sampling accuracy is reduced due to noise, or if the sampling density is sparse. Given that these issues occur in range images acquired by AMCW based imaging systems, their registration by this pipeline can be challenging. While the pipeline does have moderate success registering AMCW range images, it is recommended that they first be filtered and preprocessed to improve their sample point accuracy, and thus the quality of their normals. The flow on effect is that their overall registration accuracy will be improved.

In this thesis, each step of the coarse registration pipeline was analysed using ground-truth data to provide exact measures of overlap, rotational misalignment and translational misalignment. This allowed the ability of each step to be quantified as well as permitting a means of measuring the benefit provided by potential enhancements. The remainder of this section summarises the performed analysis and the enhancements applied at each step of the pipeline and details the key findings that were made.

10.1.1 Binning Normals

The manner in which sets of normals are discretely represented in preparation for spherical-harmonic conversion and correlation affects the calculated rotation, both in accuracy and in efficiency. Therefore, it was important to identify and use an appropriate binning scheme. By investigating three binning schemes (equiangle, icosahedron and Fibonacci) the effects of bin distribution on the sphere were determined. Intuitively it would be expected that the icosahedron and Fibonacci schemes should be preferred due to their near uniform bin distribution. However, in practice they are unsuitable as they do not have a natural affinity for spherical-harmonic conversion, making their use impracticably expensive. With respect to their accuracy, they are only marginally better than equiangle binning at a bandwidth of 64.

Equiangle binning was found to be the optimal scheme as its bin distribution both efficiently collects normals and has a natural affinity for spherical-harmonic conversion. The non-uniform distribution of its bins were mitigated by applying correct normalisation, a step that is not stated by Makadia *et al.* (2006) as being included. The analysis also found that orientating the normals to the pole of the sphere, instead of the equator, improved registration performance. This is because equiangle binning has an increased bin density at the pole, allowing the normals to be sampled more finely. The predominant benefit of equiangle binning is its efficient conversion to spherical harmonics at higher bandwidths. Increasing the bandwidth improves the representation of normals and reduces rotation error as there are more selectable rotations. This improvement is exemplified by the 50% increase in correct rotational alignments when doubling the bandwidth from 64 to 128.

10.1.2 Weighting Normals

Weighting the normals of partially overlapping point-clouds is an essential step for improving the reliability of the registration, as simply binning the normals is unlikely to provide a faithful representation of the surfaces. This is because the representation varies with sample distribution, noise and disconnected surfaces contributing normals to the same bins. A range of different weighting schemes and parameters were used to help mitigate these detriments, with these schemes falling into two categories. The first reweighted prominent bins (bins that contained a sufficient number of normals), while the second weighted and culled normals based on their surrounding surface curvature.

The representation of normals and subsequently the rotational performance were improved by both weighting scheme categories.

The best representation of the normals was found by introducing the idea of encoding both the reweighting of prominent bins and the surface curvature value into a complex value. This was possible as the spherical-harmonic correlation inherently handles complex values. By maintaining the mean surface curvature weighting of the normals kept by a bin, each bin received an additional distinguishing aspect, instead of simply indicating that it contains normals. Representing the equiangle bins using these complex values increased registration performance by over 100% in comparison to the representation proposed by Makadia *et al.* (2006). As such, this novel combination significantly enhanced registration performance, making this coarse registration pipeline have state-of-the-art capabilities.

10.1.3 Translational Alignment

The translational alignment of point-clouds had exceptional performance, as it aligns any point-cloud pair if their overlap is greater than 5% and they have correct rotational alignment. This ability is linked to point-clouds only depicting the surfaces of a scene, resulting in many voxels remaining empty. It is this aspect that allows the correlation to easily identify the optimal coarse translation. The translational alignment does have the weakness in that its accuracy is fundamentally linked to the rotational alignment accuracy; the greater the rotation accuracy, the greater the translation accuracy. It is for this reason that if the translational alignment fails, the rotational alignment should be investigated. In the unlikely event that translational alignment fails when the rotation is correct, it can be presumed that a repetitive structure exists within the point-clouds. While not explored as part of this research, this step could incorporate a weighting scheme similar to that of the normals, helping to mitigate this situation.

10.1.4 Verification of Correct Alignment

Visually checking to confirm successful alignment is undesirable as this step of the pipeline is preferably automated. Reducing the required manual intervention was achieved by checking that the aligned point-clouds met predefined criteria stipulating what defines correct alignment. Three verification strategies (surface orientation, TCV thresholding, and transform transitivity), each specifying a different criterion, were evaluated to determine their classification reliability. Because each of their criterion are different, it is possible to use all three strategies in conjunction with each other. Makadia *et al.* (2006) used verification for identifying if an alternative rotation should be selected, but the conducted research found that alternative rotation selection provided minimal improvement considering its computational expense. The benefit of verification is that it identifies which transforms should be kept or discarded when bringing all point-clouds into the same coordinate system.

The surface orientation and the TCV thresholding strategies had similar classification performances. The advantage of the TCV is that it is computationally free, as it is a by-product of the translation correlation. It does have the limitation of being linked to overlap, as overlap decreases the TCV corresponding to the correct alignment will also decrease. Identifying a means of automatically adjusting the threshold will mitigate this issue. Transform transitivity produced the best results, with it being capable of correctly classifying all point-cloud pairs. This ability is dependent upon there being three or more point-clouds and the percentage of correct point-cloud alignments; correct alignments are necessary for forming consistent triplets. If the classification performance of transform transitivity is inadequate, then it is better to focus on repairing limitations further up the pipeline. The verification step completes the coarse registration pipeline, with any necessary refining of alignment being performed by a fine registration algorithm.

10.2 Contributions to Knowledge

The major contributions to knowledge produced throughout this dissertation are now listed in a concise manner:

- A comprehensive analysis which quantified registration performance at each step of the pipeline.
- Evaluated an extensive number of modifications and parameters to these steps to ascertain which provide the most benefit.
- A new algorithm was developed for assigning points on a sphere to the closest Fibonacci bin.
- Equiangle binning was found to only have a marginal decrease in accuracy when compared against more uniform bin distributions, but its bin-

ning efficiency and affinity for conversion to spherical harmonics makes it the superior choice.

- Established that the orientation of normals on the sphere can impact registration; orientating the normals to the pole of the spherical coordinate system improves registration performance as bin density is greater.
- Applied the plane distance scheme for weighting normals to the spherical-harmonic correlation, showing it to be a beneficial addition.
- Introduced the novel combination of using two weighting schemes together by encoding them as complex values, providing a significant increase in registration performance.
- Showed that the quality of the translation correlation is fundamentally dependent upon the rotation accuracy, if rotation is correct, the translation is rarely incorrect.
- Introduced the alignment verification strategy based on thresholding the translation correlation value.
- Introduced the transform transitivity verification strategy and analytically identified its limitations; this strategy was the most accurate if particular criteria were met.
- Revealed that alternative rotation selection does not provide much advantage due to its high computational cost and minimal gain.
- Showed that the generic nature of the coarse registration pipeline makes it applicable to the majority of 3-D point-clouds, while being able to produce state-of-the-art results.

10.3 Future Directions

Improving the coarse registration pipeline with new and novel enhancements will further increase its robustness for consistently aligning point-clouds which have minimal overlap and increasing levels of noise. In order to achieve this, the focus of any future work should be directed towards the two steps used to calculate the rotational alignment. This is because selecting the correct rotation is both the most important and difficult step, as the subsequent steps in the pipeline rely upon the accuracy of the rotation. With respect to this, the future direction that should be predominately explored is the binning and weighting of normals to provide a more faithful representation of the point-cloud surfaces on the unit sphere. However, solely investigating alternative binning schemes is unlikely to offer any substantial improvements to registration ability, as equiangle binning has been shown capable of performing at a level similar to that provided by more uniform bin distributions. If a new binning scheme is deemed necessary, it should be accompanied by appropriate weighting schemes. However, the recommendation is to continue using equiangle binning and to concentrate on improving the weighting schemes.

The weighting schemes currently used are limited by their static thresholds, which do not readily cater to the variations in scene structure or sampling quality. The incorporation of a measure which examines either the global surface structure of the two point-clouds or the distribution of normals on the unit sphere will help provide a means of dynamically adjusting the thresholds used with the weighting schemes. The advantage of dynamic thresholds is that they would inherently handle a greater variety of scenes and imaging systems, therefore intrinsically increasing the likelihood of correct alignment. Enhancing or replacing these weighting schemes is also possible, as alternative schemes applicable to spherical-harmonic correlation could be developed. This is especially true if two weighting schemes which complement each other are combined via the proposed complex values weighting scheme. An example of a future enhancement that should be suitable for the complex values weighting is to incorporate the sign into the measure of surface curvature about each vertex. Surfaces that are concave would receive a negative value, planar surface would still be zero, while convex surfaces would be positive. Because different types of surface structure are distinguished, an improved representation of a point-cloud will likely be attained.

The mathematics underlying the spherical-harmonic correlation fundamentally determines the achieved registration performance. Increasing the transform bandwidth improves the representation of normals in the frequency domain, which combined with a high correlation bandwidth enhances registration accuracy. It is however unknown whether there is an upper limit to the benefit provided by further increasing these two bandwidths. It is presumed that there is an upper limit, but identifying it was restricted by the available computational resources. Further analysis in this area will determine if a more efficient algorithm or more computational resources sufficiently benefits registration. There is also the possibility of expanding the base mathematics to further extend the idea of using multiple pieces of information, as established by the complex values weighting. This would allow additional measures of the surface structure to be incorporated, allowing the characteristics of the point-clouds to be better distinguished and correctly matched by the correlation.

Currently, both the rotation correlation and translation correlation are coarse registration algorithms, but there is the potential of converting them into fine registration algorithms. Sub-pixel registration has been shown feasible in the frequency domain (Balci and Foroosh, 2006; Tzimiropoulos *et al.*, 2011) for correlating 2-D images. There is the possibility of using similar techniques for either spherical-harmonic correlation or 3-D correlation. At present translation correlation is exceptionally robust; it is unlikely that any enhancements, other than sub-voxel registration, will produce additional benefit.

An area that was not investigated, but is applicable to the verification step is the incorporation of constraints that restrict which rotations or translations can be selected. These constraints would be based on prior knowledge of the image acquisition process, such as when acquisition order is known or when the maximum rotational difference between two poses must be less than a particular angle. There are any number of usable verification strategies that could be designed for a particular scene or acquisition scenario. Enhancing and expanding the available strategies would further guarantee that the calculated transforms are the ones desired for achieving correct alignment.

The future directions suggested here are by no means the only ones worth exploring, as any additional analysis will help refine the characterisation of the coarse registration pipeline. The incorporation of additional enhancements will further expand its state-of-the-art capabilities, allowing it to draw ever closer to the goal of truly automated 3-D registration.

References

Althloothi, S., M. H. Mahoor, and R. M. Voyles. A robust method for rotation estimation using spherical harmonics representation. *Image Processing, IEEE Transactions on*, **22(6)**, pp. 2306–2316 (2013).

Arfken, G. B., H. J. Weber, and F. E. Harris. *Mathematical methods for physicists: a comprehensive guide.* Academic Press, Waltham, MA, 7th edition (2013).

Arius 3D. Arius 3D color scanners. Retrieved April 9, 2014 from `http://www.ariustechnology.com/color-scanners` (2014).

Balci, M. and H. Foroosh. Subpixel estimation of shifts directly in the Fourier domain. *Image Processing, IEEE Transactions on*, **15(7)**, pp. 1965–1972 (2006).

Becerik-Gerber, B., F. Jazizadeh, G. Kavulya, and G. Calis. Assessment of target types and layouts in 3D laser scanning for registration accuracy. *Automation in Construction*, **20(5)**, pp. 649–658 (2011).

Besl, P. J. and N. D. McKay. A method for registration of 3-D shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **14(2)**, pp. 239–256 (1992).

Bi, Z. and L. Wang. Advances in 3D data acquisition and processing for industrial applications. *Robotics and Computer-Integrated Manufacturing*, **26(5)**, pp. 403–413 (2010).

Blais, F. Review of 20 years of range sensor development. *Journal of Electronic Imaging*, **13(1)**, pp. 231–243 (2004).

Blais, F. and J.-A. Beraldin. Recent developments in 3D multi-modal laser imaging applied to cultural heritage. *Machine Vision and Applications*, **17(6)**, pp. 395–409 (2006).

Blais, J. A. R. and M. A. Soofi. Spherical harmonic transforms using quadratures and least squares. In: *Sixth International Conference on Computational Science (ICCS) - Volume Part III*, pp. 48–55 (2006).

Blumenthal, S., E. Prassler, J. Fischer, and W. Nowak. Towards identification of best practice algorithms in 3D perception and modeling. In: *Robotics and Automation (ICRA), IEEE International Conference on*, pp. 3554–3561 (2011).

Bonarrigo, F., A. Signoroni, and R. Leonardi. A robust pipeline for rapid feature-based pre-alignment of dense range scans. In: *Computer Vision (ICCV), IEEE 13th International Conference on*, pp. 2260–2267 (2011).

Bornaz, L., A. Lingua, and F. Rinaudo. Engineering and environmental applications of laser scanner techniques. In: *Photogrammetric Computer Vision (PCV)*, volume 34 of *International Society for Photogrammetry and Remote Sensing (ISPRS)*, pp. 40–43 (2002).

Brechbühler, C., G. Gerig, and O. Kübler. Parametrization of closed surfaces for 3-D shape description. *Computer Vision and Image Understanding*, **61(2)**, pp. 154–170 (1995).

Brusco, N., M. Andreetto, A. Giorgi, and G. M. Cortelazzo. 3D registration by textured spin-images. In: *3-D Digital Imaging and Modeling (3DIM), Fifth International Conference on*, pp. 262–269 (2005).

Buchholz, D., D. Kubus, S. Winkelbach, and F. M. Wahl. 3D object localization using single camera images. In: *Pattern Recognition (ICPR), 21st International Conference on*, pp. 821–824 (2012).

Bülow, H. and A. Birk. Spectral 6DOF registration of noisy 3D range data with partial overlap. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **35(4)**, pp. 954–969 (2013).

Burel, G. and H. Henocq. Determination of the orientation of 3D objects using spherical harmonics. *Graphical Models and Image Processing*, **57(5)**, pp. 400–408 (1995).

Castellani, U. and A. Bartoli. 3D shape registration. In: *3D Imaging, Analysis and Applications*, pp. 221–264. Springer London (2012).

Chen, C.-S., Y.-P. Hung, and J.-B. Cheng. RANSAC-based DARCES: a new approach to fast automatic registration of partially overlapping range images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **21(11)**, pp. 1229–1234 (1999).

Chen, Y. and G. Medioni. Object modeling by registration of multiple range images. In: *Robotics and Automation (ICRA), IEEE International Conference on*, volume 3, pp. 2724–2729 (1991).

Choi, S., T. Kim, and W. Yu. Performance evaluation of RANSAC family. In: *The British Machine Vision Conference (BMVC)*, pp. 81.1–81.12 (2009).

Corsini, M., M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello, and R. Scopigno. Fully automatic registration of image sets on approximate geometry. *International Journal of Computer Vision*, **102(1-3)**, pp. 91–111 (2013).

Creaform. Technical Specifications: Handyscan 3D Handheld 3D Scanner. Retrieved April 9, 2014 from `http://www.creaform3d.com/en/metrology-solutions/products/portable-3d-scanner/technical-specifications-handyscan-3d` (2014).

De Falco, I., A. D. Cioppa, D. Maisto, U. Scafuri, and E. Tarantino. Adding chaos to differential evolution for range image registration. In: *Applications of Evolutionary Computation*, volume 7835 of *Lecture Notes in Computer Science*, pp. 344–353. Springer Berlin Heidelberg (2013).

Díez, Y., F. Roure, X. Lladó, and J. Salvi. A qualitative review on 3D coarse registration methods. *ACM Computer Surveys (CSUR)*, **47(3)**, pp. 45:1–45:36 (2015).

Dinh, H. Q. and S. Kropac. Multi-resolution spin-images. In: *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, volume 1, pp. 863–870 (2006).

Dorai, C., G. Wang, A. K. Jain, and C. Mercer. Registration and integration of multiple object views for 3D model construction. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **20(1)**, pp. 83–89 (1998).

Dorrington, A. A., M. J. Cree, A. D. Payne, R. M. Conroy, and D. A. Carnegie. Achieving sub-millimetre precision with a solid-state full-field heterodyning range imaging camera. *Measurement Science and Technology*, **18(9)**, pp. 2809–2816 (2007).

Driscoll, J. R. and D. M. Healy, Jr. Computing Fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics*, **15(2)**, pp. 202–250 (1994).

Dutagaci, H., C. P. Cheung, and A. Godil. Evaluation of 3D interest point detection techniques via human-generated ground truth. *The Visual Computer*, **28(9)**, pp. 901–917 (2012).

Eiben, A. E. and J. E. Smith. *Introduction to Evolutionary Computing.* Natural Computing Series. Springer, first edition (2003). Corrected Reprint, 2007.

Ekekrantz, J., A. Pronobis, J. Folkesson, and P. Jensfelt. Adaptive iterative closest keypoint. In: *Mobile Robots (ECMR), European Conference on*, pp. 80–87 (2013).

Fantoni, S., U. Castellani, and A. Fusiello. Accurate and automatic alignment of range surfaces. In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), Second International Conference on*, pp. 73–80 (2012).

Fischler, M. A. 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)**, pp. 381–395 (1981).

Fletcher, R. *Practical methods of optimization.* John Wiley & Sons, Inc., second edition (1987).

Foix, S., G. Alenya, and C. Torras. Lock-in time-of-flight (ToF) cameras: A survey. *IEEE Sensors Journal*, **11(9)**, pp. 1917–1926 (2011).

Fortenbury, B. R. and G. Guerra-Filho. Robust 2D/3D calibration using RANSAC registration. In: *Advances in Visual Computing*, volume 7431 of *Lecture Notes in Computer Science*, pp. 179–188. Springer Berlin Heidelberg (2012).

Franaszek, M., G. S. Cheok, and C. Witzgall. Fast automatic registration of range images from 3D imaging systems using sphere targets. *Automation in Construction*, **18(3)**, pp. 265–274 (2009).

Gerig, G., M. Styner, D. Jones, D. Weinberger, and J. Lieberman. Shape analysis of brain ventricles using SPHARM. In: *Mathematical Methods in Biomedical Image Analysis (MMBIA), IEEE Workshop on*, pp. 171–178 (2001).

Godbaz, J. P. *Ameliorating systematic errors in full-field AMCW lidar*. Ph.D. thesis, The University of Waikato (2012).

GOM. ATOS ScanBox - optical 3D measuring machine. Retrieved April 9, 2014 from `http://www.gom.com/metrology-systems/system-overview/atos-scanbox.html` (2014).

González, A. Measurement of areas on a sphere using Fibonacci and latitude–longitude lattices. *Mathematical Geosciences*, **42**, pp. 49–64 (2010).

Gorski, K. M., E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. Healpix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere. *Astrophysical Journal*, **622(2)**, pp. 759–771 (2005).

Gu, X., Y. Wang, T. F. Chan, P. M. Thompson, and S.-T. Yau. Genus zero surface conformal mapping and its application to brain surface mapping. *Medical Imaging, IEEE Transactions on*, **23(8)**, pp. 949–958 (2004).

Guarnieri, A., A. Vettore, M. Camarda, and C. Domenica. Automatic registration of large range datasets with spin-images. *Journal of Cultural Heritage*, **12(4)**, pp. 476–484 (2011).

Gutman, B., Y. Wang, T. Chan, P. M. Thompson, and A. W. Toga. Shape registration with spherical cross correlation. In: *Mathematical Foundations in Computational Anatomy (MFCA), Second MICCAI Workshop on*, pp. 56–67 (2008).

Gutman, B. A. *Shape analysis methods for 3D brain and skull imaging*. Ph.D. thesis, University of California, Los Angeles (UCLA) (2013).

Halma, A., F. ter Haar, E. Bovenkamp, P. Eendebak, and A. van Eekeren. Single spin image-ICP matching for efficient 3D object recognition. In: *Proceedings of the ACM Workshop on 3D Object Retrieval (3DOR)*, pp. 21–26 (2010).

Healy, D. M., Jr., D. N. Rockmore, P. J. Kostelec, and S. Moore. FFTs for the 2-sphere - improvements and variations. *Journal of Fourier Analysis and Applications*, **9(4)**, pp. 341–385 (2003).

Herbort, S. and C. Wöhler. An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. *3D Research*, **2(3)** (2011).

Hexagon Metrology. Romer Absolute Arm product brochure (October 15, 2012). Retrieved April 9, 2014 from `http://apps.hexagon.se/downloads123/hxrom/romer/general/brochures/ROMER%20Absolute%20Arm_brochure_en.pdf` (2014).

Ho, H. T. and D. Gibbins. Multi-scale feature extraction for 3D surface registration using local shape variation. In: *Image and Vision Computing New Zealand (IVCNZ), 23rd International Conference* (2008).

Hołowko, E. and R. Sitnik. View integration technique using multi-resolution point clouds in 3D modeling. In: *Videometrics, Range Imaging, and Applications XII; and Automated Visual Inspection*, volume 8791 of *Proceedings of the SPIE* (2013).

Horn, B. K. P. Extended Gaussian images. *Proceedings of the IEEE*, **72(12)**, pp. 1671–1686 (1984).

Horn, B. K. P. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America A*, **4(4)**, pp. 629–642 (1987).

Houshiar, H., J. Elseberg, D. Borrmann, and A. Nüchter. A study of projections for key point based registration of panoramic terrestrial 3D laser scans. *Geo-spatial Information Science*, **In Press** (2013). Accepted for Publication on December 21, 2012.

Hrgetić, V. and T. Pribanić. Surface registration using genetic algorithm in reduced search space. arXiv:1310.0302 [cs.CV] (2013). Computer Vision and Pattern Recognition.

Huang, H., L. Shen, R. Zhang, F. Makedon, B. Hettleman, and J. Pearlman. Surface alignment of 3D spherical harmonic models: application to cardiac MRI analysis. In: *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 3749 of *Lecture Notes in Computer Science*, pp. 67–74. Springer Berlin Heidelberg (2005).

Huber, D. F. and M. Hebert. Fully automatic registration of multiple 3D data sets. *Image and Vision Computing*, **21(7)**, pp. 637–650 (2003).

Isenburg, M., Y. Liu, J. Shewchuk, and J. Snoeyink. Streaming computation of Delaunay triangulations. *ACM Transactions on Graphics*, **25(3)**, pp. 1049–1056 (2006).

Jacq, J.-J. and C. Roux. Registration of 3-D images by genetic optimization. *Pattern Recognition Letters*, **16(8)**, pp. 823–841 (1995).

Jiang, J., J. Cheng, and X. Chen. Registration for 3-D point cloud using angular-invariant feature. *Neurocomputing*, **72(16-18)**, pp. 3839–3844 (2009).

Jin, S., R. R. Lewis, and D. West. A comparison of algorithms for vertex normal computation. *The Visual Computer*, **21(1-2)**, pp. 71–82 (2005).

Johnson, A. E. *Spin-images: a representation for 3-D surface matching*. Doctor of philosophy in the field of robotics, Carnegie Mellon University, Pittsburgh, Pennsylvania (1997).

Johnson, A. E. and M. Hebert. Using spin images for efficient object recognition in cluttered 3D scenes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **21(5)**, pp. 433–449 (1999).

Jolliffe, I. *Principal Component Analysis*. Springer, second edition (2002).

Kabsch, W. A solution for the best rotation to relate two sets of vectors. *Acta Crystallographica Section A*, **32(5)**, pp. 922–923 (1976).

Kakarala, R., P. Kaliamoorthi, and V. Premachandran. Three-dimensional bilateral symmetry plane estimation in the phase domain. In: *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, pp. 249–256 (2013).

Kim, C., J. Lee, M. Cho, and C. Kim. Fully automated registration of 3D CAD model with point cloud from construction site. In: *Automation and Robotics in Construction, 28th International Symposium on*, pp. 917–922 (2011).

King, B. J., T. Malisiewicz, C. V. Stewart, and R. J. Radke. Registration of multiple range scans as a location recognition problem: Hypothesis generation, refinement and verification. In: *3-D Digital Imaging and Modeling (3DIM), Fifth International Conference on*, pp. 180–187 (2005).

Klasing, K., D. Althoff, D. Wollherr, and M. Buss. Comparison of surface normal estimation methods for range sensing applications. In: *Robotics and Automation (ICRA), IEEE International Conference on*, pp. 3206–3211 (2009).

Koay, C. G. Analytically exact spiral scheme for generating uniformly distributed points on the unit sphere. *Journal of Computational Science*, **2(1)**, pp. 88–91 (2011a).

Koay, C. G. A simple scheme for generating nearly uniform distribution of antipodally symmetric points on the unit sphere. *Journal of Computational Science*, **2(4)**, pp. 377–381 (2011b).

Kolb, A., E. Barth, R. Koch, and R. Larsen. Time-of-flight cameras in computer graphics. *Computer Graphics Forum*, **29(1)**, pp. 141–159 (2010).

Kondo, N. Automation on fruit and vegetable grading system and food traceability. *Trends in Food Science & Technology*, **21(3)**, pp. 145–152 (2010).

Kostelec, P. J. and D. N. Rockmore. FFTs on the rotation group. *Journal of Fourier Analysis and Applications*, **14**, pp. 145–179 (2008).

Kulkarni, N. A. and S. Kumar. Vote based correspondence for 3D point-set registration. In: *Computer Vision, Graphics and Image Processing, Eighth Indian Conference on*, pp. 22:1–22:8 (2012).

Larkins, R. L., M. J. Cree, and A. A. Dorrington. Analysis of ICP variants for the registration of partially overlapping time-of-flight range images. In: *Image and Vision Computing New Zealand (IVCNZ), 25th International Conference* (2010).

Larkins, R. L., M. J. Cree, and A. A. Dorrington. Analysis of binning of normals for spherical harmonic cross-correlation. In: *Three-Dimensional Image Processing (3DIP) and Applications II*, volume 8290 of *Proceedings of SPIE* (2012).

Larkins, R. L., M. J. Cree, and A. A. Dorrington. Analysis of weighting of normals for spherical harmonic cross-correlation. In: *Three-Dimensional Image Processing (3DIP) and Applications*, volume 8650 of *Proceedings of SPIE* (2013).

Lee, S. Time-of-flight depth camera motion blur detection and deblurring. *Signal Processing Letters, IEEE*, **21(6)**, pp. 663–666 (2014).

Levin, A., R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM Transactions on Graphics*, **26(3)**, p. Article No.70 (2007).

Liu, Y., R. R. Martin, L. de Dominicis, and B. Li. Using retinex for point selection in 3D shape registration. *Pattern Recognition*, **47(6)**, pp. 2126–2142 (2014).

Liu, Y.-S. and K. Ramani. Robust principal axes determination for point-based shapes using least median of squares. *Computer-Aided Design*, **41(4)**, pp. 293–305 (2009).

Mahmood, N., C. Omar, and T. Tjahjadi. Multiviews reconstruction for prosthetic design. *Information Technology, The International Arab Journal of*, **9(1)**, pp. 49–55 (2012).

Makadia, A. and K. Daniilidis. Spherical correlation of visual representations for 3D model retrieval. *International Journal of Computer Vision*, **89(2-3)**, pp. 193–210 (2010).

Makadia, A., A. Patterson IV, and K. Daniilidis. Fully automatic registration of 3D point clouds. In: *Computer Vision and Pattern Recognition (CVPR), IEEE Conference on*, volume 1, pp. 1297–1304 (2006).

Martínez-Carranza, J., A. Calway, and W. Mayol-Cuevas. Enhancing 6D visual relocalisation with depth cameras. In: *Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on*, pp. 899–906 (2013).

Mesa Imaging. Mesa SwissRanger SR4500 datasheet Rev 1.0 (July 12, 2013). Retrieved March 24, 2014 from `http://www.mesa-imaging.ch/dlm.php?fname=pdf/SR4500_DataSheet.pdf` (2014).

Minguez, J., L. Montesano, and F. Lamiraux. Metric-based iterative closest point scan matching for sensor displacement estimation. *Robotics, IEEE Transactions on*, **22(5)**, pp. 1047–1054 (2006).

Moreda, G. P., J. O.-C. navate, F. J. García-Ramos, and M. Ruiz-Altisent. Non-destructive technologies for fruit and vegetable size determination - a review. *Journal of Food Engineering*, **92(2)**, pp. 119–136 (2009).

Morell, V., M. Cazorla, D. Viejo, S. Orts, and J. Garcia. A study of registration techniques for 6DoF SLAM. In: *Artificial Intelligence Research and Development: Proceedings of the 15th International Conference of the Catalan Association for Artificial Intelligence*, edited by D. Riaño, E. Onaindia, and M. Cazorla, Frontiers in Artificial Intelligence and Applications, pp. 111–120. IOS Press (2012).

Newman, T. S. and H. Yi. A survey of the marching cubes algorithm. *Computers & Graphics*, **30(5)**, pp. 854–879 (2006).

Oliensis, J. A critique of structure-from-motion algorithms. *Computer Vision and Image Understanding*, **80(2)**, pp. 172–214 (2000).

Payne, A., A. Daniel, A. Mehta, B. Thompson, C. S. Bamji, D. Snow, H. Oshima, L. Prather, M. Fenton, L. Kordus, P. OConnor, R. McCauley, S. Nayak, S. Acharya, S. Mehta, T. Elkhatib, T. Meyer, T. ODwyer, T. Perry, V.-H. Chan, V. Wong, V. Mogallapu, W. Qian, and Z. Xu. A 512×424 CMOS 3D time-of-flight image sensor with multi-frequency photo-demodulation up to 130MHz and 2GS/s ADC. In: *International Solid-State Circuits Conference (ISSCC)*, Image Sensors (2014).

Payne, A. D. *Development of a full-field time-of-flight range imaging system.* Ph.D. thesis, University of Waikato (2008).

Penman, D. W., R. J. Valkenburg, and N. S. Alwesh. Robust calibration of the position of reference targets for a six degrees of freedom pose sensor. In: *Image and Vision Computing New Zealand (IVCNZ), 23rd International Conference*, pp. 1–6 (2008).

Perwaß, C. and L. Wietzke. Single lens 3D-camera with extended depth-of-field. In: *Human Vision and Electronic Imaging XVII*, volume 8291 of *Proceedings of SPIE* (2012).

Pham, M.-T., O. J. Woodford, F. Perbet, A. Maki, R. Gherardi, B. Stenger, and R. Cipolla. Scale-invariant vote-based 3D recognition and registration from point clouds. In: *Machine Learning for Computer Vision*, volume 411 of *Studies in Computational Intelligence*, pp. 137–162. Springer Berlin Heidelberg (2013).

Pomerleau, F., F. Colas, R. Siegwart, and S. Magnenat. Comparing ICP variants on real-world data sets. *Autonomous Robots*, **34(3)**, pp. 133–148 (2013).

Pooja, A. and V. M. Govindu. A multi-view extension of the ICP algorithm. In: *Computer Vision, Graphics and Image Processing (ICVGIP), Seventh Indian Conference on*, pp. 235–242 (2010).

Pribanic, T., Y. Diez, S. Fernandez, and J. Salvi. An efficient method for surface registration. In: *Computer Vision Theory and Applications (VISAPP), International Conference on*, volume 1, pp. 500–503 (2013).

Price, M., J. Green, and J. Dickens. Point-cloud registration using 3D shape contexts. In: *Robotics and Mechatronics Conference of South Africa (ROBOMECH)*, pp. 1–5 (2012).

Reddy, B. S. and B. N. Chatterji. An FFT-based technique for translation, rotation, and scale-invariant image registration. *Image Processing, IEEE Transactions on*, **5(8)**, pp. 1266–1271 (1996).

Ritchie, D. W. and G. J. L. Kemp. Fast computation, rotation, and comparison of low resolution spherical harmonic molecular surfaces. *Journal of Computational Chemistry*, **20(4)**, pp. 383–395 (1999).

Rusinkiewicz, S. and M. Levoy. Efficient variants of the ICP algorithm. In: *3-D Digital Imaging and Modeling (3DIM), Third International Conference on*, pp. 145–152 (2001).

Saff, E. B. and A. B. J. Kuijlaars. Distributing many points on a sphere. *The Mathematical Intelligencer*, **19(1)**, pp. 5–11 (1997).

Salvi, J., C. Matabosch, D. Fofi, and J. Forest. A review of recent range image registration methods with accuracy evaluation. *Image and Vision Computing*, **25(5)**, pp. 578–596 (2007).

Santamaría, J., O. Cordón, and S. Damas. A comparative study of state-of-the-art evolutionary image registration methods for 3D modeling. *Computer Vision and Image Understanding*, **115(9)**, pp. 1340–1354 (2011).

Santamaría, J., S. Damas, O. Cordón, and A. Escámez. Self-adaptive evolution toward new parameter free image registration methods. *Evolutionary Computation, IEEE Transactions on*, **17(4)**, pp. 545–557 (2013).

Segundo, M. P., L. Gomes, O. R. P. Bellon, and L. Silva. Automating 3D reconstruction pipeline by surf-based alignment. In: *Image Processing (ICIP), International Conference on*, pp. 1761–1764 (2012).

Seo, J. K., G. C. Sharp, and S. W. Lee. Range data registration using photometric features. In: *Computer Vision and Pattern Recognition (CVPR). IEEE Conference on*, volume 2, pp. 1140–1145 (2005).

Shen, L. and M. K. Chung. Large-scale modeling of parametric surfaces using spherical harmonics. In: *3D Data Processing, Visualization, and Transmission (3DPVT), Third International Symposium on*, pp. 294–301 (2007).

Shen, L., H. Huang, F. Makedon, and A. J. Saykin. Efficient registration of 3D SPHARM surfaces. In: *Computer and Robot Vision (CRV), Fourth Canadian Conference on*, pp. 81–88 (2007).

Shen, L. and F. Makedon. Spherical mapping for processing of 3D closed surfaces. *Image and Vision Computing*, **24(7)**, pp. 743–761 (2006).

SoftKinetic. DS325 datasheet V4.0 (May 27, 2013). Retrieved March 24, 2014 from `http://www.softkinetic.com/Portals/0/Documents/PDF/WEB_20130527_SK_DS325_Datasheet_V4.0.pdf` (2014).

Steder, B., R. B. Rusu, K. Konolige, and W. Burgard. Point feature extraction on 3D range scans taking into account object boundaries. In: *Robotics and Automation (ICRA), IEEE International Conference on*, pp. 2601–2608 (2011).

Sun, H., Z. Zhao, X. Jin, L. Niu, and L. Zhang. Depth from defocus and blur for single image. In: *Visual Communications and Image Processing (VCIP)*, pp. 1–5 (2013).

Taguchi, Y., Y.-D. Jian, S. Ramalingam, and C. Feng. Point-plane SLAM for hand-held 3D sensors. In: *Robotics and Automation (ICRA), IEEE International Conference on*, pp. 5182–5189 (2013).

Tam, G. K. L., Z.-Q. Cheng, Y.-K. Lai, F. C. Langbein, Y. Liu, D. Marshall, R. R. Martin, X.-F. Sun, and P. L. Rosin. Registration of 3D point clouds and meshes: a survey from rigid to nonrigid. *Visualization and Computer Graphics, IEEE Transactions on*, **19(7)**, pp. 1199–1217 (2013).

Teanby, N. A. An icosahedron-based method for even binning of globally distributed remote sensing data. *Computers & Geosciences*, **32(9)**, pp. 1442–1450 (2006).

Thabet, A. K., E. Trucco, J. Salvi, W. Wang, and R. J. Abboud. A dynamic 3D foot reconstruction system. In: *Engineering in Medicine and Biology Society (EMBS), 33rd Annual International Conference of the IEEE* (2011).

Theiler, P. W. and K. Schindler. Automatic registration of terrestrial laser scanner point clouds using natural planar surfaces. *Photogrammetry, Remote Sensing and Spatial Information Sciences, ISPRS Annals of*, **I(3)**, pp. 173–178 (2012).

Tian, Z., F. Gao, Z. Jin, and X. Zhao. Dimension measurement of hot large forgings with a novel time-of-flight system. *International Journal of Advanced Manufacturing Technology*, **44(1-2)**, pp. 125–132 (2009).

Tippetts, B., D. J. Lee, K. Lillywhite, and J. Archibald. Review of stereo vision algorithms and their suitability for resource-limited systems. *Journal of Real-Time Image Processing*, **January**, pp. 1–21 (2013).

Tombari, F., S. Salti, and L. DiStefano. Performance evaluation of 3D keypoint detectors. *International Journal of Computer Vision*, **102(1-3)**, pp. 198–220 (2013).

Torre-Ferrero, C., J. R. Llata, S. Robla, and E. G. Sarabia. 3D registration of point clouds from time-of-flight cameras. *Electronics Letters*, **47(3)**, pp. 175–177 (2011).

Tzimiropoulos, G., V. Argyriou, and T. Stathaki. Subpixel registration with gradient correlation. *Image Processing, IEEE Transactions on*, **20(6)**, pp. 1761–1767 (2011).

Valanis, A. and M. Tsakiri. Automatic target identification for laser scanners. In: *Technical Commision V*, volume 35 of *International Society for Photogrammetry and Remote Sensing (ISPRS)*, pp. 1–6 (2004).

Valsecchi, A., J. Dubois-Lacoste, T. St utzle, S. Damas, J. Santamaría, and L. Marrakchi-Kacem. Evolutionary medical image registration using automatic parameter tuning. In: *Evolutionary Computation (CEC), IEEE Congress on*, pp. 1326–1333 (2013).

Wang, D. and H. Qian. 3D object recognition by fast spherical correlation between combined view EGIs and PFT. In: *Pattern Recognition (ICPR), 19th International Conference on* (2008).

Williamson, D. L. The evolution of dynamical cores for global atmospheric models. *Journal of the Meteorological Society of Japan*, **85B**, pp. 241–269 (2007).

Xiao, J., B. Adler, and H. Zhang. 3D point cloud registration based on planar surfaces. In: *Multisensor Fusion and Integration for Intelligent Systems (MFI), IEEE Conference on*, pp. 40–45 (2012).

Xiong, Y., A. Chakrabarti, R. Basri, S. J. Gortler, D. W. Jacobs, and T. Zickler. From shading to local shape. arXiv:1310.2916 [cs.CV] (2013). Computer Vision and Pattern Recognition.

Yang, S.-W., C.-C. Wang, and C.-H. Chang. RANSAC matching: simultaneous registration and segmentation. In: *Robotics and Automation (ICRA), IEEE International Conference on*, pp. 1905–1912 (2010).

Yao, J., M. R. Ruggeri, P. Taddei, and V. Sequeira. Automatic scan registration using 3D linear and planar features. *3D Research*, **1(3)**, pp. 1–18 (2011).

Zhang, R., P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: a survey. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, **21(8)**, pp. 690–706 (1999).

Zhang, Z., S. H. Ong, and K. Foong. Improved spin images for 3D surface matching using signed angles. In: *Image Processing (ICIP), International Conference on*, pp. 537–540 (2012).