IMPROVED ALGORITHMS FOR AUTOMATIC REGISTRATION OF 3D POINT CLOUDS

A thesis submitted
in partial fulfilment for the degree of

Doctor of Philosophy

by

DHANYA S PANKAJ

Department of Earth and Space Sciences

INDIAN INSTITUTE OF SPACE SCIENCE AND TECHNOLOGY
Thiruvananthapuram - 695547

February 2017

CERTIFICATE

This is to certify that the thesis entitled **IMPROVED ALGORITHMS FOR AUTOMATIC REGISTRATION OF 3D POINT CLOUDS** submitted by **Dhanya S Pankaj** to the Indian Institute of Space Science and Technology, Thiruvananthapuram, in partial fulfilment for the award of the degree of **Doctor of Philosophy** is a *bona fide* record of research work carried out by her under my supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institution or University for the award of any degree or diploma.

Dr. Rama Rao Nidamanuri
Supervisor
Department of Earth and Space Sciences

Thiruvananthapuram
February 2017

Counter signature of HOD with seal

DECLARATION

I declare that this thesis entitled **IMPROVED ALGORITHMS FOR AUTOMATIC REGISTRATION OF 3D POINT CLOUDS** submitted in partial fulfilment of the degree of Doctor of Philosophy is a record of original work carried out by me under the supervision of **Dr. Rama Rao Nidamanuri**, and has not formed the basis for the award of any other degree or diploma, in this or any other Institution or University. In keeping with the ethical practice in reporting scientific information, due acknowledgements have been made wherever the findings of others have been cited.

Thiruvananthapuram-695547 Dhanya S Pankaj

February 2017 (SC12D014)

ACKNOWLEDGEMENTS

My Ph.D. journey has witnessed all versions of me and I would not have emerged alive out of it without the understanding and support of the people around me. It has been a blissful journey of learning both in professional and personal life and I owe my gratitude to many individuals.

First and foremost I would like to thank my supervisor Dr. Rama Rao Nidamanuri, whose support and guidance throughout the years have made this journey possible. I am grateful to him for believing in me and giving me the freedom to explore a variety of ideas. I also express my sincere gratitude to him for supporting me during the times when I had to be away from the campus. I am thankful to him for helping me to be an independent researcher.

I would like to express my sincere gratitude to Dr. Gorthi R.K.S.S. Manyam, my Doctoral Committee member, for his whole hearted help and guidance whenever I approached him. I would like to thank all my Doctoral Committee members - Dr. Ramakrishnan Krishnan, Prof. P. K. Garg, Dr. Raju K. George, Dr. A. Chandrasekar and Dr. Anandmayee Tej for their support and encouragement throughout my research. I would like to acknowledge my deep gratitude to Dr. Bhanu Prasad, who has initiated me into the wonderful world of 3D imaging. I would like to thank the thesis reviewers and the anonymous journal reviewers for their valuable suggestions.

I would like to acknowledge Director, IIST for providing me an opportunity to be a part of this institute. I am extremely grateful to Dr. Ramiya for all the help and discussions we had during my stay in IIST. I also like to thank all faculty members of IIST, especially Dr. L. Gnanappazham and Dr. Anand Narayan for their support and goodwill. I would like to thank IIST for the infrastructure and financial support during my research work. I take this opportunity to thank the department staff members Ms. Divya and Ms. Celin for their help. My stay in IIST was made pleasant by a bunch of fellow students and friends - (Maneesha, Gayathri, Srujana, Richa, Keerthi, Vidya, Arun, Aneesha, Harsha, Dhanya, Sarah, Sarika, to name a few) and all my lab mates -

I thank them for their companionship and all the wonderful memories. I thank Adheen for helping me take the initial steps in Kinect programming. I would like to mention my gratitude to Dr. Ambili for the pleasant discussions and friendship. I am thankful to my seniors and fellow Ph.D. scholars for providing me insights into the Ph.D. life and guiding me through the right path.

The love and support of my dear friends who have seen my tears of joy and agony always help me move forward. I am thankful to my wonderful friends - Deepa, Surya, Divya H, Febin, Elizabeth, Anjali, Reji, Kasthury, Divya, Susan, Smitha, Aruna, Sreejith, Ram, Vinoth, Prithivi - for always being there for me. Special thanks to my ever-understanding friend Arun Sankar for his care and support. I would like to thank my teacher Ms. K Lalitha Kumari for all the love and prayers which made my life blissful. I express my sincere gratitude to all my teachers for shaping my mind.

I would like to thank my beloved parents, Mr. G. Pankajakshan and Mrs. V. G. Sumohini, and my brother, Mr. Sruthi S Pankaj, for all the love, care, encouragement and support they have showered upon me which chiselled me to the person I am today. I cannot thank them enough for providing everything I always needed. I thank all my family members and my in-laws for understanding and supporting me throughout this path. I would like to thank my cousin brother Mr. Surej, for his love, care and support.

No words would do justice to the tremendous love and support offered by my dear husband Mr. Deepak Sreenivasan, that enabled me to walk this path of ups and downs. His care, encouragement and understanding have made it possible for me to continue my education through four of the five years of our marriage. I thank him for always believing in me and pushing me to hang on to my journey even when I lost hope occasionally. I am extremely thankful to him for being so understanding to my emotional roller-coasters and helping me find peace and happiness.

Last, but not the least, I thank God for guiding me through all the darkness and all the blessings and love. To the many others who have assisted in one way or the other I say thank you.

Dhanya S Pankaj.

ABSTRACT

A 3D scanner/sensor is a device that captures shape (and possibly appearance or colour) information of a real-world object or scene. With the introduction of practical 3D systems and the availability of off-the-shelf sensors, the 3D scanning technology has gained strong momentum in its research and development phase. The introduction of low cost 3D sensors such as Microsoft Kinect into consumer market has shifted the direction of 3D research from developing low cost 3D sensors to developing applications using these sensors. The 3D technology market has been growing rapidly with an estimated Compound Annual Growth Rate (CAGR) of 21% foraying into new application areas. The wide array of application areas which make use of the 3D scan data include rapid prototyping, garment fitting and design, reverse engineering, gaming, animation movies, architecture, real estate, medical, archaeology, inspection and quality control, robotic vision etc.

Due to spatial constraints, an object has to be scanned from multiple view points to create a complete 3D model of it. It is essential to bring the scan data (saved as point clouds, one of the popular digital representations of the scan data) to a reference coordinate system. This process of aligning the scans is called registration. Registration task involves estimation of the rigid body transformation parameters (in the case of rigid body modelling) to align the scans to the reference coordinates. Traditionally, registration task is done with the help of manual intervention using external markers on the object or using precisely controlled machinery. The need for robust automatic registration algorithms which avoid expert intervention or supervision in scanning and registration stages becomes prominent due to the ubiquitous nature of problem. However, due to the complexity of 3D registration problem, the potential of existing algorithms for automated 3D registration is confined to only some stages of the 3D registration process. This thesis explores the development of a complete automatic registration pipeline for 3D modelling that eliminates the need for external markers or tracking equipment.

The point cloud matching problem is challenging due to various factors like ex-

tend of overlap, availability of distinctive geometric features and unstructured nature of the point clouds. This thesis presents a complete framework for obtaining a single point cloud from a sequence of 3D scans of an object or a scene. The 3D registration problem is usually tackled in two steps - pairwise registration and multiview registration. While the pairwise registration aims to align the overlapping pair of scans, the multiview registration aims to align all of the scans to a common coordinate system using the information obtained from pairwise registration. The pairwise registration is performed in a coarse to fine registration strategy. This thesis proposes a novel robust estimation algorithm named 'ProLoSAC' for coarse pairwise registration. Experiments on various 3D datasets indicate that the proposed ProLoSAC algorithm outperforms the existing RANSAC algorithm in terms of accuracy and computational time. This thesis also presents a novel robust multiview registration algorithm (robust motion averaging) which utilises the Lie group structure of the rigid body motion to find global transformations by averaging the redundant relative transformations. The wrong relative transformations are correctly identified and filtered out before motion averaging. The performance of the developed algorithms is compared with the existing approaches on various 3D datasets (acquired using laser scanners, Kinect etc.). Validation of the results indicate superior performance of the proposed algorithm. The generalizability of the developed pairwise registration algorithm has been assessed by adapting it for automatic registration of several multispectral satellite images. An end-to-end automatic framework for 3D point cloud registration for 3D modelling is proposed and tested for different 3D objects in this thesis. The complete framework for 3D registration is implemented in an object oriented open source platform which enables efficient sharing. The cases of object reconstruction as well as scene reconstruction are considered and accurate registration using the proposed framework is achieved. The robust parameter estimation algorithm for pairwise alignment and robust motion averaging algorithm for multiview alignment are promising tools for high precision and automatic registration of 3D point clouds.

ABBREVIATIONS

3D	Three dimensional
3DSC	3D Shape Context
6DoF	6 Degree of Freedom
AWiFS	Advanced Wide Field Sensor
BRAND	Binary Robust Appearance and Normal Descriptor
CAD	Computer Aided Design
CAGR	Compound Annual Growth Rate
CMM	Coordinate Measuring Machine
CMOS	Complementary Metal Oxide Semiconductor
CNES	Centre national d'études spatiales
CPU	Central Processing Unit
CSHOT	Colour Signature of Histograms of OrienTations
DLR	German Aerospace Center
DOG	Difference of Gaussian
EM	Expectation Maximisation
FLANN	Fast Library for Approximate Nearest Neighbours
FPFH	Fast Persistent Feature Histogram
GAH	Geometric Attribute Histogram
GigE	Gigabit Ethernet
GIS	Geographic Information Systems
GMM	Gaussian Mixture Models
GPU	Graphics Processing Unit
HKS	Heat Kernel Signature
HMM	Hidden Markov Model
ICCP	Iterative Clustered Closest Points
ICP	Iterative Closest Point
IR	Infrared
ISS	Intrinsic Shape Signature
IVC	Industrial Vision Camera

kd-tree	k-dimensional tree
KPQ	Keypoint Quality Index
LiDAR	LIght Detection And Ranging
L-M	Levenberg-Marquardt
LoSAC	Locally Optimized RANSAC
LRA	Local Reference Axis
LRF	Local Reference Frame
LSP	Local Surface Patches
LUM	Lu and Milios
MAPSAC	Maximum APosteriori Sample Consensus
MeshDoG	Mesh Difference of Gaussian
MeshHoG	Mesh Histogram of Gaussian
MLESAC	Maximum Likelihood Estimate SAmple Consensus
MST	Minimum Spanning Tree
NAPSAC	N Adjacent Points SAmple Consensus
NARF	Normal Aligned Radial Feature
NCC	Normalized Cross Correlation
NDT	Normal Distributions Transform
NMS	Non-maximal suppression
NRSC	National Remote Sensing Centre
OGH	Oriented Gradient Histogram
OLI	Operational Land Imager
PCA	Principal Component Analysis
PCL	Point Cloud Library
PFH	Persistent Feature Histogram
ProLoSAC	Progressive sampled Locally Optimised RANSAC
PROSAC	PROgressive SAmple Consensus
PS	Point Signature
RANSAC	RANdom SAmple Consensus
RGB-D	Red-Green-Blue-Depth
RMSE	Root Mean Square Error
RoPS	Rotational Projection Statistics
R-RANSAC	Randomised RANSAC

SDH	Spatial Distribution Histogram
SHOT	Signature of Histograms of OrienTations
SI	Spin-Image
SIFT	Scale Invariant Feature Transform
SI-SIFT	Shape Index - Scale Invariant Feature Transform
SLAM	Simultaneous Localisation And Mapping
SO	Special Orthogonal
SRG-NDT	Segmented Region Growing NDT
SS	Surface Signature
SSSP	Single Source Shortest Path
SURF	Speeded Up Robust Features
SUSAN	Smallest Univalue Segment Assimilating Nucleus
SVD	Singular Value Decomposition
TFT	Thin Film Transistor
TLS	Terrestrial Laser Scanner
TriSi	Tri-Spin-Image descriptor
ToF	Time of Flight
UAV	Unmanned Aerial Vehicle
USC	Unique Shape Context
USGS	United States Geological Survey
UWA	University of Western Australia
VC	Vision Components
VD-LSD	Variable Dimensional Local Shape Descriptor
VGA	Video Graphics Array

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

Introduction

Documenting the natural environment is innate to human nature and can be observed in early inscriptions and cave drawings. With the invention of photography in 1826, new possibilities for archiving and documenting opened doors to the human kind. However, any 2D representation has its own limitations in reproducing the 3D world. Even though the basic principles of obtaining 3D data was known to the human kind, and first 3D scanning technology was created in 1960's, practical 3D scanners evolved only in the last three decades (Breuckmann, 2014; Ebrahim and Mostafa, 2011). The initial optical systems used light projectors and cameras to capture 3D data which later evolved to use white light and lasers. Meanwhile in 1980's, the tool making industry developed contact probes to capture surface shape. This was not only slow, but also could not be used to scan delicate objects as it may cause surface damage by contact. This led to the adaptation of optical technology which included point laser, laser stripes and area scanners. Ebrahim and Mostafa (2011) provided a brief history on the evolution of 3D laser scanners.

The 3D scanner market is a rapidly growing one and according to a market report by 'Markets and Markets', it is expected to grow from USD 3.41 Billion in 2015 to USD 5.90 Billion by 2022, at a Compound Annual Growth Rate (CAGR) of 9.6% between 2016 and 2022 (Markets and Markets, 2016). The 3D scanner hardware market size, in terms of volume, is expected to reach 114.9 thousand units by 2022, at a CAGR of 10.9% between 2016 and 2022. The 3D technology market which covers the hardware and software was valued at $46.0 billion according to a report by Person (2014), and is expected to reach $175.1 billion by 2020, supported by a CAGR of 21%. These trends shows that the 3D scanning and modelling market is expected to be highly dynamic and experience rapid product innovation. Development of technologies to handle this huge amount of data efficiently is the need of the hour.

The following sections provide necessary background information for 3D registration. Various 3D acquisition techniques, off-the-shelf 3D scanners and some of the applications of 3D imaging are discussed in the following sections.

1.1 3D image Acquisition Techniques

3D imaging generally refers to the technique of acquiring some properties of an object as a function of (x, y, z) coordinates. Volumetric techniques used in medical imaging like Computed Tomography and Magnetic Resonance Imaging can reconstruct the 'true 3D image' including the internal structure of the object (Geng, 2011). Here in contrast, the term 3D imaging or 3D scanning refers to surface imaging or shape sensing where the (x, y, z) coordinates of points on the surface of an object are measured with respect to a reference coordinate system.

The 3D acquisition techniques can be broadly classified into contact and non-contact techniques. Contact techniques include the contact probe based systems like Coordinate Measuring Machines (CMMs) which move a mechanical probe over the surface of the object in accurately controlled motion to obtain the shape information. Although very accurate measurements are obtained, this technique is slow and cannot be applied on delicate objects, which led to the adaptation of non-contact techniques. Among the non-contact methods, the optical methods are the most popular. The optical non-contact 3D acquisition techniques can be broadly classified into active and passive methods (Moons et al., 2009). Specifically controlled lighting or illumination is used to obtain the 3D data in active methods whereas the passive methods basically depend only on the ambient light. For surfaces without adequate well defined features, passive methods require high-contrast targets to be added to the surface for easy localization (Domey and Picard, 2002). They are also highly affected by the variations in ambient light. The active methods surpass this limitation by controlled illumination. However, they face difficulties in dealing with varying surface finish or sharp discontinuities (El-Hakim et al., 1995). Another classification is based on the number of vantage points from where the images are acquired or illuminated. The single vantage point methods include the time of flight, shape from x (shading, texture, defocus, contact, occlusion, contour) methods etc. The multiple vantage points methods include stereo, structure from motion, structured light etc. The 3D acquisition techniques are summarised in Fig.1.1. Hussmann et al. (2013) presented an overview of the 3D acquisition techniques used in some of the commercially available 3D sensors along with their application scenarios. A comparison of the different sensors was presented by Sansoni et al. (2009). Some of the popular techniques are briefly explained

in the below section [1].

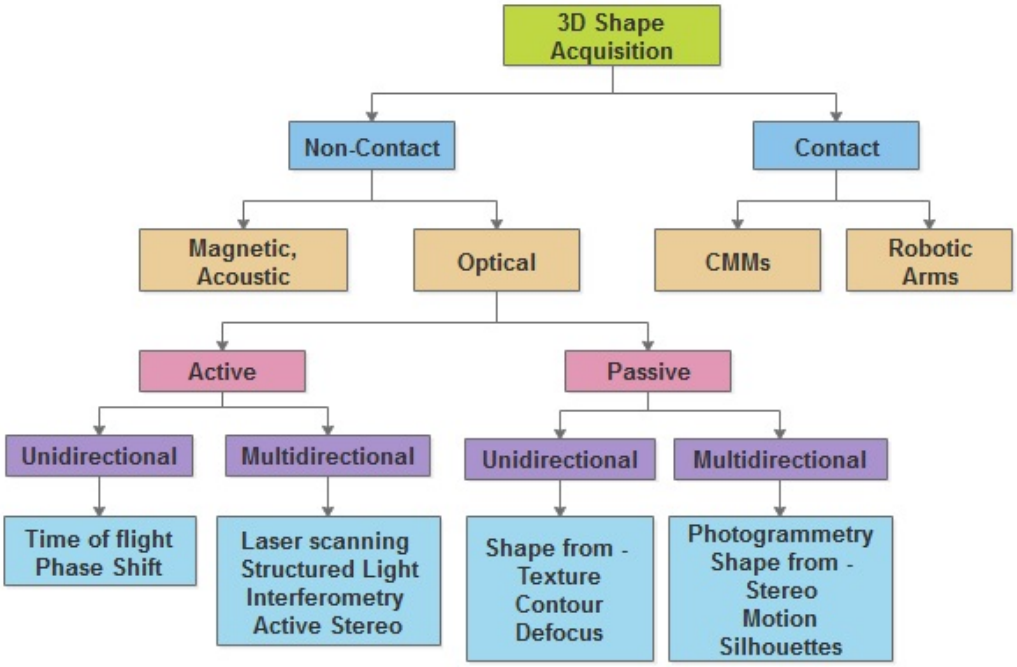


Figure 1.1: Taxonomy of 3D acquisition methods.

1.1.1 Laser Scanning

Laser scanning is an active triangulation method and the system consist of three main parts viz. the transmitter, receiver and the processing part. The transmitter consists of a laser diode which projects the laser light on the object to be measured. The reflected light is received by the receiver and directed towards a detector. The position of the reflected light on the detector varies according to the height of the object. The arrangement is shown in Fig.1.2. The scanners are classified into point, line and area scanners based on the object area scanned at single shot. The sensor (diffuse or specular) is selected based on the reflective properties of the object to be measured. Properly designed laser triangulation sensors offer extremely high resolution and stability, often approaching that of expensive and complex laser interferometer systems. Their ability to detect such small motions is utilized in many demanding, high-precision

[1] Some of the sections in this chapter are published in *Proceedings of International Conference on Innovations in Computer Science and Engineering*, 2013. Authors: Dhanya S Pankaj, Rama Rao Nidamanuri, Prasad, P. B

measurement applications. The primary factor in determining resolution is the system electrical noise (Pavlidis et al., 2007).

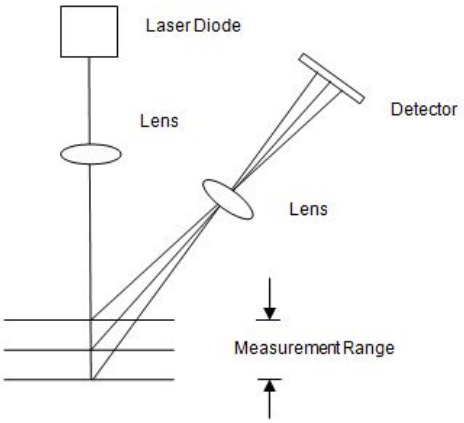


Figure 1.2: Laser triangulation set-up

Many of the commercial scanners in use today work on laser triangulation principle. Examples of industrial scanners include Rangers, Rulers, Scanning Rulers, IVC3D and TriSpector1000 by SICK, DS1000 3D displacement sensors by Cognex, VC Nano3D by Vision Components, Faro sensors etc.

1.1.2 Time of Flight and Phase Shift techniques

Time of flight (ToF) technique can obtain the 3D surface information of an object in real time. The LiDAR (Light Detection and Ranging) systems based on ToF principle are used in the areas of computer vision, remote sensing, autonomous navigation and robotics. The LiDAR imaging is well suited for capturing 3D images from long ranges. The LiDAR instrument sends rapid laser pulses at the object at rates up to 150,000 pulses per second. A sensor on the instrument calculates the amount of time taken by each pulse to bounce back. Light moves at a constant, known speed and the LiDAR instrument can calculate the distance between itself and the target with high accuracy as Distance = (Speed of Light x Time of Flight) / 2. As these systems capture 3D images of an object point by point, scanning is required. LiDAR gives excellent precision at large distances. Typically, errors at tens of meters will be within the range of a few centimetres (LiDAR, 2016).

Another technique which is similar in working principle to ToF systems is the

phase-shift technique. The working principle of phase shift systems is shown in Fig.1.3 (Hansard et al., 2012). The object is actively illuminated with an incoherent light signal which is intensity modulated by a cosine-shaped signal of frequency f_{mod}. The reflected light is captured by the sensor and the phase difference between the emitted and reflected light is used to calculate the depth of the object. Depth is given by Eq.1.1, where c represents the velocity of light, d the distance travelled by the light, f_{mod} is the modulation frequency and $\Delta\psi$ the phase shift.

$$d = \frac{c}{2}\frac{\Delta\psi}{2\pi f_{mod}} \tag{1.1}$$

The range is dependent on the modulation frequency of the camera, as this frequency defines the wave length of the emitted signal. Due to their high lateral resolution, ToF cameras have very dense depth information at constant resolution and high frame rates without the need of interpolation (Schaller, 2011).

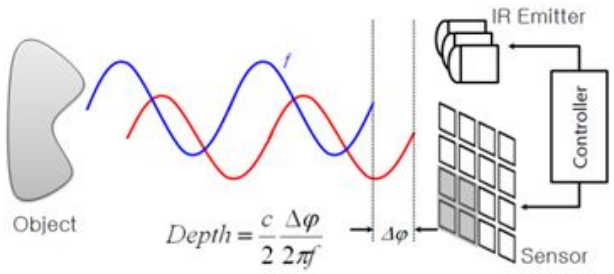


Figure 1.3: Phase-shift measurement principle (Hansard et al., 2012).

Commercially available ToF scanners include the SICK 3vistor-T, ifm electronic 3D-camera O3D200, MESA Imaging SR4000, Fotonic E40/E70 camera, Microsoft Kinect for XBox One etc.

1.1.3 Shape from Structured Light

This method works based on the principle of triangulation and is often confused with laser scanning methods (Pavlidis et al., 2007). A specific pattern is projected onto the surface of an object and its geometry information is obtained by recording and processing the deformations of this pattern caused by the object surface. Under controlled environment, this method can yield very accurate 3D shape measurement

and several variations of this technique are available in literature (Salvi et al., 2004; Guehring, 2000). The various structured light techniques and applications were surveyed by Geng (2011). Fotonic P70 camera, Microsoft Kinect for XBox 360 etc. are commercial 3D scanners based on structured light.

1.1.4 Shape from Stereo

Stereo vision is a passive 3D acquisition technique widely used in research and industries. The depth information is obtained by using two or more images of the scene each acquired by multiple cameras displaced in space or by a single camera in motion. Depth of a point can be calculated by means of a technique called triangulation. There are two main problems to be solved in stereo vision - the correspondence problem and the reconstruction problem. Finding the corresponding image points of a scene point in the two images forms the correspondence problem. Once the corresponding points are obtained, disparity can be calculated. If the geometry of the stereo set up and disparity are known, the depth can be calculated, which forms the reconstruction problem. Theoretical stereo camera depth resolution is influenced by imaging geometry and quantization resolution. However, the depth resolution obtained from a stereo camera is affected by multiple factors such as calibration and matching errors of correspondence points (Kytö et al., 2011). The two camera parameters - focal length and baseline affect the depth resolution obtainable in stereoscopic systems. Proper calibration of the cameras is critical in achieving accurate results. Shape from motion and photogrammetry also depend on the basic principle of stereo vision. SICK 3vistor-P is a commercial scanner based on the stereo vision principle.

1.1.5 Shape from Texture

Textures on an object surface can be thought of as made up of instances of elements, which are distributed in some fashion. The information about the geometry of the surface can then be inferred based on the distortion of this distribution or the distortion of the elements or both (Lobay and Forsyth, 2006). The transformation estimated based on the distortions are used to extract the 3D surface geometry. The method is photographic and is of low cost, but has only limited applicability (Pavlidis

et al., 2007).

1.1.6 Shape from Shading

This technique is used to recover the 3D information of an object from a single view. The human visual system is capable of inferring depth from the difference in shading and this technique tries to simulate the same. Shape is inferred from a gradual variation in shading in the image (Zhang et al., 1999). Given a gray level image, shape from shading technique tries to recover the light source and surface shape at each pixel. This technique captures both texture and geometry. The sensors using this technique are of low cost, but less accurate and have difficulties in capturing texture in shaded areas (Pavlidis et al., 2007).

1.1.7 Shape from Photometry

Photometric stereo methods recover the shape and albedo of an object by capturing images from a single view point by varying the direction of incident illumination. The recovery of shape is based on the principle that the amount of light reflected by a surface is dependent on the orientation of the surface in relation to the light source and the observer. This provides sufficient information to estimate the surface normals. Since there is no change in the imaging geometry, each pixel in the captured images consists of the same object point and this provides correspondence. By varying the direction of incident illumination, it is the reflectance map that get changed and this information can be processed to retrieve the surface orientations (Woodham, 1980).

1.1.8 Shape from Focus/Defocus

In this method, the shape of the scene is estimated from a set of two or more images of a scene from the same point of view, obtained by changing the camera parameters (typically the focal setting or the image plane axial position) (Favaro and Soatto, 2002). Depth from defocus has been proven to be effective for small distance i.e. microscopic images. The depth of field information in images are efficiently utilised to extract the surface geometry of the object.

1.2 3D Imaging Sensors

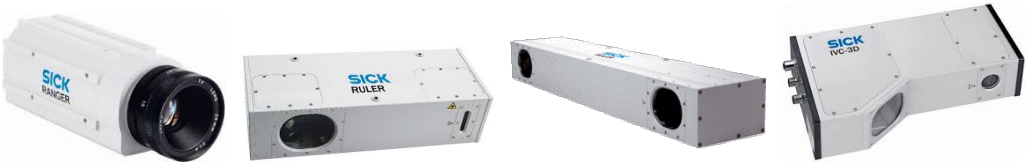
This section presents some of the commonly used 3D acquisition devices in market which are based on different working principles. The purpose of this section is to briefly describe the available 3D imaging sensors and does not endorse any of the camera models or make.

1.2.1 SICK 3D Cameras

SICK, one of the major manufacturers of industrial 3D solutions, has a range of 3D cameras in market. They are classified into different families based on their features and use viz. Ranger, Ruler, Scanning Ruler, IVC-3D, 3vistor-P, TriSpector1000, 3vistor-T etc. (SICK, 2016). The Fig.1.4 shows the range of SICK 3D cameras.

The sensors which work on the laser triangulation principle are Rangers, Rulers, Scanning Rulers, IVC3D and TriSpector1000. The Ranger family of sensors provides fast 3D profile measurement and multi scan, and with their high flexibility and versatility, they are well suited for advanced industrial solutions. Ranger cameras extract the true 3D shape of an object, regardless of its contrast or colour. It is used to measure object height and volume, to detect shape defects, and for quality grading and size sorting. These cameras have a Multi-scan tool which can capture a multitude of object features such as contrast, gloss, and scatter at the same time thus enabling reliable inspection results and more cost-efficient solutions since it only takes one Ranger to perform it all. The cameras feature sensor resolutions of up to 1,536 pixels in 3D and 3,072 pixels in grayscale and colour and also provide in-machine 3D calibration which helps to attain mm dimensions.

The Ruler range of cameras are suited for in-line industrial applications and are robust to very harsh environments and temperatures as low as $-30°C$. They provide GigE interface making them suitable for remote operation over long distances. They provide high speed precise measurements (sub-mm range) and also provide grayscale intensity and laser scatter measurements. The Scanning Ruler cameras are suitable for robot-aided gripping applications. They provide mm-precise measurements and capture 2D and 3D images making them suitable for part identification and gripping.



(a) SICK Ranger (b) SICK Ruler (c) SICK Scanning Ruler (d) SICK IVC-3D

(e) SICK 3vistor-P (f) SICK TriSpector1000 (g) SICK 3vistor-T

Figure 1.4: SICK 3D sensors.

(a) Cognex DS1000 3D (b) iDS Ensenso 3D (c) VCNano3D (d) Gocator 3D

Figure 1.5: 3D sensors.

(a) FARO3D sensor (b) Kinect for XBox 360 (c) Kinect for XBox One

Figure 1.6: Faro and Kinect 3D sensors.

The IVC-3D camera is acclaimed as the first 3D smart camera which is a self-contained factory-calibrated smart camera that combines imaging, lighting and analysis into one camera housing. The TriSpector1000 is a stand-alone configurable sensor for cost-efficient 3D inspections. It has factory calibrated 3D data with true mm values

in all dimensions. It has embedded image analysis for easy configuration for a variety of tasks.

The SICK 3vistor-P is based on the stereoscopic principle. It is a highly intelligent driver assistance system for collision awareness in industrial vehicles and is well suited for working in outdoor environments, e.g. in mines, ports and construction sites. The 3vistor-P supports drivers manoeuvring large vehicles by continuously monitoring blind spots in three dimensions. The SICK 3vistor-T 3D vision sensors work on an innovative snapshot time-of flight technique and provide real-time depth information for each pixel, even for stationary applications. They offer maximum flexibility for indoor use and record up to thirty 3D images per second. 3vistor-T provides more than 25,000 distance and intensity values in a single shot. Hence 3D information is readily available for stationary applications without the need for an actuator.

1.2.2 Congnex DS1000

DS1000 3D displacement sensors (Fig.1.5(a)) by Cognex are industrial 3D sensors which work on the principle of laser triangulation. These factory-calibrated 3D sensors measure features such as length, width, height, tilt or volume relative to any surface in real world units with micron-level accuracy. They are widely employed in areas like food and pharmaceutical production, electronics and consumer product manufacturing, automotive industry etc. (Cognex, 2016). These are factory calibrated vision systems with proprietary software and can withstand harsh factory environments.

1.2.3 iDS Ensenso 3D sensors

The Ensenso 3D cameras by iDS (Fig.1.5(b)) work according to the projected texture stereo vision principle. The projected pattern is used in addition to the stereo cameras to capture details even in areas of poor texture. A camera has two integrated CMOS sensors and a projector that casts a random point pattern onto the object to be captured, allowing structures that are not visible or only faintly visible on the surface to be enhanced or highlighted. The cameras are factory calibrated and the housing is suitable for industrial use. It can capture both stationary and moving objects (iDS,

2016).

1.2.4 VC Nano3D sensors

VC Nano 3D smart cameras by Vision Components (Fig.1.5(c)) work according to laser triangulation. Measuring merely 140 x 70 x 35 mm, the camera housing includes an intelligent camera and a line laser which enables the real-time recording of images at a scan rate up to 400 Hz. The images can be analysed by the digital signal processor of the smart camera (VisionComponents, 2016). The different models available cover a wide range of application scenarios.

1.2.5 Gocator sensors

Gocator All-In-One 3D Smart Sensors by LMI technologies (Fig.1.5(d)) combine 3D scanning, measurement and control in a single device with no external PCs or control required. The Gocator leverages both laser triangulation and fringe pattern projection thus providing optimal 3D scanning for both high speed and stationary 3D part inspection. The sensors are available in single point profile, line profile and snapshot technologies. The sensors feature scanning rates up to 32000 Hz and micron level resolution (LMI, 2016). More than just a sensor, each Gocator houses a complete 3D inspection ecosystem.

1.2.6 FARO 3D sensors

FARO features a number of 3D sensors (Fig.1.6(a)) suitable for a wide range of applications. The laser scanner series called FARO Focus^{3D} are used for 3D documentation and surveying. The sensors vary in the range of distance they cover and provide distance accuracy up to ± 2 mm. Another 3D scanner by FARO is the FARO Scanner Freestyle3D which is a smart hand held 3D laser scanner for professionals. It offers fast data acquisition, real-time visualisation and the largest scan volume on the market. The scanner can produce 3D point accuracy up to 1.5 mm (FARO, 2016).

1.2.7 Microsoft Kinect

Kinect sensor is a mass market low cost RGB-D sensor (Fig.1.6(b)) originally developed as part of Microsoft Xbox game console and now being extensively used for 3D computer vision applications. The basic principle is structured light where the IR projector projects an IR speckle dot pattern onto the object and the IR-pass filtered CMOS camera captures the reflected light. The depth is calculated from the deformation of the known pattern, caused by object surface irregularities. In addition to the depth sensor, Kinect also has an RGB camera which returns the image intensity. It provides a frame rate of 30 Hz and a spatial resolution of 640×480 (VGA). Spatial resolution is around 3 mm and depth resolution is 1 cm at 2 m distance. The field of depth sensing is around 0.8 to 3.5 m (Andersen et al., 2015). The second generation of Kinect as part of XBox One (Fig.1.6(c)) uses ToF principle and also features a high resolution video camera.

1.3 Applications of 3D imaging

The application areas of 3D scan data encompass a wide spectrum of industries and academic research. 3D scanning finds application in rapid prototyping, reverse engineering, industrial inspection and quality control, metrology, animation in movies and gaming, 3D modelling, robotic vision etc. and the services benefited include medical and dental, museum and archaeology, architecture, automotive, fashion and textiles, defense, security, entertainment etc. The recent introduction of many low-cost 3D data acquisition systems and the technological advancements in visualizing and processing 3D data have accelerated the growth of interest in academia and industry. The rapid technological developments and the growing adoption of 3D scanners to enhance the quality of products and to reduce their manufacturing time are some of the significant growth drivers for the 3D scanner market (Markets and Markets, 2016). Some of the typical application scenarios of 3D data are detailed here (Pankaj et al., 2013).

1.3.1 Surface Inspection

Examination of the surface of manufactured parts is increasingly being done using 3D vision. Using 3D vision, contrast independent inspection is possible. Some typical surface inspection tasks include break pad inspection, mandrel bar surface inspection, inspection of wood surfaces for knots, geometry and texture, detection of surface defects in paintings, TFT display glass inspection, leather inspection etc. 3D vision sensors which can provide micron level accuracy may be required for surface finish inspection tasks.

1.3.2 Assembly Lane Inspection and Volume Estimation

Various robotic vision based tasks such as assembly lane inspection can be made easy using 3D vision. The 3D information can be used to identify the defective parts in the assembly lane. 3D vision system uses the 3D coordinate information to identify and locate the parts and then sends the id and 3D location to the picking robot. The system then instructs the robot to pick up the identified objects which are located at the top. Another typical application is the classification of objects based on size and volume. For example, in food processing industry, the volume of oysters is calculated by making use of depth data and the oysters are directed to separate bins according to their volume (SICKRanger, 2016).

1.3.3 Motion Tracking

Motion tracking applications are made easy with the use of third dimension. For example, the Microsoft Kinect sensor (Andersen et al., 2015) is used in Microsoft XBox gaming applications where the player motion is identified using the 3D data. The system developed for human motion tracking (Zhang et al., 2012) makes use of the depth information provided by multiple Microsoft Kinect sensors. The system creates a joint point cloud from the depth images obtained from multiple cameras.

1.3.4 3D Digitization of objects

The 3D digitization of objects finds applications in fields like 3D printing, preservation of cultural artefacts, reverse engineering etc. The digitization of models or monuments of cultural importance is generally performed for conservation and creation of a digital museum for remote study and appreciation of these objects. Digital archives are durable and unalterable and can be used as a reference for degradation monitoring and restoration works (Pieraccini et al., 2001).

1.4 3D data Representation

The 3D data can be represented in different formats - as a set of unstructured points or as a matrix of depth values. In a more general case, some properties of the object surface like surface reflectance (intensity or RGB values) are also captured along with the coordinate information. 3D data acquired from the acquisition devices are commonly represented in one of the standard formats mentioned below which are then used for further processing.

Range images: Range images are a class of digital images which represent depth information. Each pixel of a range image expresses the distance between a known reference frame and a point in the scene. Hence, a range image reproduces the 3D structure of a scene. Range images are also referred to as depth images, depth maps, xyz maps, surface profiles and 2.5D images. A range image is a matrix of depth values of points along the directions of the x, y image axes with explicit spatial organisation. A sample range image is given in Fig.1.7

RGB-D images: RGB-D images are similar to the range images except for the fact that they have colour information in addition to the depth information. The images are captured with the help of RGB-D sensors which capture RGB data and depth. The colour camera and the depth sensor have to be carefully calibrated so as to obtain an aligned RGB-D image.

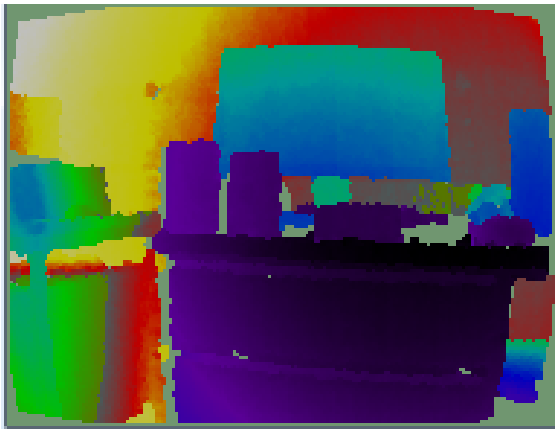
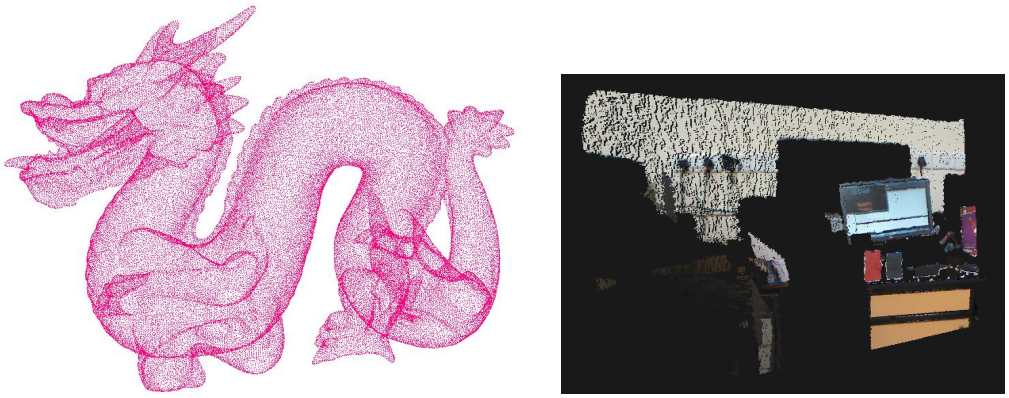


Figure 1.7: A sample range image.



(a) Point cloud (b) Point cloud with texture

Figure 1.8: Snapshot of sample point clouds.

Point clouds: Most of the laser scanners output 3D data in the form of point clouds. A point cloud is nothing but a collection of data points in some coordinate system. In a three-dimensional coordinate system, these points are usually defined by x, y, z coordinates and generally there is no particular ordering for the points (Fig.1.8(a)). Some range scanners provide point clouds which are organized (with fixed width and height) similar to range images. A variety of the scanners also provide surface normal information of each of the points. Such points are called oriented points. 3D scanning systems sometimes provide the texture information of points synchronized along with the depth information. Such point clouds contains R, G, B information in addition to the x, y, z coordinates as in Fig.1.8(b).

In this thesis, the 3D point cloud representation is adopted.

1.5 3D Registration

3D modelling using the acquired 3D scan data is widely used in many of the applications. 3D registration is an important stage in the 3D modelling pipeline after the acquisition of 3D data (Fig.1.9). The success of 3D registration is critical in determining the performance of the subsequent stages of processing. This section briefly explains the essential background and mathematics involved in the problem of 3D registration.

To obtain a complete 3D model of an object, the 3D data should be acquired from multiple view points. This is required because of the scanning limitation of the devices or due to self occlusion caused by the object surface. In this thesis, 3D modelling of rigid objects is considered. The complete surface of an object can be captured either by moving the object in front of the camera or by moving the camera around the object. Fig.1.10 shows a simple scanning set up where the camera is moved so as to obtain the scans from multiple view points. The camera sensor plane is shown in rectangle and the arrow indicates the z-axis direction.

The motion of camera around the object or object in front of the camera can be modelled by rigid body motions. A brief description of this formulation is given here.

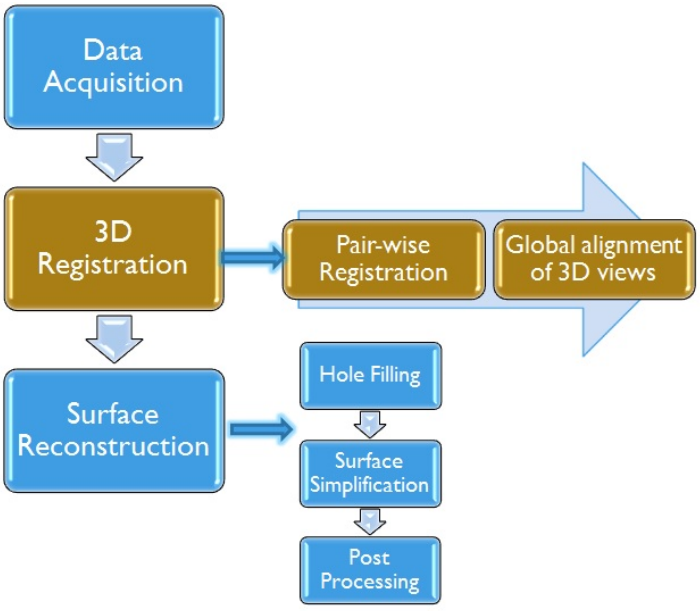


Figure 1.9: 3D modelling pipeline.

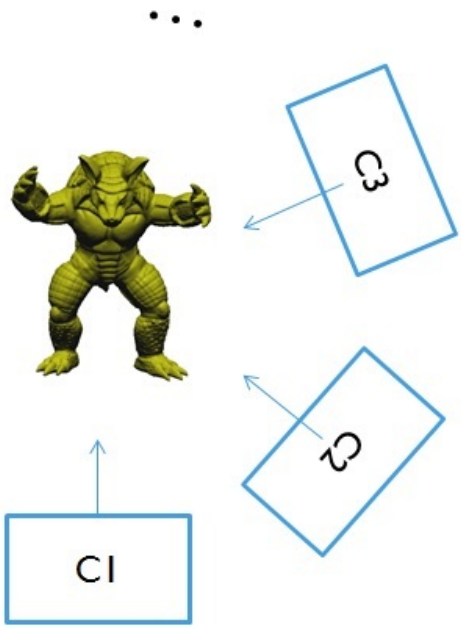


Figure 1.10: 3D scanning setup.

1.5.1 Rigid body motion

As the overarching goal of this thesis, the 3D registration problem is studied from the rigid body modelling perspective, a brief description of the rigid body motion is presented here before describing the 3D registration problem.

Consider an object moving in front of a camera. In order to describe its motion, one should specify the trajectory of every single point on the object. Fortunately, for rigid bodies, it is sufficient to specify the motion of one point and the motion of three coordinate axes attached to that point. This is because, for every rigid object, the distance between any two points on it does not change over time as the object moves (Ma et al., 2012).

If $\boldsymbol{X}(t)$ and $\boldsymbol{Y}(t)$ are the coordinates of any two points p and q on the object, then distance between them is constant as given by Eq.1.2.

$$\|\mathbf{X}(t) - \mathbf{Y}(t)\| \equiv \text{constant}, \quad \forall t \in \mathbb{R} \tag{1.2}$$

A *rigid-body motion* or rigid body transformation is a family of maps that describe

how the coordinates of every point on a rigid body change in time while satisfying Eq.1.2. Such a map can be denoted by Eq.1.3. If instead of considering the complete path, we concentrate on the map between the initial and final configurations, we have a *rigid-body displacement*, denoted as in Eq.1.4.

$$g(t) : \mathbb{R}^3 \longrightarrow \mathbb{R}^3; \qquad \mathbf{X} \mapsto g(t)(\mathbf{X}) \tag{1.3}$$

$$g : \mathbb{R}^3 \longrightarrow \mathbb{R}^3; \qquad \mathbf{X} \mapsto g(\mathbf{X}) \tag{1.4}$$

The map g also acts on vectors. Suppose v is a vector defined by two points p and q with co-ordintes $v = \mathbf{Y} - \mathbf{X}$; then, g acting on v will give the new vector u given by Eq.1.5. Since g preserves the distance between points, we have $\|g_*(v)\| = \|v\|$, for all free vectors $v \in \mathbb{R}^3$.

$$u = g_*(v) \doteq g(\mathbf{Y}) - g(\mathbf{X}) \tag{1.5}$$

A map that preserves the distance is called a *Euclidean transformation* and the set of Euclidean transformations in 3-D space is denoted by $E(3)$. We also require that any rigid body motion, besides preserving distances, preserves orientation as well. Hence, in addition to preserving the norm of vectors, it must also preserve their cross product. The map or transformation induced by a rigid-body motion is called a *Special Euclidean transformation* where the word "special" indicates orientation-preserving.

Definition 1.5.1. ***(Rigid-body motion or special Euclidean transformation)****. A map $g : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ is a rigid-body motion or a special Euclidean transformation if it preserves the norm and the cross product of any two vectors,*

(i) norm: $\|g_*(v)\| = \|v\|, \quad \forall v \in \mathbb{R}^3$

(ii) cross product : $g_*(u) \times g_*(v) = g_*(u \times v), \quad \forall u, v \in \mathbb{R}^3.$

The collection of all such motions or transformations is denoted by SE(3).

The angle between the vectors is also preserved according to the above definition because the inner product can be expressed in terms of the norm by *polarization identity* as in Eq.1.6 and since $\|u + v\| = \|g_*(u) + g_*(v)\|$, we can see that for any rigid

body motion g, Eq.1.7 is satisfied.

$$\langle u, v \rangle = \frac{1}{4}(\|u+v\|^2 - \|u-v\|^2) \tag{1.6}$$

$$\langle u, v \rangle = \langle g_*(u), g_*(v) \rangle , \qquad \forall u, v \in \mathbb{R}^3 \tag{1.7}$$

Thus, a rigid-body motion can be defined as one that preserves both the inner product and the cross product. Since distances and orientations are preserved by a rigid body motion, the individual parts cannot move relative to each other. Hence a rigid body motion can be described by the motion of a point on the body and the rotation of a coordinate frame attached to that point. One can represent the *configuration* of a rigid body by attaching a Cartesian coordinate frame to some point on the rigid body and keeping track of the motion of this frame relative to a fixed *world (reference) frame*.

Consider a coordinate frame with its principal axes given by three *orthonormal* vectors $e_1, e_2, e_3 \in \mathbb{R}^3$ such that they satisfy the equation Eq.1.8.

$$e_i^T e_j = \delta_{ij} \doteq \begin{cases} 1 & \text{for} \quad i = j, \\ 0 & \text{for} \quad i \neq j. \end{cases} \tag{1.8}$$

The vectors are ordered to form a right-handed frame: $e_1 \times e_2 = e_3$. The rigid body motion g acting on the object can be given by Eq.1.9.

$$g_*(e_i)^T g_*(e_j) = \delta_{ij}, \quad g_*(e_1) \times g_*(e_2) = g_*(e_3) \tag{1.9}$$

We can see that the resulting vectors still form a right-handed orthonormal frame. Hence a rigid object can always be associated with a right handed orthonormal frame, which can be called the *object coordinate frame* or the *body coordinate frame*, and its rigid body motion can be entirely specified by the motion of such a frame.

Consider an object moving relative to a fixed world reference frame $W : (X, Y, Z)$. In order to specify the configuration of the object relative to the world frame W, consider a point o on the object and attach to it an object frame, $C : (x, y, z)$. When the object moves, the object frame also moves along with it. The configuration of the object with respect to W is then decide by two components:

(i) the vector between the origin w of the world frame and that of the object frame o, called the "translational" part and denoted by T;

(ii) the relative orientation of the object frame C, with coordinate axes (x, y, z), relative to the fixed world frame W with coordinate axes (X, Y, Z), called the "rotational" part and denoted by R.

In the 3D image acquisition scenario considered in this thesis, the world reference frame and object reference frame can be the camera coordinate frame or the object coordinate frame. What matters is the relative motion between the object or scene and the camera.

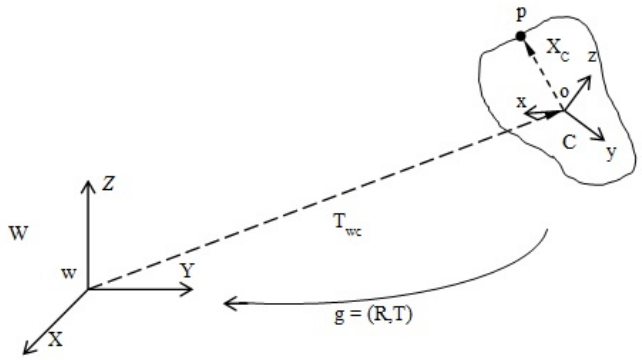


Figure 1.11: A rigid-body motion between object frame C and world frame W.

Fig.1.11 illustrates a moving rigid object with a coordinate frame C attached to it. For a point p on the object, the coordinates $\mathbf{X}_w$ with respect to the world frame W is given by Eq.1.10. R_{wc} is the orthogonal rotation matrix which represents the rotational transformation from coordinates $\mathbf{X}_c$ relative to frame C to the coordinates $\mathbf{X}_w$ relative to the frame W. $T_{wc} \in \mathbb{R}^3$ is the translation of the origin of the frame C relative to that of the frame W.

$$\mathbf{X}_w = R_{wc}\mathbf{X}_c + T_{wc} \tag{1.10}$$

Homogeneous representation: The coordinate transformation of a rigid body is affine and it can be converted to a linear one by using homogeneous coordinates. Appending a "1" to the coordinates $\mathbf{X} = [X_1, X_2, X_3]^T \in \mathbb{R}^3$ of a point p yields a

vector in $\mathbb{R}^4$, denoted by $\bar{\mathbf{X}}$ as given by Eq.1.11.

$$\bar{\mathbf{X}} \doteq \begin{bmatrix} \mathbf{X} \\ 1 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix} \in \mathbb{R}^k \tag{1.11}$$

Using this homogeneous representation, the rigid body transformation can be represented as a linear transformation as given by Eq.1.12.

$$\bar{\mathbf{X}}_w = \begin{bmatrix} \mathbf{X}_w \\ 1 \end{bmatrix} = \begin{bmatrix} R_{wc} & T_{wc} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_c \\ 1 \end{bmatrix} \doteq \bar{g_{wc}} \bar{\mathbf{X}}_w \tag{1.12}$$

Here the 4×4 matrix $\bar{g}_{wc} \in \mathbb{R}^{4\times4}$ is called the *homogeneous representation* of the rigid body motion $g_{wc} = (R_{wc}, T_{wc}) \in SE(3)$. In general, if $g = (R, T)$, then its homogeneous representation is given by Eq.1.13.

$$\bar{g} = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \in \mathbb{R}^{4\times4} \tag{1.13}$$

Now the rigid body transformation of coordinates can be represented by a linear matrix multiplication.

From the above discussion, one can see that the motion of the camera around the object for scanning amounts to motion of the camera coordinate system around the object. This motion can be represented by the rigid body transformation matrix. Thus finding the transformation between the coordinates of the two scan views is equivalent to finding the transformation between the two camera coordinate systems. The registration process aims to find this transformation. The same logic applies to the case when the object is moved before the camera.

1.5.2 Registration

3D data is acquired from multiple view points by moving the camera around the stationary rigid body object. The data is represented in the form of point clouds where the X, Y, Z coordinates are with respect to the moving camera coordinate frame. The

coordinate system changes every time the camera is moved to a different view point and the data from multiple views are related by rigid body transformations.

The goal of registration is to bring all of the partial point clouds of an object to a common coordinate system by estimating the rigid body transformations involved.

Sufficient overlap is maintained between adjacent point clouds to facilitate registration. The extent of overlap required depends on the methods used for registration. Alternatively, the object can be moved in front of a stationary camera with the help of a turn-table set-up. In this case also, the rigid body transformations relate the captured scans. The registration of the partially overlapping scans is usually performed in multiple stages (Gomes et al., 2014), as depicted in Fig.1.12. Since the adjacent scans overlap, they are registered using a pair-wise registration method. Once the pairwise registration is performed, this information is used in a global alignment stage to bring all of the point clouds to a common coordinate system.

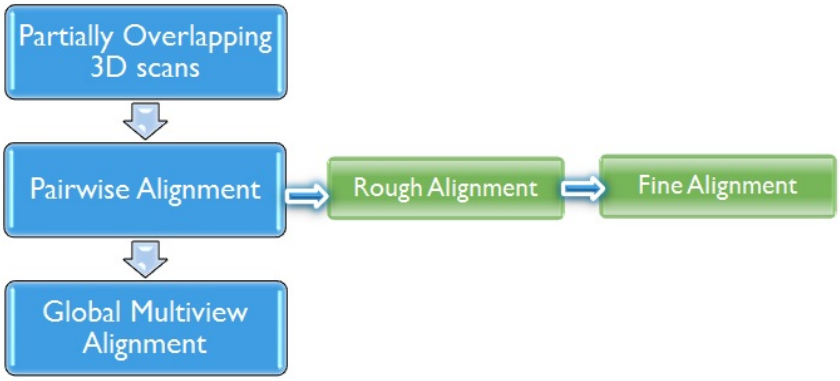


Figure 1.12: Stages in registration of point clouds.

Different methods are attempted in literature to perform the pairwise alignment of point clouds. These methods can be broadly classified into two - rough alignment methods and fine alignment methods. The methods which can provide an alignment of two arbitrarily located (with rigid body transformation) point clouds are called rough alignment methods. These methods aim to find a coarse alignment of the point clouds which serves as an initial estimate for the transformation. The fine alignment methods aim to refine an initial estimate of the transformation. These methods are employed when the point clouds to be registered are already aligned to some extent. The registration is usually performed in a coarse to fine approach - rough alignment followed

by a fine alignment. The decision to employ whether a fine alignment method alone or both methods depends upon the extend of overlap or the transformation between the point clouds. Most fine alignment methods can handle transformations only up to certain degrees of rotation or translation. The number of view points from which scans are to be captured to build a complete 3D model depends on a variety of factors and is an active topic of research (Hébert et al., 2016).

The initial alignment problem is handled in different ways. One of the common methods, especially in industrial scenarios, is to manually align the clouds (Wyngaerd, 2003). Graphic markers are also widely used in industry to establish reliable corresponding points (Wu et al., 2015). Another method for initial registration is to use carefully calibrated motion control instruments which can measure the transformation. The automatic initial registration of point clouds is a vibrant topic of research because it overcomes the limitations of the existing manual or controlled environment methods. The most common automatic alignment method is usually performed by feature matching. In this method, features extracted from the point clouds (or their other representations) are matched so as to obtain corresponding points (the common object points in both the point clouds). Once corresponding pairs are obtained by feature matching, the transformation between the clouds can be estimated by minimization algorithms (Arun et al., 1987; Horn, 1987). However, in majority of the cases a large number of false correspondences is present because of partial overlap, noise in the scan, local nature of features, inaccuracies in the previous stages etc. Hence robust estimators are typically employed to find the good correspondences for estimating the transformation (Han et al., 2015). The widely used robust estimator RANSAC (Random SAmple Consensus) exhibits slow convergence when the percentage of inliers in the set of corresponding points is low. In this thesis, a novel robust estimator which outperforms RANSAC in terms of accuracy and computational time is proposed.

Once the pairwise alignment is performed on the overlapping clouds, the next step is to globally align all the point clouds. One straight forward method to find the global alignment is to register the clouds sequentially by multiplying the transformation matrices. However, this will lead to accumulation of errors (Chen and Medioni, 1992) and hence different methods are adopted in literature to minimize this error and to find an accurate global alignment (Shih et al., 2008). Govindu and Pooja (2014) utilised the redundant overlap between nearby scans to form additional constraints in order to

estimate global transformations by averaging relative transformations. The averaging approach cannot handle the presence of wrong relative motions and leads to corrupt global transformations. A novel algorithm to filter out the wrong relative transformations and to perform robust multiview registration is proposed in this thesis.

1.6 Hypothesis and Research Questions

Due to the ubiquitous nature of 3D data, the availability of automatic registration algorithms has become essential. Pairwise registration is challenged by the presence of outliers in corresponding point pairs generated by matching feature points. This necessitates the development of robust pairwise approaches that can deal with the wrong feature matches. In the case of multiview registration, the complexity of the highly non-linear optimisation problem has led to the development of a variety of approaches. One such approach is to make use of additional constraints introduced by redundant overlap information from the scan sequence. However, the correctness of transformation estimates from such additional overlaps is not predictable. In the case of model reconstruction from a scan sequence, an end-to-end automatic framework is essential to abstract the complexities of registration and make the whole processing simple enough to be handled by non-experts. On the light of these arguments, following are the research questions addressed by this thesis.

1.6.1 Research Questions

- What is the role of pairwise registration and multiview registration stages in object reconstruction from 3D scans?
- How the development of a robust estimator can improve the pairwise registration stage?
- What is the potential of a robust algorithm in improving the multiview registration?
- What is the possibility of developing an automatic framework for registration of 3D point clouds in the case of object and scene reconstruction?

- What is the potential of a robust 3D transformation estimation algorithm for solving problems in other domains like satellite image registration?

1.7 Objectives of the thesis

Development of novel robust algorithms for the automatic registration of 3D point clouds is attempted in this thesis. The 3D object or scene to be modelled is scanned from multiple view points in partially overlapping fashion. Alternatively, the object can be moved in front of the camera in a turn-table set-up. The data captured is represented in the form of point clouds. With an overarching goal of developing an open source algorithmic framework for 3D object representation and scene reconstruction, this thesis work aims to develop automatic robust algorithms to solve the problems of

(i) pairwise alignment of the partially overlapping pair of point clouds.

(ii) global alignment of all the partially overlapping point clouds to form a complete 3D point cloud of the object or scene,
and

(iii) development of fast and accurate robust estimation algorithms for automatic registration of point clouds for object modelling

The registration problem addressed is mainly focussed on the alignment of the clouds captured for 3D modelling of objects. However, the approaches developed are general and are employable in other applications. As a test case, one of the algorithms developed is applied for automatic registration of different remote sensing satellite imagery. In addition to the object reconstruction, scene modelling using scan point clouds is also considered.

1.8 Structure of the thesis

The thesis is organised as follows. Chapter 2 presents a detailed literature survey on 3D image registration and discusses the sub-problems and current solutions to address these problems. Chapter 3 presents a novel algorithm to address the problem of pairwise registration. A robust estimation method is proposed, implemented and

evaluated on scans of various 3D models. The problem of multiview registration is investigated in Chapter 4. The challenges in multiview alignment and a novel algorithm to address the challenges are detailed. Chapter 5 details a novel framework for addressing the 3D registration problem of obtaining a complete 3D point cloud from partially overlapping sequence of scans of an object. The proposed pipeline is evaluated on different model scans varying in acquisition techniques, information available (colour and/or coordinate information), object shapes, object size etc. The case of object reconstruction and scene reconstruction are considered and evaluated to show the effectiveness of the framework. The robust estimation technique developed in Chapter 3 is general and can be applied to other similar estimation problems. In Chapter 6, this technique is adapted for automatic registration of remote sensing satellite images. Chapter 7 provides a summary of general conclusions of the thesis and give insights into the future directions that can be explored.

CHAPTER 2

Literature Review

Prelude: The extent of literature in 3D registration is vast and distributed across the application domains. This chapter presents a detailed review on the various approaches in literature to solve the 3D registration problem. The stages involved in solving 3D registration problem as well as the different solutions attempted in literature to solve the challenges involved are examined. Specifically, the stages involved in coarse registration pipeline are explained in detail.

The registration problem is cast as "*Orthogonal Procrustes*" problem in early literature (Hurley and Cattell, 1962). Procrustes was a rogue smith in Greek mythology who physically assaulted people by stretching them or amputating them so as to fit them to the size of his bed (Pomerleau et al., 2015). This throws light on the registration problem where a transformation estimate is to be fitted to the captured data.

Some of the early and seminal works in registration are discussed here. In one of the early works, Faugeras and Hebert (1986) identified that the fundamental constraint to be exploited in data analysis is that of rigidity. 3D primitives like planes and curves were extracted and matched to obtain the transformation. In an early survey, Besl (1988) observed the importance of identifying 3D objects using a network of geometric primitives like volume, surface, curve and point primitives. The paper touched on many geometric matching problems in computer vision using the geometric primitives. Over the next years, various solutions to solve the registration problem evolved and one of the seminal works was by Besl et al. (1992), which is popularly known as the "Iterative Closest Point (ICP)" algorithm. It is considered as the gold standard registration algorithm and it still finds further research interest (Han et al., 2016; Marani et al., 2016). Zhang (1994); Chen and Medioni (1992) also developed similar algorithms during this time. These approaches were not suited for partial overlapping scenario. The algorithms based on higher order primitives were mostly prone to noise in the datasets and the interest shifted to matching point sets directly. Early attempts to build 3D models from range images were by Dorai et al. (1996) and Curless and

Levoy (1996).

A number of surveys have studied the evolution of approaches in literature to solve the 3D registration problem (Salvi et al., 2007; Castellani and Bartoli, 2012; Tam et al., 2013; Gomes et al., 2014; Pomerleau et al., 2015; Dalley and Flynn, 2002; van Kaick et al., 2011; Mian et al., 2005). Closely related to 3D point cloud registration problem are the *range image registration* (Salvi et al., 2007), *model fitting* and *3D object recognition* problems (Guo et al., 2014). In the *range image registration* problem, range images are aligned instead of the 3D point clouds. The problem to be solved remains the same where one needs to estimate the 3D rigid body transformations between the 3D point clouds represented by the range images. In *model fitting*, the transformation between a partial view and a 3D model of an object is to identified. This is typically used in robotic object picking (Park et al., 2010) and object tracking. In *object recognition*, the 3D objects in a scene are to be identified from among a set of 3D models available in a library. The scene objects in the form of 3D point clouds are segmented and registered to the 3D models (point clouds) to estimate the pose (Chua and Jarvis, 1996). Registration becomes more challenging in cluttered environments. The pose or transformation estimation is a prerequisite to a multitude of applications like robotic grasping of objects.

Two major application scenarios of 3D point cloud registration are *scene reconstruction* and *object reconstruction* (Morell-Gimenez et al., 2014). The *scene reconstruction* refers to the robotic map building used to register the point sets obtained by the robot at different positions in order to get a map of its environment. A similar case is the autonomous vehicle navigation using 3D data. The *object reconstruction* is focussed on the alignment of object scans acquired from multiple views to a common coordinate system in order to form a 3D model. Other application areas of 3D registration include medical and surgery support (Markelj et al., 2012), 3D face recognition (Bagchi et al., 2016; Spreeuwers, 2011) etc. Each application scenario has different constraints and challenges to be satisfied (Pomerleau et al., 2015). For example, many of the medical image applications may have to deal with challenges like deformable objects (non-rigid registration), multi-modal sensing, high impact due to risk on human life etc. However they may have the advantages of precise sensors and controlled environments. The object reconstruction field often has the advantage of low real time requirements, manual or semi-automated environment etc. and face challenges like

loop closure and realistic model. The major challenge faced by robotic applications is having to deal with an unbounded object - the scene which may have uncontrolled dynamic elements. The main focus is the localization of the robot and the stability of the map.

3D registration techniques for object reconstruction can be broadly classified into *pairwise registration* and *multiview registration* (Salvi et al., 2007). The next sections review the existing works in these areas.

2.1 Pairwise Registration

Finding the transformation between two overlapping or partially overlapping point clouds is generally called *pairwise registration*. Let P and Q be the two point clouds to be registered and $\mathbf{T}$ be the transformation that aligns Q to P. Then the aim of pairwise registration is to estimate the rigid body transformation $\mathbf{T}_{min}$ that minimises the error function given by Eq.2.1

$$\mathbf{T}_{min} = \arg\min_{\mathbf{T}} \left(error \left(\mathbf{T} \left(Q \right), P \right) \right) \tag{2.1}$$

In geometric registration, the error function represents the distance between the corresponding points of the two datasets. Corresponding points are the points which belong to the same object point in the two datasets. The manner in which the corresponding points are detected varies with the kind of methods used for registration. Ideally, for a pair of corresponding points (p, q), $p \in P$ and $q \in Q$, the equation $p = \mathbf{T}(q)$ holds.

Pairwise registration methods can be broadly divided into two - *coarse registration* and *fine registration* (Campbell and Flynn, 2001). Coarse registration methods aim to find a rough initial alignment between two point clouds whereas fine registration methods refine the initial transformation estimate to form an accurate estimate. Point clouds which are arbitrarily transformed according to a rigid body motion can be aligned by coarse registration methods while fine alignment methods require an initial estimate of the actual transformation.

2.1.1 Coarse Registration Methods

Various methods to obtain the rough initial alignment of a pair of point clouds can be broadly classified into *manual, controlled* and *automatic* methods (Langming et al., 2014). Human intervention is required in manual registration methods and is often used in some commercial systems where the users are asked to select corresponding points from the point clouds or roto-translate them (Gomes et al., 2014). Many industrial systems make use of a carefully calibrated scanning environment where object and scanner motions are controlled. The motion can be measured from the scanning set up and this is used as an initial estimate. Some other controlled solutions use graphic markers placed over the object. The corresponding points are estimated using these markers (Tao and Jiyong, 2007). The automatic alignment methods make use of geometric or intensity information calculated from the point clouds to perform the registration.

Tam et al. (2013) presented an excellent survey on the various methods used for rigid and non-rigid registration. The existing methods can be classified based on the optimisation methods used. **Global deterministic methods** like *branch and bound* (Gelfand et al., 2005; Li and Hartley, 2007) and *game theory* based methods (Albarelli et al., 2011) have been used for coarse registration. **Stochastic optimisation** methods are another class of methods widely employed for obtaining a rough alignment. *Evolutionary* algorithms like Genetic Algorithm were employed by Meng et al. (2014); Lomonosov et al. (2006). Particle filtering, Hough transform and RANSAC are the other stochastic optimisation methods used in literature (Tam et al., 2013). **Constrained search** methods like Geometric hashing was used for identifying the best transformation by Gal and Cohen-Or (2006).

The initial registration algorithms can be classified according to the methods used for finding the correspondences. Estimating corresponding points by matching extracted 3D local features is a popular choice (Guo et al., 2016; Alexandre, 2012). Another strategy is to extract geometric primitives like lines, curves and planes from the point clouds and match them to find the correspondences (Yang and Zang, 2014; Stamos and Leordeanu, 2003). 4-Point congruent sets were extracted for finding the correspondences by Aiger et al. (2008).

Spectral matching techniques use algorithms like Principal Component Analysis (PCA) for obtaining a rough alignment. PCA based registration simply aligns the clouds based on the principal directions. Although this method is fast, it is not suitable in the case of low overlap and is very sensitive to outliers (Castellani and Bartoli, 2012; Salvi et al., 2007). **Graph based solutions** have been used for point-set matching. The point sets can be formulated as graphs and graph matching algorithms can employed for alignment (Gold and Rangarajan, 1996). Leordeanu and Hebert (2005) formulated the correspondences as graph nodes and a spectral clustering method was used for registration. Solutions based on **graphical models** and probabilistic formulations have also been employed to solve the registration problem (Caetano et al., 2004; Sun et al., 2012; Caetano and Caelli, 2006). Another recent approach to solve the registration problem is the use of **Gaussian mixture models**(GMM). There are two approaches for registration using GMMs. In one approach, one of the point sets is modelled as GMM and the other point set as realization of this GMM (Myronenko et al., 2010; Yang et al., 2014). In the second approach, both the point sets are modelled as GMMs and the distance between the distributions are minimised to obtain the registration (Jian and Vemuri, 2011; Campbell and Petersson, 2015).

A very popular technique for automatic point cloud registration is by **feature matching** (Gomes et al., 2014). This is justified by the presence of a large number of keypoint detection and feature extraction methods in literature (Tombari et al., 2013; Guo et al., 2016). Guo et al. (2014) presented an excellent summary of the methods which make use of feature matching to estimate the transformation. Many methods exploit the geometric information extracted from the point clouds for registration where others make use of intensity or colour information. A number of methods which use RGB-D images as inputs utilize a combination of geometric and colour information for registration. In this thesis, feature matching technique is used for automatic coarse registration and the steps involved are explained in detail here.

Coarse Alignment using Feature Matching

To estimate the rigid transformation between two point clouds, a minimum of three corresponding point pairs is required (Arun et al., 1987). The correspondences are found out by matching features extracted from input point clouds. The features must

be invariant to rigid transformation. From the set of correspondences, transformation is estimated using any of the optimisation methods discussed. Many works which review the various 3D feature descriptors are available in literature (Alexandre, 2012; Guo et al., 2016). The transformation estimation using feature matching is generally carried out in a number of stages.

Since the point clouds involved in object reconstruction are only partially overlapping, global features which represent the whole point cloud (Aldoma et al., 2011, 2012; Marton et al., 2011; Rusu et al., 2009b; Wohlkinger and Vincze, 2011; Rusu et al., 2010) cannot be used for matching. The solution is to make use of local features which are calculated by defining a small (local) neighbourhood around each point. The concept of local neighbourhood is explained here.

Local Neighbourhood: The concept of local neighbourhood is useful in multiple scenarios in 3D image processing. For a given point p, consider a set of points $Q = \{q_1, q_2, \ldots, q_n\}$ in the vicinity of p as its neighbourhood if it satisfies the relation given by Eq.2.2.

$$\|q_i - p\|_x \leq d_{max} \tag{2.2}$$

Here d_{max} represents the maximum allowable distance from the neighbours to the query point. And $\|\cdot\|_x$ is the L_x norm and Euclidean L_2 norm is the most commonly used norm. The local neighbourhood can alternately be defined by fixing the number of points in the vicinity of the query point p. The calculation of nearest neighbours being a computationally intensive task led to the development of many approximate nearest neighbour algorithms like Fast Library for Approximate Nearest Neighbours (FLANN) (Muja and Lowe, 2009). In order to calculate the nearest neighbours, the *k dimensional tree (kd-tree)* structure is used. The size of the neighbourhood depends on the task for which it is to be used.

The various steps involved in automatic transformation estimation using feature matching is depicted in Fig.2.1. Each of the stages involved and the current approaches are explained in detailed here.

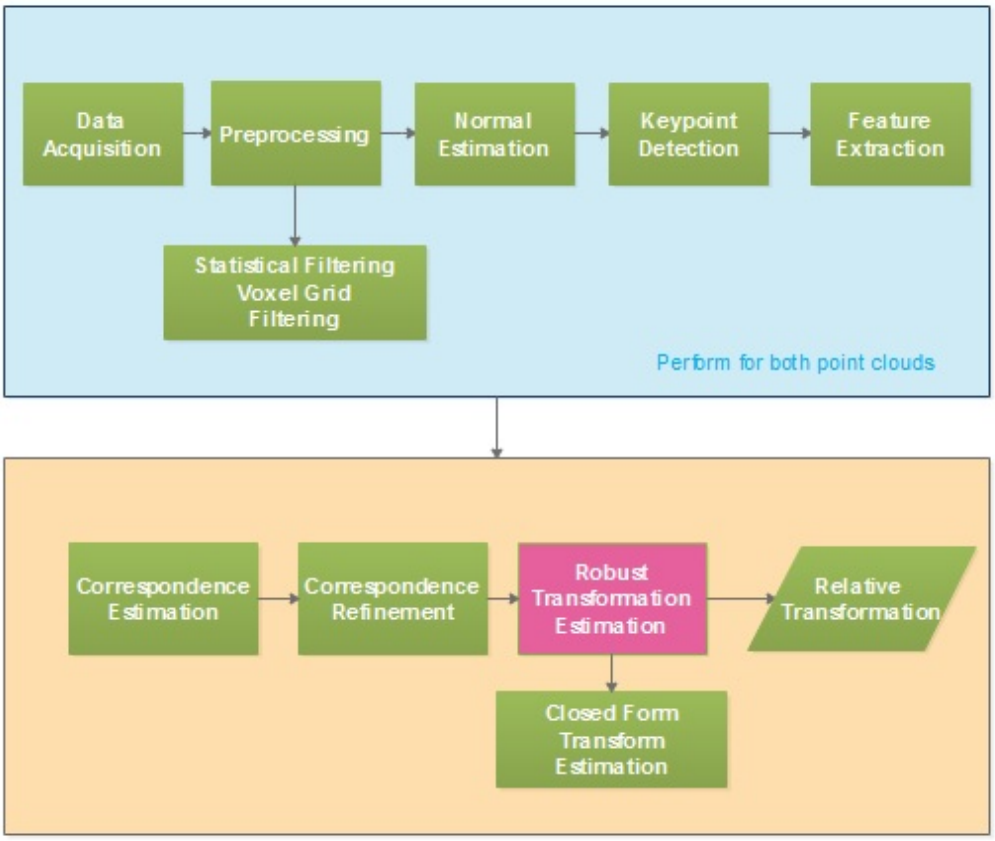


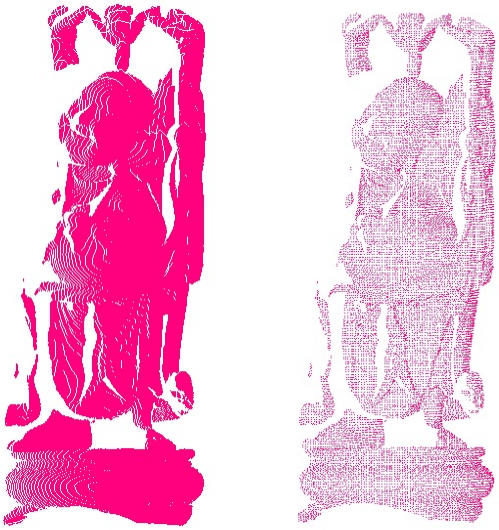
Figure 2.1: The coarse pairwise registration pipeline.

(i) **Preprocessing - Point cloud down-sampling and filtering**:

The point cloud data usually consists of tens of thousands of points and this leads to high computational requirement by the processing algorithms. The point cloud data can be down-sampled in order to reduce the execution time of the algorithms at the expense of accuracy. A voxelized grid approach was used by Rusu and Cousins (2011) to downsample the point clouds without compromising on the geometry of the underlying surface. In this approach, a 3D voxel grid (3D box) was assumed over the input point cloud. The points inside each voxel were then approximated with their centroid. This approach, called voxel grid filtering, is used to reduce the number of points to simplify further processing. Fig.2.2 shows a point cloud before and after filtering.

The point cloud data obtained from the sensors may be prone to varying point densities and sparse outliers near the edges. This may corrupt the surface normal and curvature estimation which may lead to wrong features. A filtering approach

is based on the statistical properties of a point in its local neighbourhood (Rusu and Cousins, 2011). The main idea is to filter out the points with sparse neighbourhood compared to the average neighbourhood density of the point cloud. The distribution of 'point to neighbour distances' for all points in the input point cloud is calculated. The mean distance from a point to its neighbours is computed. Approximating the resulting distribution to a Gaussian with a mean and a standard deviation, all points whose mean distances outside an interval (defined by the global mean and standard deviation) can be considered as outliers and can be trimmed from the dataset. The points having mean and standard deviation comparable to the global statistics are retained and the rest are rejected as outliers.



(a) Input Point Cloud (b) After filtering

Figure 2.2: Voxel grid filtering.

(ii) **Normal Estimation**:

The most basic information that describes the geometry of a point is its orientation in a given coordinate system, i.e its surface normal. Surface normals are important properties of a surface and are used to define the light sources and shading effects in computer graphics. A number of different normal estimation techniques have been proposed in literature. Klasing et al. (2009) and Dey et al. (2005) provided a comparison of different normal estimation techniques. A simple and fast method is to use first order 3D plane fitting (Berkmann and Caelli, 1994; Rusu, 2010). In this method, a small neighbourhood around each point is approximated

to a plane and the plane normal is assigned as the surface normal. This is done by fitting a plane to the neighbourhood by performing Principal Component Analysis (PCA) (Jolliffe, 2014). The eigen vector corresponding to the smallest eigen value is approximated as the surface normal at the point. Since the direction of normal is ambiguous, all normals are oriented consistently towards the view point (Rusu, 2010). Another useful surface measure at a point, which can be inferred using PCA, is the surface variation value defined by Eq.2.3

$$\sigma(p) = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \tag{2.3}$$

where λ_i are the eigen values of the covariance matrix calculated at the neighbourhood of the point with λ_0 being the smallest. The surface variation value is closely related to the curvature value at a point (Pauly et al., 2002).

(iii) **Keypoint Detection**:

Extracting 3D local features for establishing correspondences at every point of the input point cloud is computationally inefficient. Hence, feature extraction is usually carried out at a subset of the input points rather than at the complete set. The traditional way to obtain this subset is to down-sample the point clouds to reduce the computational cost at the expense of accuracy. However this method is inefficient to highlight the discriminative areas of the point cloud. Detecting a subset of points called keypoints can significantly reduce the computational expense as well as make feature matching more accurate. Mian et al. (2010) defined keypoints as the points which satisfy three constraints:

- high repeatability - keypoints should be repeatedly detectable under different conditions like view point changes, point density variations, clutter, sensor noise etc. (Tombari et al., 2013) (Distinctiveness and repeatability are regarded as the main traits of a 3D keypoint detector),
- definition of a unique coordinate system at the neighbourhood of the keypoint - most of the 3D descriptor algorithms make use of a unique coordinate system to describe the feature (Tombari et al., 2010a,b), and

- sufficient descriptive information in the neighbourhood - to define discriminative features.

The 3D keypoint detectors are mainly classified based on how they define the *scale* of the keypoints. Tombari et al. (2013) pointed out that the definition of a proper and repeatable scale in 3D data is different from the scale invariance notion in 2D. In the case of 2D, scale invariance notion helps in detecting keypoints invariantly across the differently sized images that occur due to the projection of 3D data onto a 2D plane. However, as 3D sensors provide metric data, selecting a characteristic scale helps to obtain distinctive features by defining a salient neighbourhood around a keypoint. Based on whether the scale is predefined or adaptively detected, the keypoint detectors are classified as **fixed-scale** and **adaptive scale** detectors. Fig.2.3 depicts the classification of 3D keypoint detectors proposed by Guo et al. (2014). The basic idea of identifying a keypoint is to select the *local extremum* of a *saliency measure* which determines its robustness and repeatability.

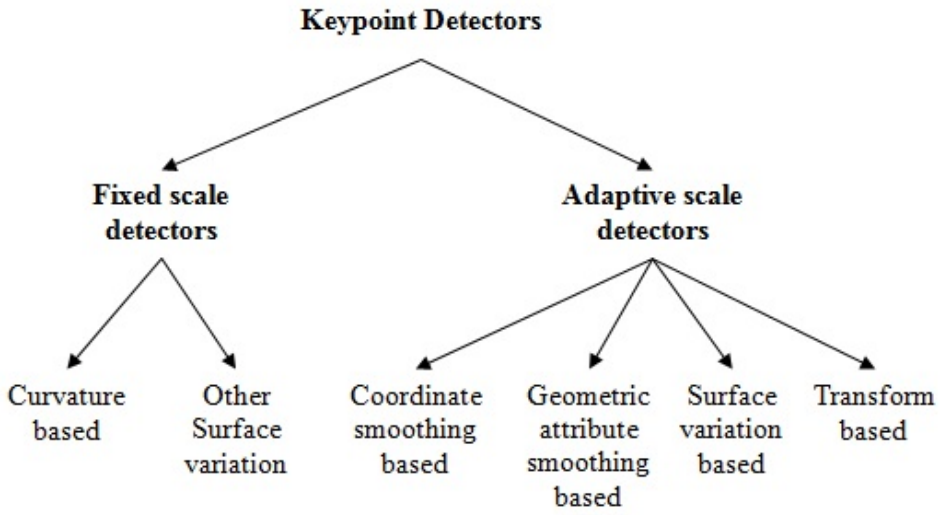


Figure 2.3: Taxonomy of 3D keypoint detectors.

(a) Fixed scale keypoint detectors: These detectors find keypoints based on a fixed neighbourhood, the size of which is passed as a parameter to the algorithm. The general structure of a fixed scale keypoint detector as in Tombari et al. (2013) is given in Fig.2.4 (dashed line indicates optional block). The

initial pruning step is optional and this helps in rejecting points that do not satisfy a quality criterion. A non-maximal suppression (NMS) is carried out in the neighbourhood to select the point with maximum saliency. The saliency measure can be *point-wise*, if it is calculated based on the characteristics at a point, or *region-wise*, if it is calculated based on the property of a neighbourhood around the point. For point-wise saliency measure, the input scale is used for defining the NMS region. For region-wise saliency measure, the input scale usually defines the support region for calculating saliency and a separate NMS scale may be defined, if required. The fixed scale keypoint detectors can further be classified based on the saliency measure used for identifying the keypoint as *curvature based methods* and *other surface variation based methods* (Guo et al., 2014).

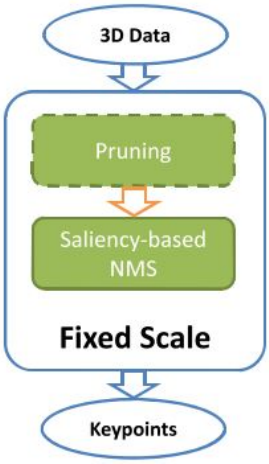


Figure 2.4: General structure of fixed scale 3D keypoint detector.

Various curvature measurements are used as saliency measures in *curvature based* methods. Gaussian curvatures and mean curvatures were used for selecting keypoints in 3D meshes by Mokhtarian et al. (2001). Simplex angles, directly related to the mean curvature, were used as saliency measure in Yamany and Farag (2002). Another method defined by Gal and Cohen-Or (2006) was based on a saliency grade. It was based on curvature values at the point and its neighbourhood. Chen and Bhanu (2007) presented a Local Surface Patches (LSP) method which was based on Shape Index value at a point, which was dependent on the maximum and minimum principal

curvatures at the vertex. This method performed poorly in presence of noise as it was based on second-order derivatives (Tombari et al., 2013).

The *other surface variation based* methods utilise geometric information other than curvature values for identifying the keypoints. The scatter matrix or covariance matrix calculated at the support region around a point carries much information on geometry. The smallest eigen value of the covariance matrix calculated in the support region is indicative of the surface variation and there are a few works which made use of this (Matei et al., 2006; Zhong, 2009). In Intrinsic Shape Signatures (ISS) (Zhong, 2009), a pruning stage which used the ratios of the successive eigen values to exclude points, was carried out before performing the NMS based on the saliency measure (smallest eigen value). Guo et al. (2013a) presented a method that worked on rotated and projected range images built from 3D point clouds. The ratio of eigen values calculated on the images were used as saliency measure. Tombari et al. (2013) proved that these methods had good repeatability and were computationally efficient. Extension of the popular 2D keypoint detector Harris (Harris and Stephens, 1988) to 3D was suggested by Głomb (2009) and Sipiran and Bustos (2011). Mian et al. (2010) presented a keypoint detector called Key Point Quality index (KPQ) which made use of the scatter matrix. The local surface was aligned with the principal directions (computed by PCA) of the scatter matrix and pruning was done on the ratio of lengths of data along the first two principal directions. Saliency measure was based on a combination of principal and Gaussian curvature values calculated on a smoothed surface fitted to the local support. Although the smoothing of the surface made it less affected by noise, the performance deteriorated in presence of clutter or partial data (Tombari et al., 2013).

The fixed scale keypoint detectors are fast owing to their calculation in a single region of support. However, the main drawbacks include the detection of less number of keypoints in low-curved regions and non-utilisation of the inherent scale information present in local geometric structures.

(b) Adaptive Scale Keypoint Detectors: The main stages involved in adaptive scale detection are highlighted in Fig.2.5 (dashed line indicates optional

block). A scale space can be defined in 3D or by embedding the 3D data in 2D and by defining it in 2D. A characteristic scale is assigned to each point by selecting the maximum value of a saliency measure along the scale space. Then keypoints are selected at the characteristic scale by NMS. The classification of adaptive scale detectors is depicted in Fig.2.3.

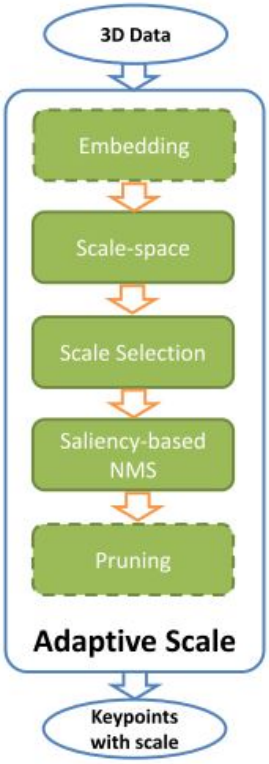


Figure 2.5: General structure of adaptive scale 3D keypoint detector.

In *coordinate smoothing based* methods, the scale space is constructed by successively smoothing the 3D coordinates similar to the idea of 2D scale space theory (Lindeberg, 1998). The scale space was constructed using Gaussian pyramids of the surface by Akagunduz and Ulusoy (2007). Connected components were extracted at the scale space and a weighted average of the curvature values of the voxels covered by the connected components was estimated to determine the location and scale of the keypoint. This detector was invariant to scale and rotation and exhibited better performance than fixed scale keypoints in the case of scale varying datasets. Li and Guskov (2005) created a scale space using smoothed versions of the surface by point set surface projection. A 2.5D extension to the 2D SIFT keypoints (Lowe, 2004) was presented by Lo and Siebert (2009). These

methods tried to extend the 2D scale space concept to 3D by replacing intensities with pixel coordinates. However, this approach could not preserve the extrinsic geometry and topology of a 3D shape, leading to the violation of causality property (Novatnack et al., 2006).

The geometric attributes are smoothed to create the scale space in *geometric attribute smoothing based* methods. Hence causality property is not violated (Guo et al., 2014). Surface density map, normal field, Gaussian curvatures, geodesic metrics etc are some of the geometric attributes which are smoothed to find the keypoints (Novatnack et al., 2006; Flint et al., 2007; Zou et al., 2008, 2009).

Surface variation based methods are developed based on the assumption that the neighbourhood size can be considered as a scale parameter and increasing the neighbourhood size is equivalent to creating a scale space by smoothing (Pauly et al., 2003). The extremum on the surface variation space was selected as keypoints by Pauly et al. (2003). The standard deviation of shape index values and curvedness in different neighbourhood spaces were used as saliency measures by Ho and Gibbins (2008, 2009). Laplace Beltrami Scale Space was created by Unnikrishnan and Hebert (2008) by computing an invariant at different scales. Even though their method was robust to noise, its repeatability was relatively low and detected only very few keypoints (Tombari et al., 2013).

Transform based methods transform the point from spatial domain to some other domain and detect the keypoints in the transformed space. Hu and Hua (2009) transformed the shape to Laplace Beltrami spectral domain to detect the keypoints. The 3D shape was considered as a compact Riemannian manifold and the heat diffusion process over it was shown by Sun et al. (2009). They detected the keypoints as the local maxima by restricting the heat kernel to the temporal domain. This method captured the extremities of long protrusions on the surface and the keypoints were highly stable and robust to noise. However, heavy computational requirement was a drawback.

Keypoint detection by geometric methods has been discussed so far. A recent trend in extracting keypoints is by learning keypoints from available data using

machine learning techniques. Learning keypoints in the case of 2D images was suggested by Strecha et al. (2009); Verdie et al. (2015) and Rosten et al. (2010). An algorithm for learning keypoints from a depth image was presented in Holzer et al. (2012). Random forests were used for training and the interest points were learnt to exhibit high repeatability. Keypoint detection in face images was performed using a learning approach in Creusot et al. (2013). Linear discriminant analysis and AdaBoost (Viola and Jones, 2004) algorithms were used in training. Teran and Mordohai (2014) formulated 3D interest point detection as a binary classification problem. It had the ability to adapt to task specific requirements in selecting the keypoints.

A recent evaluation of 3D keypoint detectors via human generated ground truth was presented by Dutagaci et al. (2012). Benchmark surveys on detectors and descriptors were presented by Bronstein et al. (2010) and Boyer et al. (2011). A good comparison on the performance of the keypoint detectors was done by Tombari et al. (2013). Based on the keypoints evaluated, Keypoint Quality Index (KPQ) (Mian et al., 2010) was identified as a good technique based on repeatability, distinctiveness and robustness to noise. Intrinsic Shape Signature (ISS) (Zhong, 2009) was shown as an effective method which possess a good trade-off between absolute and relative repeatability. The important benefit of ISS was that it was highly efficient. Mesh Difference of Gaussian (MeshDoG) and the adaptive scale version of KPQ also produced good results.

The open source library PCL (Rusu and Cousins, 2011) implements a number of 3D keypoint detectors and a comparison of these detectors is available in Filipe and Alexandre (2014). The considered keypoints included ISS (Zhong, 2009), 3D SIFT (Flint et al., 2007), 3D extension of SUSAN (Smith and Brady, 1997), 3D Harris detector which used different saliency measures based on Harris 2D (Harris and Stephens, 1988), Kanade-Lucas-Tomasi detector (Tomasi and Kanade, 1991), Curvature, Lowe and Noble etc. The survey concluded that SIFT3D and ISS showed the best repeatability and ISS was more invariant to the transformations. Based on the recent evaluations (Tombari et al., 2013; Filipe and Alexandre, 2014; Salti et al., 2012; Guo et al., 2016), *ISS* (Zhong, 2009) is selected for keypoint detection in this work.

(iv) **Feature Extraction:**

The 3D features or descriptors are compact representations of a 3D scan view and are used for detecting similarities with other 3D scan views. They can be classified as *point-wise, local* or *global* descriptors based on the region of support. The point-wise descriptors encode compact and discriminative properties like curvature of a point. Even though simple and efficient, they are often less descriptive and are susceptible to noise. Local descriptors encode the information present in a small neighbourhood centred at the point in order to describe it. They are often descriptive enough for the various applications like registration, segmentation, object recognition in clutter and occlusion etc. The size of the neighbourhood is an important factor in deciding the trade-off between descriptiveness and robustness to clutter and occlusion. Global descriptors define the object surface as a whole and are used for tasks like 3D object recognition and retrieval (Aldoma et al., 2011, 2012; Marton et al., 2011; Rusu et al., 2009b; Wohlkinger and Vincze, 2011; Rusu et al., 2010). For the problem of 3D registration for model reconstruction, the 3D scans usually have only partial overlap. The descriptors suited for this task are the local or regional descriptors. Descriptiveness and robustness of a feature descriptor are the crucial factors for accurate feature matching (Bariya et al., 2012; Guo et al., 2013a) and hence registration. This is because the accuracy of feature matching directly influences the quality of the estimated transformation and the computational requirements for further stages of registration. Moreover, the feature descriptors should be robust to a set of factors like noise, varying mesh resolution, occlusion, rigid transformations etc. (Guo et al., 2013a).

The descriptors employed for feature matching for registration should exhibit the property of rigid transformation invariance and should be robust to noise and occlusion. Rigid transformation invariance property is achieved by most feature descriptor algorithms by identifying a Local Reference Axis/Frame (LRA/LRF) in the local neighbourhood. The Local Reference Axis considered usually is the surface normal at the point (Johnson and Hebert, 1999). The LRF is obtained by defining three orthogonal unit vectors defined upon the local support. The definition of a unique and repeatable LRF is challenging (Tombari et al., 2010b).

Different taxonomies for the 3D shape descriptors are available in literature. Lara López et al. (2016) presented different taxonomies and an evaluation of different 3D descriptors. The taxonomy presented by Guo et al. (2014) and Tombari et al. (2010b) is considered here. 3D feature descriptors can be broadly classified into three based on the approaches employed for construction as *Signature based, Histogram based* and *Transform based* methods as shown in Fig.2.6.

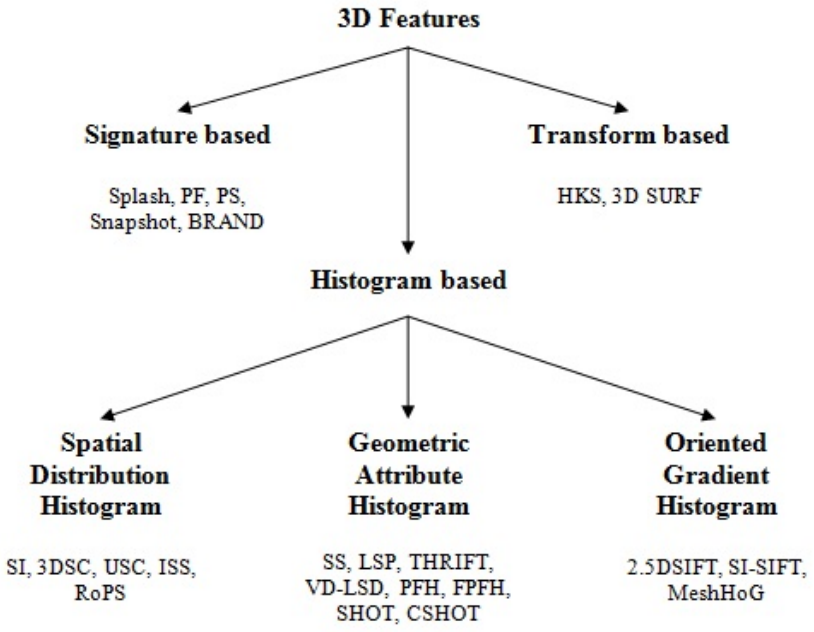


Figure 2.6: Taxonomy of 3D feature descriptors.

(a) Signature Based Methods: The descriptors are defined by encoding one or more geometric measures computed individually at each point of a subset of the neighbourhood (Tombari et al., 2010b). One of the early attempts in describing a local 3D surface was by Stein and Medioni (1992). The descriptor named "*Splash*" was based on a LRF defined by the normal and the tangent plane at the point p. The angular distances between the normals at the point p and the neighbouring points in a circular slice with geodesic radius r were encoded into a 3D vector, which described a curve in 3D space. The curvatures and torsion angles of a fitted segment were encoded to form the descriptor. Another early descriptor is the "*Point Signature (PS)*" by Chua and Jarvis (1997). The surface was intersected with a sphere centred at p to obtain a 3D space curve C and a surface normal n was obtained by plane fitting. The fitted plane was translated along n to p and the points were

projected on it to obtain another curve C'. A reference direction n_2 was defined as the unit vector from p to the projected point on C' which gave the largest positive distance. Then the points sampled along C were then represented using the signed distance d between the point and its correspondence on C', and the clockwise rotation angle θ from n_2. The limitations include sensitivity to mesh resolution, non-uniqueness of reference direction and the difficulty in computing the intersection. "*Point Fingerprint*" descriptor defined by Sun and Abidi (2001) formed a pattern similar to human fingerprint and it encoded the information of the normal variation along geodesic circles centred at the point. The LRF was defined by the normal vector and tangent plane at the point. The geodesic circles projected on to the tangent plane form contours, which, together with the radius and normal variations, formed the descriptor.

Malassiotis et al. (2007) and Mian et al. (2010) formed the LRF by Eigen decomposition of the scatter matrix. Malassiotis et al. (2007) formed the "*Snapshot*" descriptor by considering a pinhole camera at a distance d along the z-axis and by projecting the points on to the image plane. The distances of points from the image plane were encoded to form the descriptor. Snapshot was very robust to occlusion. Mian et al. (2010) also used the depth values compressed using PCA to form the feature vector. Castellani et al. (2008) extracted the various attributes of a point like saliency, surface normal variation, minimum-maximum curvature etc. along a defined spiral path and used a discrete time Hidden Markov Model (HMM) to encode it to form the descriptor. This descriptor was robust to rotation, non-uniform sampling and varying mesh resolution (Guo et al., 2014). Novatnack and Nishino (2008) formed the "*Exponential Map descriptor*" by encoding the surface normals in a 2D geodesic polar coordinate frame. Masuda (2009) formed the "*Log-polar Height Map*" descriptor by defining a log-polar coordinate frame on the tangent plane at p and encoding the height map of the projected points. Steder et al. (2010, 2011) defined the "*Normal Aligned Radial Feature (NARF)*" descriptor by encoding the pixel variation in a star pattern overlaid onto a local range image patch aligned with the normal. The descriptor was aligned to an extracted unique direction to make it rotation

invariant. This descriptor made use of the border information in range images and it outperformed the spin images in object matching experiments. Do Nascimento et al. (2013) defined the "*Binary Robust Appearance and Normal Descriptor (BRAND)*" for RGB-D images using both appearance and geometric information. The BRAND outperformed SIFT, SURF, spin image and CSHOT (Colour-SHOT) in terms of matching precision and robustness (Guo et al., 2014).

(b) Histogram Based Methods: These methods form the descriptor by accumulating geometric or topological measurements (e.g., point numbers, mesh areas) into histograms according to a specific domain (e.g., point coordinates, geometric attributes) (Tombari et al., 2010b). The methods can be broadly classified into spatial distribution histogram (SDH), geometric attribute histogram (GAH) and oriented gradient histogram (OGH) based methods (Guo et al., 2014).

In SDH methods, the spatial distribution or the coordinate information of points in the local neighbourhood is accumulated in histograms to form the descriptor where the coordinate is formed by defining an LRF/axis at the keypoint. One of the successful descriptors in this category is the "*Spin-image (SI)*" descriptor formed by Johnson and Hebert (1997, 1999) and is used as a benchmark for 3D descriptors (Tombari et al., 2010b). Several variants of the descriptor are also present in literature (Dinh and Kropac, 2006; Carmichael et al., 1999; Assfalg et al., 2007; Ruiz-Correa et al., 2001; Pasqualotto et al., 2013; Darom and Keller, 2012; Guo et al., 2013b) that tried to address its limitations like varying mesh resolutions, non-uniform sampling, limited descriptiveness etc. Frome et al. (2004) formed the "*3D Shape Context (3DSC)* " which outperformed Spin Image in the case of recognition in cluttered scenes (Guo et al., 2014). To deal with the reference axis ambiguity of 3D Shape Context, a descriptor named "*Unique Shape Context (USC)*" with a unique repeatable LRF was defined by Tombari et al. (2010a), thus reducing memory requirements. Zhong (2009) proposed the "*Intrinsic Shape Signature (ISS)*" by constructing an LRF by the Eigen decomposition of the scatter matrix. "*Rotational Projection Statistics (RoPS)*" proposed by Guo et al. (2013a) is another successful descriptor in this cate-

gory.

In GDH methods, the geometric attributes of the points in a local neighbourhood are binned to form the histogram. The "*Surface Signature (SS)*" formed by accumulating simplex angles by Yamany and Farag (2002) was proved to be more descriptive than spin images, splash and point signatures. A 2D histogram with shape index values and angle between normals in the neighbourhood as the two axes, was formed into the "*Local surface patches (LSP)*" descriptor by Chen and Bhanu (2007). The "*THRIFT*" descriptor was a 1D histogram formed from the surface normal angles in a local neighbourhood. Two normals at a point were defined using two differently sized windows and the angle between them formed the surface normal angle. A generalized descriptor "*Variable Dimensional Local Shape Descriptor (VD-LSD)*" that subsumes a large number of descriptors was defined by Taati et al. (2007) where a feature selection algorithm was used to select a subset of the features which were then binned to form the histogram. Rusu et al. (2008, 2009a) proposed two descriptors "*Persistent Feature Histogram (PFH)*" and "*Fast Persistent Feature Histogram (FPFH)*". The distance between points and angle between normals were extracted and binned to form the PFH descriptor. FPFH was a fast version of the PFH and retained most of the discriminative power of PFH. Tombari et al. (2010b) defined the "*Signature of Histograms of OrienTations (SHOT)*" descriptor which was a hybrid of the signature and histogram based methods. This descriptor was descriptive as well as robust to noise, however it was susceptible to variation in mesh densities. The authors also developed another descriptor called "*CSHOT*" which included colour information also (Tombari et al., 2011).

In the OGH methods, the oriented gradients in the local neighbourhood of a keypoint were accumulated to form histograms. The descriptors by Hua et al. (2008); Lo and Siebert (2009) and Darom and Keller (2012) formed 3D descriptors adapting ideas from the 2D SIFT. Lo and Siebert (2009) divided the local surface into nine elliptical subregions and generated two histograms using shape index values and gradient of orientations. All the histograms were concatenated to form the "*2.5D SIFT*" descriptor. "*Mesh-HOG*" defined by Zaharescu et al. (2009) is a popular descriptor in this class.

A 3D descriptor proposed by Bayramoglu and Alatan (2010) based on the concept of Shape Index and SIFT called "*SI-SIFT*," also belongs to this class of descriptors. SIFT algorithm wss applied on to the extracted Shape Index image to form the descriptor.

(c) Transform Based Methods: In these methods, the 3D image is transformed into another domain and the neighbourhood is defined by encoding the information in the transformed domain. Hu and Hua (2009) used a histogram of the Laplace Beltrami spectrum of the local surface to as the feature descriptor. This descriptor had salient spectral geometric features invariant to spatial transforms such as translation, rotation, and scaling. Sun et al. (2009) proposed "*Heat Kernel Signature (HKS)*," which was a restriction of the heat kernel to the time domain. It captured all information contained in the heat kernel and preserved all the shape information up to isometric transformations. An extension to the 2D SURF (Bay et al., 2006) descriptor was proposed as "*3D SURF*" by Knopp et al. (2010b). The mesh was voxelized and Haar wavelet transform was applied to the voxels. Haar wavelet responses were used to define the LRF and the local neighbourhood volume was divided into bins. The responses were grouped into a vector and the binned vectors formed the 3D SURF descriptor.

The histogram based methods are relatively robust to noise because of the inherent smoothing in histogram calculation. The extend of the neighbourhood considered is an important factor in deciding the descriptiveness of the feature descriptor. The performances of different 3D feature descriptor algorithms were evaluated in many works. A benchmark comparison of feature detectors and descriptors were presented by Bronstein et al. (2010); Boyer et al. (2011). Alexandre (2012) compared various features available in the PCL library (Rusu and Cousins, 2011) for object recognition task. They opined that the descriptors which use both colour and geometry information provided the best results when compared to methods using any one of these. Kim and Hilton (2013) presented a comparison of features for multimodal registration. Restrepo and Mundy (2012) evaluated local shape descriptors- FPFH, SHOT, SI, 3DSC - for the task of object classification. In a latest work, Lara López et al. (2016) presented a good comparison of feature descriptors belonging to six different taxonomies.

Guo et al. (2016) presented a comprehensive evaluation of the performance of 3D descriptors for 3D object recognition, 3D shape retrieval, and 3D modelling tasks in eight different datasets. They stated that FPFH was the best choice for time-critical applications (e.g., real-time systems) on point clouds with a small number of points and it provided a good balance between feature matching accuracy and computational efficiency. When the number of points was high, SHOT provided a good performance in terms of descriptiveness as well as computational efficiency. When memory footprint was crucial, FPFH was the best choice as the length of each descriptor was less. For more challenging registration scenarios RoPS was a good choice as it was more descriptive. It also provided satisfactory performance in all kinds of datasets. When the number of models were more, TriSi (Guo et al., 2013b), USC and 3DSC exhibited the best scalability. However, their storage cost was high. One of the major considerations was the quality of data as most of the descriptors exhibited low performance with data from low-resolution scanners. The spatial histogram based descriptors (3DSC, USC, SI, TriSI) were more robust to noise whereas the geometric attribute histogram based descriptors were more sensitive to noise. Salti et al. (2012) evaluated the combination of various keypoint detectors and feature descriptors for the tasks of registration and object recognition. They concluded that the pairs ISS/PS, ISS/3DSC, ISS/KPQ were the best choices for registration. In this thesis, ISS/3DSC combination is used for object reconstruction.

(v) **Correspondence Estimation by Feature Matching:**

Once 3D features are extracted at the keypoints, the correspondences can be estimated by matching the features in the two point clouds. Given a set of features in the target point cloud $F^1 = \{f_1^1, f_2^1, \ldots, f_n^1\}$ and source point cloud $F^2 = \{f_1^2, f_2^2, \ldots, f_m^2\}$, one needs to estimate the closest match of each source feature in the set of target features. Since this is a nearest neighbour estimation in high dimensional space, kd-trees (Friedman et al., 1977) are employed for speed. Feature kd-trees formed with k equal to the dimension of the feature are employed. For each source feature, its nearest neighbour in the target feature tree is found out. The respective keypoints are saved as corresponding point pairs in

the form $C = \{(p_1, q_1, d_1), (p_2, q_2, d_2), \ldots, (p_m, q_m, d_m)\}$. Here m is the number of keypoints in the source point cloud. (p_i, q_i) indicates a pair of corresponding points with p_i being the source keypoint, q_i the target keypoint whose feature is the closest to that of p_i and d_i, the distance between them.

(vi) **Correspondence Refinement:**

The set of corresponding point pairs obtained may contain many wrong matches due to partial overlap, symmetry of the surface, noise etc. Hence correspondence refinement algorithms are required to deal with the outliers in the set of correspondences. A number of correspondence refinement strategies are being used. The Table 2.1 summarises some of the popular correspondence refinement methods (Rusu and Cousins, 2011). The nearest neighbours were considered as corresponding points in some methods (Besl et al., 1992). In such cases, the correspondences with distance greater than a certain threshold were rejected as these may attribute to non-overlapping regions. Another option is to reject correspondences if the corresponding distance is greater than the median of the distances of all the correspondences. In one-to-one matching strategy, the target matches are examined and if one target point corresponds to more than one source point, then all the duplicates except for the one with lowest distance are eliminated. Another strategy is to find the matches from source to target as well as from target to source and retain only those correspondences which satisfy the correspondence in both cases. The surface normals at the points are preserved after a rigid body transformation. This rationalizes the rejection strategy based on the angle between the normals. The correspondences are analysed by comparing the angle between their normals and are rejected if this crosses a certain threshold.

Chetverikov et al. (2002) suggested to retain only the best k correspondences. The best correspondences were obtained by sorting them in the order of their distances and by selecting only k correspondences. The parameter k was obtained based on the range of overlap (Chetverikov et al., 2002). Phillips et al. (2007) proposed to retain only a fraction ($k\%$) of the correspondences by suggesting a method to estimate the k value. The corresponding points in a rigid transformation are bound by the geometric constraints of the transformation. A rigid body transfor-

Table 2.1: Correspondence refinement methods.

Correspondence Refinement Method	**Explanation**
Distance	Threshold by distance between correspondences
Median Distance	Threshold by median distance between correspondences
One to One match	Duplicate matches are eliminated
Reciprocal	Accept only reciprocal correspondences
Angle between normals	Threshold by the angle between normals at corresponding points
Trimmed Correspondences	Retains 'k' best correspondences (Chetverikov et al., 2002)
Best k% Correspondences	Retains 'k' percentage of correspondences with the least distances. (Phillips et al., 2007)
Polygon Based	Using rigid body geometric constraints (Buch et al., 2013)

mation should preserve the distances between different points of an object. This constraint was made use of in the correspondence rejection method suggested by Buch et al. (2013). In this work, this correspondence rejection method is used as it provides the best set of correspondences compared to other methods.

(vii) **Closed form solutions for Transformation Estimation:**

If a set of at least three correspondences is available, least squares methods can be used to estimate the rigid transformation. The solutions in literature vary based on the representations of the transformation. Using transformation matrix formulation, Arun et al. (1987) presented a SVD based least squares approach for transformation estimation from a set of corresponding points. This solution might give a reflection instead of rotation when the data is corrupted. Umeyama (1991) suggested a modified solution which provided a correct transformation matrix even in presence of corrupted data. Horn (1987) laid out a solution based on unit quaternions for rotation estimation. Walker et al. (1991) presented a solution for

transformation estimation using dual quaternions. A comparison of four different transformation estimation methods was given by Lorusso et al. (1995). Analytic solutions for 3D rotation fitting was discussed by Kanatani (1994). Recent approaches include transformation estimation from plane correspondences instead of point correspondences (Khoshelham, 2016) and an iterative solution instead of a closed form solution by Hersch et al. (2012). The transformation estimation method proposed by Umeyama (1991) is employed in this work.

(viii) **Transformation Estimation:**

The closed form solutions presented for transformation estimation will fail considerably when large number of outliers is present in the set of correspondences. Outliers refer to the correspondences which are at a distance greater than a threshold after the transformation is applied. The closed form solutions presented above will lead to corrupted results in this case. Hence robust estimation methods like RANdom SAmple Consensus (RANSAC) are required for finding out the transformation from a set of correspondences with outliers (Taati et al., 2007; Taati and Greenspan, 2011). Other methods make use of the geometric constraints to arrive at a good correspondence set (Gelfand et al., 2005). Pose clustering is another commonly used method for transformation estimation especially in the case of object recognition problem (Mian et al., 2010; Drost et al., 2010; Guo et al., 2013b,a). Generalized Hough transform which performs voting in parametric space offers another solution (Tombari and Stefano, 2012; Knopp et al., 2010a). Geometric Hashing is also used to find the transformation from a set of correspondences (Lamdan and Wolfson, 1988).

The robust motion estimation using RANSAC depends on a number of parameters like the number of data points, inliers to the model, size of the sampling set etc. Convergence of RANSAC can be very slow in the case of large number of outliers. This thesis presents a novel algorithm which provides accurate results compared to RANSAC with lesser computational expense in Chapter 3.

2.1.2 Fine Registration Methods

The fine registration methods aim to align two point clouds when an initial estimate of transformation between them is known. Most of the methods in this category make use of a local optimisation method to minimise an objective function. Gradient descent, Newton, (damped) Gauss-Newton, quasi-Newton, Levenberg-Marquardt (L-M) and Expectation - Maximization are the common optimisation techniques used (Tam et al., 2013). The two major algorithms in solving the fine registration problem are Iterative Closest Point (ICP) and Normal Distributions Transform (NDT). These and the related methods are briefly described below.

Iterative Closest Point (ICP)

One of the benchmark algorithms in the area of 3D registration is the Iterative Closest Point (ICP) algorithm proposed by Besl et al. (1992). (Closely related to ICP are the methods proposed by Chen and Medioni (1992) and Zhang (1994)). In these methods, the set of correspondences between point clouds are estimated by identifying the distance between them. The popular distance measures used to estimate the corresponding points are *point-to-point* and *point-to-plane* distances. Once the corresponding points are obtained, the transformation which minimizes the total distance between the point clouds is estimated using an optimization algorithm. This entire process is repeated until an accurate transformation estimate is obtained. The basic steps involved in ICP are included below.

(i) Let P be the target data set with N_p points $\{p_i\}$ and Q be the source dataset to be registered to P containing N_q points $\{q_i\}$.

(ii) Initialise transformation parameters (by initial registration or by using robotic odometry data). Let T_0 be the initial transformation which registers Q to P. Apply the transformation to Q to obtain Q_0 i.e. points $\{q_i'\}$. (Let d_k be the distance between corresponding point pairs in each iteration k and $d_0 = 0$)

(iii) Compute the closest points (in P) of the points $\{q_i'\}$ to form correspondences $C = (q_i', p_i')$ where p_i' are the corresponding points in P

(iv) Estimate the transformation Δt using the correspondences C using a Least Squares

algorithm

(v) Apply the transformation $T_{k+1} = \Delta t T_k$ to current Q_k form Q_{k+1}

(vi) Terminate when $d_k - d_{k-1} < \tau$. Else repeat steps (iii) to (v).

There are three major limitations to the ICP algorithm. The first limitation is that it is not directly extendible to multiple views, because of the absence of overlap between the views. (This is handled by using multiview registration algorithms, which is the focus of section 2.2). Secondly, it provides only a local solution. Hence it has the chance of falling into a local minimum and requires a good initialization for the algorithm to succeed. Coarse registration techniques are used to initialize the solution to deal with this problem. The next limitation is the implicit assumption that one data set is a subset of the other (complete overlap). This is not the case in many application scenarios especially in partial view matching. In order to deal with various limitations, modifications to the basic algorithm have been proposed in literature (Rusinkiewicz and Levoy, 2001; Pomerleau et al., 2013, 2015). The variations can be classified based on the various steps involved in ICP (Rusinkiewicz and Levoy, 2001).

(i) **Point Selection**: Different strategies exist for selecting points from the two datasets such as

- using all the points (Besl et al., 1992)
- uniform subsampling of points (Turk and Levoy, 1994)
- random sampling at each iteration (Masuda et al., 1996)
- selecting points with high intensity gradient (Weik, 1997)
- selecting points such that distribution of normals is as large as possible (Rusinkiewicz and Levoy, 2001), and
- geometrically stable sampling (Gelfand et al., 2003)

(ii) **Correspondence Estimation**: Several methods exist to find the corresponding points such as

- finding the closest point in the other mesh (Besl et al., 1992) **(point-to-point)**

- finding the intersection of the source point normal with the tangent plane at the target surface (Chen and Medioni, 1992) **(point-to-plane)**
- projecting the source point onto the destination mesh, from the point of view of the destination mesh's range camera (Neugebauer, 1997a; Blais and Levine, 1995)
- projecting the source point onto the destination mesh, then perform a search in the destination range image using a metric (Dorai et al., 1998; Benjemaa and Schmitt, 1999), and
- matching points based on colour (Godin et al., 1994) or angle between normals (Pulli, 1999)

(iii) **Point-pair weighting**: Different weighting strategies include

- constant weight
- lower or zero weights to pairs with large point-to-point distances (Godin et al., 1994)
- weighting based on compatibility of normals or colours (Godin et al., 1994), and
- weighting based on scanner noise effect

(iv) **Pair Rejection**: Different methods to refine the correspondences by removing the wrong ones encompass

- rejection based on distance, median distance
- rejection of worst $n\%$ pairs (Pulli, 1999)
- rejection of pairs at a distance greater than a multiple of standard deviation of distances (Masuda et al., 1996)
- rejection based on geometric constraints (Dorai et al., 1998; Buch et al., 2013)
- rejection of boundary points (Turk and Levoy, 1994), and
- rejection of pairs other than the best 'k' pairs (Chetverikov et al., 2002)

(v) **Error Metric**: The various error metrics for minimisation in literature include

- sum of point-to-point distances: sum of Euclidean distance between corresponding points (Besl et al., 1992). Closed form solutions are available (Arun et al., 1987; Horn, 1987; Walker et al., 1991; Umeyama, 1991)
- sum of point-to-plane distances: sum of squared distance from source point along its normal to the tangent plane at the destination surface (Chen and Medioni, 1992). Non-linear minimization techniques are used.
- Fractional Root Mean Squared Distance (FRMSD): root mean squared distance between a fraction of the corresponding points (Phillips et al., 2007), and
- objective function based on point-to-point distance as well as difference between surface normals and tangents (Serafin and Grisetti, 2014)

Apart from the various methods based on the above categories, interest in ICP has triggered many modified versions. Segal et al. (2009) presented a generalized probabilistic frame work which combined the point-to-point and point-to-plane ICP to form an algorithm called Generalized ICP. Many modified algorithms for registering coloured point clouds evolved due to the advent of sensors which provide RGB information also (Druon et al., 2006; Joung et al., 2009). ICP algorithm for affine transformation was proposed by Du et al. (2010). Yang et al. (2013) combined a branch and bound technique with ICP to form a global ICP algorithm called Go-ICP. Iterative Clustered Closest Points (ICCP) (Lu et al., 2014) was another modification to ICP in which the correspondences were clustered based on the distance between the points and registered separately. The resulting motions were then averaged using unit quaternions to obtain the final transformation. Liu et al. (2015) introduced a regularization based iterative re-weighting method for evaluating point matches for ICP. Sparse ICP (Bouaziz et al., 2013) and Efficient Sparse ICP (Mavridis et al., 2015) formulated the problem using sparsity-inducing norms, significantly improving the resilience of the registration process to large amounts of noise and outliers. Han et al. (2016) presented an enhanced octree based ICP for the registration of large-scale 3D environment models which heuristically avoided local minima. Marani et al. (2016) presented a modified ICP by introducing a deletion mask concept to delete the points

that can induce erroneous registrations.

Normal Distributions Transform (NDT)

The ICP based methods rely on the explicit calculation of the corresponding points. Biber and Strasser (2003) introduced a method for laser scan matching which did not need to calculate the correspondences explicitly. This method was significantly different from ICP and was both accurate and fast. NDT method is explained here.

A set of points (model scan) is represented using a new representation called Normal Distribution Transform (NDT). The 2D space around the sensor is subdivided regularly into cells of constant size. Then in each cell (with at least 3 points), a normal distribution is assumed. The probability of measuring a sample at a point contained in the cell is modelled by a normal distribution whose parameters (mean and covariance) are calculated from the data points falling inside it. Thus a grid of normal distributions forms the NDT. The advantage of this representation is the availability of a piecewise continuous and differentiable description of the 2D plane. Overlapping grids are considered to minimize the effect of discretization and the probability in all cells are summed up to obtain the probability at a point. Once the model scan is represented using NDT, the initial transformation (from coarse registration or odometry) is applied to the source scan. For each point in the transformed source scan, the corresponding normal distribution in the first scan is found out by using a table look up. The objective function to be optimised is defined by evaluating the distribution for each mapped point and summing the result. The optimal parameters for the transformation are obtained by optimizing the objective function using Newton's algorithm. The NDT matching is summarised as follows:

(i) build the NDT of the first scan

  (a) map the points to grids, and for each grid with at least 3 points ($x_{i=1...n}$),

  (b) calculate mean $q = \frac{1}{n}\sum_i x_i$

  (c) calculate covariance matrix $\Sigma = \frac{1}{n-1}\sum_i (x_i - q)(x_i - q)^t$.
  The probability of measuring a sample at point x inside this cell is given by the normal distribution $N(q, \Sigma)$:
  $p(x) = Cexp\left(-\frac{1}{2}(x_i - q)^t \Sigma^{-1}(x_i - q)\right)$ with C being a constant $= 1$

(ii) initialise transformation parameters (by initial registration or by using robotic odometry data)

(iii) for each point in the second scan, estimate the transformed point using the transformation parameters.

(iv) determine the corresponding normal distributions for each point in second scan

(v) the distribution is evaluated for each point and the results are summed to form the score of the parameters p
$score(p) = \sum_i exp\left(-\frac{1}{2}(x_i' - q)^t \Sigma^{-1}(x_i' - q)\right)$
where x_i' indicate a point in the second scan transformed according to the transformation represented by parameter p

(vi) optimize the parameter using Newton's algorithm and update p

(vii) goto (iii) until convergence is achieved.

Based on this 2D solution, 3D NDT which could match 3D point clouds was introduced (Takeuchi and Tsubouchi, 2006; Magnusson et al., 2007, 2009). The scan is divided into 3D voxels instead of grids and normal distributions of 3D coordinate points are used. It differs from 2D NDT in the definition of transformation function and its derivatives. One of the limitations of the NDT based algorithms is its discontinuous cost function due to the rectilinear cell boundaries (Das and Waslander, 2014). This can be addressed by using overlapping grid cells (Biber and Strasser, 2003) or by using trilinear interpolation between grid cells (Magnusson et al., 2009). These strategies only mitigate the problem instead of removing them. The runtime scales linearly with the number of distributions and a better solution is to approximate the surface by fewer Gaussian distributions (Das and Waslander, 2014). Coarse-to-fine approaches have also been tried using NDT (Das and Waslander, 2012) although this solution does not scale well with 3D data. To deal with these problems of NDT, SRG-NDT (Segmented Region Growing NDT) was proposed by Das and Waslander (2014). In this method, the ground points were segmented from the scans and non-ground points were clustered using a region growing clustering algorithm and then each cluster was modelled as a Gaussian distribution. The method offered a continuous cost function and provided a fast and accurate solution. A NDT algorithm based on variable sized voxel was presented by Jun et al. (2015) which provided a more accurate solution.

An improved NDT method which made use of colour information was presented by Hong and Lee (2016). Kim and Lee (2016) introduced a super voxel based partitioning to NDT thus increasing the modelling accuracy of the normal distributions. Many probabilistic formulations were also used for solving the point set registration method. Some of the GMM based methods (Myronenko et al., 2010; Yang et al., 2014; Jian and Vemuri, 2011; Campbell and Petersson, 2015) achieved registration accuracy similar to fine registration methods.

2.2 Multiview Registration

While pairwise registration methods aim to align a pair of scans, multiview or global registration methods aim to align all of the scans to a common coordinate system. The initial pairwise alignment is performed either prior to multiview registration or it is embedded in the multiview algorithm. The following section outlines the various approaches to solve the multiview registration problem. The works are generally outlined in chronological order. Various common trends are highlighted in bold to give a broad outline of different formalisations of the problem.

Sequential and Simultaneous approaches: The multiview registration methods can be broadly classified into sequential and simultaneous methods. The sequential registration methods align scans in a sequential manner considering pair of scans at a time whereas simultaneous registration methods consider all the scans at once to find the global estimates. The global registration achieved by sequentially propagating the pairwise transformations results in accumulation of errors (Chen and Medioni, 1992). In 3D modelling, integration of the views is done once multiview registration is performed. Early approaches directly performed integration of views on top of the pairwise registration results thus skipping the global registration stage (Soucy and Laurendeau, 1992, 1995; Dorai et al., 1996). The scans were integrated using Venn diagrams or averaging after voxelization. Chen and Medioni (1992) proposed to use an intermediate model to register the scans sequentially. Instead of registering with only a neighbouring view, a scan was registered to the merged model of the previous scans thus utilising information from the merged views. Gagnon et al. (1994) pointed out that in such an integrating approach, the registration of the previous image did

not change once it was merged whereas a subsequent view might have information to improve the registration of the previously processed views. Gagnon et al. (1994) and Bergevin et al. (1996) formed a balanced star shaped network of views and each scan was registered to all other scans. The incremental transformation matrix was obtained using a linear least squares technique by small angles approximation (Bergevin et al., 1992). However, computational complexity and memory requirements of these methods were high because of the registration of each view with every other view in all iterations. Pennec (1996) defined the concept of a mean shape and registered each of the scans to the mean shape thus proposing to avoid the accumulation of errors. Another set of approaches considered a dynamic system of springs where each view was connected to other views by means of springs and the transformation was estimated by modelling the force acting in the springs (Eggert et al., 1996; Stoddart and Hilton, 1996). These methods involved initialising a set of parameters.

Multiview registration as Optimisation problem: The multiview registration can be modelled as an optimisation problem where the sum of distances between the overlapping pairs of scans is minimised. Let S_i, $i = 1 : N$ indicate the scan views and T_i, $i = 1 : N$ be the global transformation matrices which transform the scans to a common coordinate system. Without loss of generality, the initial scan is considered as the reference scan and all the other views are transformed into its coordinate system so that T_1 becomes identity. Let $D(X, Y)$ indicate the distance between the two surfaces represented by the scan views X and Y. Now multiview registration aims to find the global transformation so as to minimise the distance given by Eqn.2.4. Here the distance between pairs of non-overlapping views is set to zero.

$$\epsilon^2 = \sum_{i \neq j} D(T_i(S_i), T_j(S_j)) \tag{2.4}$$

Since the transformation matrix involves a rotation matrix component which depends non-linearly on rotation angles, the formulated optimization problem is a highly non-linear problem and cannot be solved analytically (Neugebauer, 1997b). Iterative solutions where the incremental transformation matrices can be linearised were suggested by many authors (Neugebauer, 1997a) where the problem was then solved numerically using Newton-Taylor or Levenberg-Marquardt methods (Lawson and Hanson, 1995). Another approach is to decouple translation and rotation and solve them separately

to obtain an analytic solution. Benjemaa and Schmitt (1998) proposed an analytic solution where rotations represented by unit quaternions were optimized by applying an iterative process and translations were optimized by solving a linear system which depended on the rotated centroids. An extension of ICP algorithm to multiple views was proposed by Williams and Bennamoun (2001) where the registration problem was posed as the minimisation of a least squares cost function. The rotations and translations were decoupled before optimisation and re-weighted least-squares robust estimation framework was used for optimisation.

Pulli (1999) proposed a simultaneous multiview solution which registered each scan to all the overlapping scans in an iterative manner. To reduce the memory footprint and computational time, only the pairwise constraints were used for multiview registration. The pairwise constraints included the relative transformation matrices and the uniformly sampled points from the overlapping regions. This method tried to arrive at a global solution where the pairwise constraints were met. Re-registration of the scans was avoided which saved time. However the refinement achieved by the process was overlooked. Benjemaa and Schmitt (1999) proposed an approach similar to Bergevin et al. (1996) but the incremental transformation of a view was applied as soon as it was obtained unlike the latter, where all transformations were calculated before update. This speeded up the process and the order of update was chosen randomly to avoid bias. Z-buffers were used to establish correspondences which accelerated the correspondence estimation. Cunnington and Stoddart (1999) evaluated three approaches (Pennec, 1996; Bergevin et al., 1996; Stoddart and Hilton, 1996) and presented a comparison based on their relative performances. Nishino Ko Ikeuchi (2002) presented a simultaneous registration approach for range images where the minimization of a global error function was performed using M-estimator with conjugate gradient search. The rotations were represented in the form of unit quaternions. The global closest point among all views was found out using kd-trees and transformations in each iteration were updated after estimating for all views. Masuda (2002) presented a hybrid approach where registration and integration were performed alternately to arrive at a global solution. This iterative registration-integration method was limited by large computational requirements. A real time 3D model acquisition set up was suggested by Rusinkiewicz et al. (2002) where the scans were acquired and integrated using a voxel-grid approach in real time. A post processing step was performed

where the multiview registration by Pulli (1999) was used.

View Graph Based Approaches: A large group of methods make use of view graphs to assist in global transformation. A view graph is constructed with scan views as nodes and the relative transformations as the edges. Huber and Hebert (2003) constructed a view graph and searched it to find a sub-graph (or spanning tree) which satisfied the defined global consistency measures. This search was formalized as a discrete optimization problem. A continuous optimisation was performed on the sub-graph which made use of the multiview approach by Neugebauer (1997a). The sub-graph was formed by adding one edge at a time while ensuring global consistency criteria. If a wrong edge which did not violate the global consistency was added, then this corrupted the results. Some approaches to handle this problem were also discussed. Mian et al. (2004) presented a fully automatic multiview approach using tensor representation for range image registration. A spanning tree was formed by matching tensors which was used for global multiview alignment. An open source framework for optimising graph based non-linear error functions was proposed by Kümmerle et al. (2011).

Loop closing approaches: A set of approaches for multiview registration tries to minimise the error between local coordinate frames. One such early approach was proposed by Lu and Milios (1997) (LUM), where robot pose estimation and scan registration in a 2D case (usually called simultaneous localisation and mapping - SLAM) was solved. The various spatial and odometric constraints were utilised to minimize the error and the problem was formulated by linearising the rotation. This is useful only in the case of small rotations. Sharp et al. (2004) proposed an analytical method to solve the multiview registration by using the relative motions between views as error criterion. This approach did not require linearisation or point correspondences. The basic idea was to build a view graph and then decompose it into basic cycles. The non-linear optimization problem over each basis cycle was solved in closed form, and the solutions for the constituent basis cycles were merged using an averaging technique. When a cycle occurs, the total transformation in that cycle should be identity. This constraint, called *loop closing constraint*, was utilised to distribute the error from pairwise registration among the views in a fair manner. The scan registration proposed by Lu and Milios (1997) was extended to the 6DoF case by Borrmann et al. (2008b,a) where the multiview registration was achieved by linearisation of the mo-

tion parameters and solving using a linear system of equations using sparse Cholesky decomposition (Davis, 2006) and improved kd-tree. Sprickerhof et al. (2009) used the loop closing constraint for solving 6D SLAM by explicitly registering the last scan in the cycle to the initial scan and then distributing the error. However a costly global refinement was required to improve the accuracy. An online approach for multiview registration was proposed by Matabosch et al. (2008) where the pairwise registration using a modified point-to-plane error metric was performed until a cycle was detected. Once a cycle was detected, the loop closure error was minimised using Levenberg-Marquardt optimisation (Lawson and Hanson, 1995). This approach was shown to be more accurate compared to Sharp et al. (2004). The cycle minimization approaches were categorized into analytic and statistic approaches by Batlle et al. (2007). An online loop closure algorithm for online registration was presented by Weise et al. (2011) which made use of texture information in addition to geometric constraints.

Approaches utilising Manifold structure: The global registration problem was formulated as unconstrained optimisation on a constrained manifold by Krishnan et al. (2005). This approach exploited the Lie group structure of $SO(3)$ and the simple representation of its associated Lie algebra $so(3)$ in terms of $\mathbb{R}^3$. The rotations were decoupled from translations and the Lie group structure of rotations ($SO(3)$) was utilised to find a closed form solution using singular value decomposition (SVD) for simultaneous multiview registration in noise free scenario. An iterative solution based on Newton's method on $SO(3)$ that has locally quadratic convergence was also presented. A probabilistic Bayesian framework on top of this method (Krishnan et al., 2005) was proposed by Mateo et al. (2014) to introduce robustness by identifying wrong correspondences as well as pairwise matches. The parameters were estimated using an Expectation Maximisation (EM) algorithm. Govindu (2004) presented a motion averaging solution to obtain global motion estimates from redundant relative motion estimates utilising the Lie group structure of 3D rigid body transformations. Shih et al. (2008) presented the multiview registration using a circuit model. The problem was converted into a quadratic programming problem of Lie algebra parameters. The cycle constraints of the graph was utilised to eliminate the accumulation errors. A linear solution to distribute the accumulation error to proper positions in the graph was proposed. A comparison with similar approaches highlighted its performance. The method had low time and space complexity.

Probabilistic Framework: A probabilistic generative model was used to formulate the multiview registration problem by Evangelidis et al. (2014). All the point sets were considered as realizations of a Gaussian Mixture (GMM) and registration was cast into a clustering problem. An Expectation Conditional Minimisation algorithm was formulated which estimates the GMM parameters as well as the global motion parameters. The variances provided information on the quality of registration. The method by Mateo et al. (2014) also belongs to probabilistic approaches.

A classification of multiview approaches into a) direct multiview registration b) using intermediate shape reconstruction and c) minimising error between coordinate frames was presented by Shih et al. (2008). The type c approaches have the best time and space complexity as they only handle the inconsistency between coordinate frames and need not store or process the views iteratively. However the goal of error distribution may not match with the goal of registration (Shih et al., 2008). To deal with this problem, Govindu and Pooja (2014) presented an approach where the error distribution and registration were performed alternately. They presented an extension of ICP to multiple views by making use of the Lie group structure of 3D rigid body transformations and the availability of redundant overlaps (not just the adjacent scans in a sequence overlap). Their algorithm alternately iterated the correspondence estimation and motion averaging steps. The additional constraints helped to average out the errors than to simply distribute them. However, they assumed the input relative transformations to be correct and the results were affected in presence of wrong relative transformations, since averaging could not handle outliers. An approach to robust multiview registration which filters the outliers before motion averaging is presented in Chapter 4.

2.3 Summary

The stages involved in 3D registration as well as the various existing approaches to solve the challenges in each stage are detailed in this chapter. The basic steps involved in pairwise registration by feature matching is explained in section 2.1.1.

The existing approaches to coarse transformation estimation by feature matching are limited by the high percentage of outliers in the set of corresponding points. A

novel approach to make the transformation estimation faster and accurate even in the presence of high percentage of outlier corresponding points is presented in Chapter 3. Multiview registration is a challenging problem due to the large number of constraints involved. The simple distribution of accumulated errors among the different pairs of views does not yield a smooth point cloud, in presence of large errors. A robust multiview registration technique which averages out the accumulated errors even in the presence of wrong relative transformations is presented in Chapter 4.

CHAPTER 3

Robust Pairwise Registration of Point Clouds

Prelude: This chapter presents a novel coarse registration algorithm for the pairwise registration of point clouds by feature matching. The correspondence set obtained by feature matching is prone to the presence of outliers. A novel robust estimation technique to estimate the best transformation from the outlier-corrupted set of correspondences is presented in this chapter. The performance of the proposed algorithm is compared with the existing algorithm RANSAC and other variants of RANSAC for the registration of scan pairs of objects from different databases.[1]

3.1 Introduction

The task of registration of point clouds appears in many application scenarios like object reconstruction, robotic path building and environment mapping, object recognition etc. The approaches to solve the registration problem vary according to the constraints imposed by the image acquisition, representation, preprocessing and post processing stages. In this chapter, the case of object reconstruction is considered.

Multiple scan views of an object are required to build a complete point cloud. The scan views are acquired such that there is a certain amount of overlap between the adjacent scans. The extent of overlap is decided based on the registration algorithm employed. As discussed in Chapter 2, registration of all the views to obtain the complete point cloud is performed in multiple stages viz. pairwise registration and multiview registration. The pairwise registration stage aligns two partially overlapping scan views whereas multiview registration (global) stage aligns all the scan views to a common coordinate system. The pairwise registration itself is often performed in multiple stages - coarse (rough) registration and fine registration. A combination of coarse and fine registration stages are generally employed to estimate the transformation between a pair of views. The resolution and smoothness of the resultant point

[1]Major contents of this chapter are published in *Image Analysis and Stereology* Vol: 35 (1), 2016. Authors: Dhanya S Pankaj, Rama Rao Nidamanuri

cloud depends on the registration accuracy, scanning resolution and noise.

In the absence of a coarse registration stage, closely spaced scans are to be acquired and registered using fine registration techniques. Employing a fast and accurate coarse registration helps in reducing the number of scans acquired which in turn helps in reducing memory footprint and processing time of the fine registration stage. The real time 3D modelling of objects often employs a fast and accurate coarse registration stage.

A discussion on the various approaches in literature for performing the coarse registration is available in section 2.1.1. One of the most popular methods for coarse registration is feature matching and the different steps involved are explained in section 2.1.1. Given a set of corresponding points, various closed form solutions are available in literature to estimate the rigid body transformation between the point clouds. However, the set of corresponding points may contain wrong matches (outliers) because of partial overlap, noise in the scan, local nature of features, inaccuracies in the previous stages etc. Hence the percentage of correct matches (inliers) can be very less. The least squares estimators cannot handle the presence of outliers and hence the results get corrupted. To deal with outliers in estimation, various robust estimators like Least Median of Squares (Rousseeuw, 1984), Least Trimmed Squares (Rousseeuw and Leroy, 2005) etc. were developed by the statistical community. The robustness of a detector is identified by its breakdown point (Wang, 2004). The breakdown point of an estimator can be roughly defined as the smallest percentage of outlier contamination that can cause the estimator to produce arbitrarily large values (Rousseeuw and Leroy, 2005). The more robust an estimator is, the higher is its breakdown point. Most of the above estimators are not practical in the case of many computer vision applications, as many problems require a breakdown point of more than 50%. To deal with this, the computer vision community has developed many robust estimators like Random Sample Consensus (RANSAC) (Fischler and Bolles, 1981), Hough Transform (VC, 1962) etc. In the case of 3D registration, the percentage of inliers in the set of corresponding points depends upon many factors and hence cannot be guaranteed to be greater than 50%. Hence robust estimators are usually employed to estimate the transformation.

The number of RANSAC iterations depends on the number of correspondences,

complexity of the model and percentage of inliers (Chum and Matas, 2005). If the percentage of inliers is very low, the convergence of RANSAC is generally slow. Many variants to RANSAC like MLESAC (Torr and Zisserman, 2000), Progressive Sample Consensus (PROSAC) (Chum and Matas, 2005), Locally Optimized RANSAC (LoSAC) (Chum et al., 2003), NAPSAC (Myatt et al., 2002) etc have been proposed to handle this problem. Many of these techniques are applied to 2D vision problems like epipolar geometry estimation, motion segmentation, homography estimation, object recognition, image retrieval etc. However, adaptations of these methods to the 3D registration problem are not available in literature to the best of our knowledge. This thesis proposes a novel robust estimation algorithm which provides accurate results for 3D registration in less computational time, when compared to RANSAC and its variants.

The section 3.2 discusses the motivation behind the present work. Then the overall structure of the registration method used for coarse alignment is outlined. Section 3.3.6 details the proposed algorithm developed in this chapter to solve the robust transformation estimation problem. Then the second stage i.e. fine alignment stage is explained. The various experiments conducted to evaluate the proposed robust estimation method and the results obtained are then detailed.

3.2 Motivation

The de-facto standard algorithm for robust transformation estimation in literature is RANSAC. The general outline of 3D coarse registration using RANSAC is as follows. From the set of corresponding points (containing outliers) obtained by feature matching, RANSAC identifies the best transformation which has the largest support. In RANSAC, a minimum sample (minimum of three pairs of corresponding points is required to estimate the 3D transformation matrix) is randomly selected from the set of corresponding points. Then the transformation is estimated and the number of corresponding points which are inliers to the transformation is found. The distance between the points in a corresponding point pair after applying the transformation is calculated and if it falls below a threshold, the pair is treated as an inlier to the transformation. The sampling - transformation estimation - inlier calculation cycle is repeated

and the transformation which has the maximum number of inliers is selected.

This basic method of RANSAC does not make use of problem-specific information to assist the sampling process. Utilizing additional information is more efficient than the random sampling strategy of RANSAC as pointed out by (Chum and Matas, 2005). Different guided sampling strategies have been proposed in literature. In NAPSAC (Myatt et al., 2002), the assumption of high probability of finding an inlier adjacent to another inlier is used. This is based on the presumption that inliers of a model tend to lie closely. However this assumption is problem dependent and generally does not hold in the case of 3D transformation estimation where all selected points confined to a particular area may lead to a local solution. In fact, sample points lying close together are generally avoided. They should be distributed across the shape for more accurate alignment.

In the case of 3D registration, the quality of a corresponding point pair can be obtained from the 3D feature matching stage. The distance between the corresponding features can be used as a rough indicator of the quality of the match, even though not completely reliable. This information can be efficiently utilized for guiding the sampling. To make use of the quality of 3D feature matches, a guided sampling technique (Chum and Matas, 2005) is employed in the proposed algorithm. The sampling method finds balance between the random sampling and the sampling based purely on the quality of correspondence matches. Since promising samples are considered early, this can lead to significant computational time improvement to the algorithm. However, if the quality score is not indicative of the correctness of the match, then the accuracy of the estimate may suffer. Having a lot of self similar points in the model can also lead to sub-optimal solutions as the feature match quality of these wrong matches may be high. Since only a subset of the whole set of corresponding points is considered for sampling where all combinations of matches are not explored, and since in some cases the matching score may not very reliable, this sampling can lead to sub-optimal solutions.

To address the sub-optimal solution obtained, a refinement strategy (Chum et al., 2003) is employed in the proposed algorithm. As Chum et al. (2003) pointed out, an all-inlier sample may not lead to a transformation model which finds all the inliers. The best solution obtained usually lies near to the optimal solution and can be

refined locally. The refinement of the current best transformation estimate (obtained by guided sampling) is performed by searching in the neighbourhood by varying the threshold. However, since the process is repeated each time the current best solution is obtained, this increases the computational load. This computational burden due to local optimization is compensated by the computational savings obtained by guided sampling.

3.3 Method

The various stages employed in transformation estimation by 3D feature matching is depicted in Fig.2.1. The solutions employed in each of these stages in the current work are explained here.

3.3.1 Preprocessing

The acquired scans are down-sampled using voxel-grid filtering. Statistical filtering is employed in the case of noisy point clouds. These methods are detailed in section 2.1.1.

3.3.2 Normal estimation

Normal estimation using Principal component analysis (PCA) is employed in this work. The local neighbourhood around a point is approximated to a plane and the plane normal is calculated and assigned as the point normal. PCA is carried out on the local neighbourhood around a point and the three eigen vectors are calculated. The eigen vector corresponding to the lowest eigen value is considered as the normal direction. This is justified by the basic assumption that the spread of data is minimal along the normal direction compared to the surface directions. The local neighbourhood can be defined in two ways. One way is to consider all the points inside a sphere with a specified radius, centred at the point. In this case, some of the points may not have sufficient neighbours within the sphere to define a normal. Another method is to consider a specified number of neighbouring points. In this case, a normal can

always be defined but the non-uniform sampling of neighbourhood may not lead to smoothly varying normals. This work employed the first method of normal calculation. The points where normals could not be defined due to lack of points were removed from the original point cloud. The direction of extracted normal is ambiguous in this method. The normal $(\vec{n_i})$ at each point (p_i) was then oriented consistently towards the view point (v_p) (Rusu, 2010) using Eq.3.1.

$$\vec{n_i} \cdot (v_p - p_i) > 0 \tag{3.1}$$

3.3.3 Keypoint detection

The keypoint detection method proposed by Zhong (2009) is employed in this work. Based on the various analysis of keypoint detectors in literature (Tombari et al., 2013; Filipe and Alexandre, 2014; Salti et al., 2012; Guo et al., 2016), the Intrinsic Shape Signature (ISS) keypoint detection method was selected. The ISS keypoint detector is a fixed scale local keypoint detector. Local fixed scale indicates that the detection of a keypoint is based on the property in a fixed local neighbourhood of the point. The various steps involved in ISS keypoint detection can be summarised as follows:

- The scatter matrix of the local neighbourhood at each point p_i is calculated.
- Eigen values in the order of decreasing magnitude $\{\lambda_i^1, \lambda_i^2, \lambda_i^3\}$ and their eigen vectors $\{e_i^1, e_i^2, e_i^3\}$ are computed.
- The pruning of points is done based on the ratios of eigen values. i.e. If the conditions $\lambda_i^2/\lambda_i^1 < \gamma_{21}$ and $\lambda_i^3/\lambda_i^2 < \gamma_{32}$ are not satisfied, then the point is pruned out.
- Among the selected points, a non-maximal suppression in the local neighbourhood is done based on the smallest eigen value λ_i^3.
- The points having maximum value of the saliency λ_i^3 in the neighbourhood and which satisfies the conditions listed above are selected as keypoints.

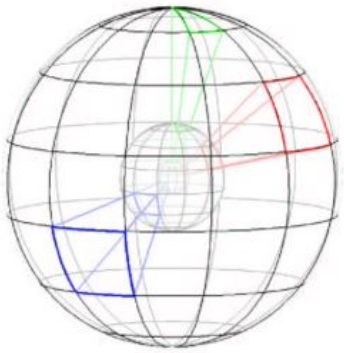


Figure 3.1: Visualisation of the bins used in 3DSC.

3.3.4 Feature extraction

The spatial histogram based descriptor 3D Shape Context (3DSC) (Frome et al., 2004) is used in this work for feature extraction. The local neighbourhood is a sphere centred on the point p with its north pole oriented with the surface normal. The support region is divided along the azimuthal and elevation dimensions by equally spaced boundaries and radial dimension by logarithmically spaced boundaries as shown in Fig.3.1. The radial, elevation and azimuthal divisions can be represented by $R = \{R_0 \dots R_J\}, \Theta = \{\Theta_0 \dots \Theta_K\}$ and $\Phi = \{\Phi_0 \dots \Phi_L\}$. Given the first radius division R_0 as the minimum radius r_{min} and R_J, the maximum radius r_{max}, the radial boundaries are calculated as in Eq.3.2.

$$R_j = \exp\left\{\ln(r_{min}) + \frac{j}{J}\ln\left(\frac{r_{max}}{r_{min}}\right)\right\} \tag{3.2}$$

The logarithmic divisions along the radial dimension help to make the descriptor robust to shape distortions with distance from the base point. A minimum radius is specified to avoid over sensitivity to small shape differences near the center of the sphere. A bin (j, k, l) accumulates a weighted count $w(p_i)$ for each point p_i whose spherical coordinates relative to p falls within the intervals of the bin $[Rj, R_{j+1}), [\Phi_k, \Phi_{k+1})$ and $[\Theta_l, \Theta_{l+1})$. Each point contributes the quantity $w(p_i)$ to the bin count and is given by Eq.3.3, where $V(j, k, l)$ represents the volume of the bin and ρ_i the local point density. The variation in bin sizes is compensated by the normalisation of the bin count by bin volume. The local point density factor accounts for the point density variations due to scanning limitations and is calculated as the count of points falling within a sphere of radius δ. The descriptor is formed by appending the bin counts for all the

bins. A degree of freedom in the azimuthal direction is accounted by storing multiple descriptors.

$$w(p_i) = \frac{1}{\rho_i \sqrt[3]{V(j,k,l)}} \tag{3.3}$$

3.3.5 Transformation estimation

The corresponding points are estimated by the feature matching using kd-trees as described in section 2.1.1. Correspondence refinement strategy by (Buch et al., 2013) is employed. From the set of corresponding point pairs (C), the proposed algorithm (described below) is used to estimate the transformation.

3.3.6 Proposed algorithm

The transformation estimation by the proposed algorithm consists of four major steps which are iterated until a stopping criterion is satisfied:

(i) sample generation by guided sampling

(ii) transformation estimation

(iii) inlier calculation, and

(iv) local refinement in the case of global best number of inliers.

The guided sampling strategy is employed to form the samples. Samples (S) are formed by randomly selecting points from a growing subset U_n of the correspondence set C sorted by the quality of correspondences. A minimum distance constraint was applied while sampling the points which ensure that sampled points are distributed across the point cloud and are not crowded together so that the estimated transformation is not local. The subset is formed initially from the top-quality correspondences, grows gradually, and eventually degenerates into the random sampling set, as the iteration progresses. The growing subset of correspondences U_n is repeatedly sampled for a minimal sample S, which is used to estimate the transformation model. Then the number of inliers to the estimated transformation is found out. This sampling - transformation estimation - inlier calculation cycle is performed iteratively and the

best transformation model is updated in each iteration based on the maximum number of inliers. If the current transformation is the best one estimated so far, then a refinement stage is carried out. The global best solution obtained so far by the outer loop is refined locally by performing an inner sampling - model estimation - inlier calculation loop. This process of local refinement leads to an optimal solution compared to the current best solution.

The proposed algorithm is given in Algorithm 1 and is explained in detail below. The set (C) of N correspondences obtained from 3D feature matching is the input to the algorithm. The set is sorted according to the quality of the matched features. A subset of corresponding points with size n and of highest quality is denoted as U_n. A sample S consists of minimum number of corresponding points ($m = 3$) required to estimate the 3D transformation. The samples are drawn from the (growing) subset U_n of the total correspondences. Quality of a sample is the quality of the corresponding point pair in it with minimum quality. Let T_N be the number of samples (sample size m) drawn by standard RANSAC from the set of N correspondences. Let T_n be the average number of samples in the original set of T_N samples, having data points only from U_n. A recurrence relation is used for finding T_{n+1}. As the values of T_n are not integer in general, T'_{n+1} is calculated. These are defined in Eq.3.4 where T'_n = 1.

$$T_n = T_N \frac{\binom{n}{m}}{\binom{N}{m}}$$

$$T_{n+1} = \frac{n+1}{n+1-m} T_n \tag{3.4}$$

$$T'_{n+1} = T'_n + \lceil T_{n+1} - T_n \rceil$$

The growth function for the sample generation set U_n is formed from T'_n using the Eq.3.5. The growth function should reflect the result of previous tests and this is indicated by k, the number of tests conducted so far, since RANSAC runs are typically characterized by a success preceded by a number of failures. Samples are drawn randomly from the subset U_n which grows in size starting from $n = m$ to a maximum of T_N according to the growth function.

$$g(k) = min \{n : T'_n \geq k\} \tag{3.5}$$

Once the minimum sample is obtained as described, a 3D transformation model is computed using a least square technique (Umeyama, 1991).

Algorithm 1 Proposed Algorithm

▷ %comment N : Number of correspondences, m : Minimum Sample Size (3 for 3D registration), η_0 : Threshold for stopping criterion % comment

$n^* \leftarrow N$, $n \leftarrow m$, $k \leftarrow 0$, $I^* \leftarrow 0$, $k^* \leftarrow 1$ $T'_n \leftarrow 1$, $T'_N \leftarrow 200000$,

while $k < k^*$ **do**

1. Select model generation set

$k \leftarrow k + 1$

if $k = T'_n$ and $n < n^*$ **then**

$n \leftarrow n + 1$

end if where T_n, T_{n+1} and T'_{n+1} are given by Eq.3.4

Sample selection of size m :

if $T'_n < k$ **then**

Select sample such that it contains $m - 1$ points at random from U_{n-1} and u_n

else

Select m points from U_n at random

end if

2. Model Generation

Generate model or estimate 3D transformation from the sample

3. Model Verification

Find inliers I_k

if $I_k > I^*$ **then**

A. Run Algorithm 2 and update best model and I_k if $I^{LO} > I_k$

B. Select termination length n^* and k^* such that equations

$k_{n^*} \geq log(\eta_0) / log(1 - \epsilon_{n^*}^m)$ and

$I_n^* \geq I_{n^*}^{min}$ are satisfied.

C. $I^* \leftarrow I_k$

D. $k \leftarrow k + 1$

end if

end while

Algorithm 2 Local Optimization

$m_{min} \leftarrow min(I_k/2, m)$

$k_{LO} \leftarrow 10$

$K \leftarrow 5$

a. Select m_{min} sized sample from the inliers I_k

b. Generate model from the sample

c. **Iterative**

Find inliers with distance $< K\theta$

Generate model from these inliers using linear algorithm

Reduce threshold and repeat step **Iterative** until threshold $= \theta$

d. Repeat steps a to c k_{LO} times and store the best model and best I as I^{LO}

For a set of corresponding point pairs $\{(x_i, y_i)\}$, where $i = 1 : N$, a rigid body transformation which minimizes the total distance between the corresponding points (after applying the transformation), is calculated. This distance is given by Eq.3.6 where R and t represent the 3D rotation matrix and translation vector respectively, x_i and y_i represent the 3D co-ordinate vectors of the point pair i.

$$E^2 = \frac{1}{N} \sum_{i=1}^{N} \|y_i - (Rx_i + t)\|^2 \tag{3.6}$$

Once transformation is estimated, the set of inliers is calculated as follows. The estimated transformation is applied to the source points and the distance between the target point and the transformed source point for each of the corresponding point pairs, is computed. This distance for a point pair i is given by Eq.3.7.

$$e^2 = \|y_i - (Rx_i + t)\|^2 \tag{3.7}$$

If this error falls within a distance threshold, then the point is considered as an inlier to the model calculated. The threshold is calculated empirically and is based on the resolution of the point cloud and sensor noise. The sampling, model generation and inlier calculation phase is repeated until a local best solution, i.e. the current best number of inliers, is obtained.

Once a local best solution is obtained, a refinement of the solution is carried out. A hybrid iterative and inner RANSAC algorithm (Chum et al., 2003) is implemented for model refinement. Once the current best solution is obtained, a sample is selected randomly from the set of inliers. This sample size can be greater than the minimal size as the sampling is done on the set of inliers and the chance of corruption due to outliers is less. In the present work, it is taken to be the three or half the number of inliers, whichever is maximum. Then the transformation model computed from this sample is further refined iteratively by varying the threshold for finding the inliers and estimating transformation from the inliers. This is repeated k_{LO} times and the best number of inliers is updated depending on the result.

The proposed solution takes care of the extra computational burden caused by local optimization steps by utilizing the computational time savings obtained by guided sampling. Also the sub-optimal solution attributed to the limited number of samples due to guided sampling is taken care of by the local optimization where samples are generated from the set of current best inliers. Thus samples other than those from the sample generation set of guided sampling are also considered.

The stopping criterion of the loop is calculated as follows. Let m be the size of the minimum sample, and ϵ be the percentage of inliers. In k iterations k samples will be drawn. The probability of not finding an all-inlier sample in k samples is $(1-\epsilon^m)^k$. k is selected such that this probability falls below η_0. This is given by Eq.3.8 where P_{I_n} represents the probability of obtaining an inlier which is approximated as the percentage of inliers, ϵ. This is the general stopping criteria of the standard RANSAC algorithm.

$$k \geq \log(\eta_0)/\log(1-P_{I_n}) \tag{3.8}$$

In the proposed algorithm, the samples are generated from a small subset of correspondences. Hence an additional check is also employed to ensure that the model obtained is not randomly endorsed by a set of outliers (Chum et al., 2003). Assuming binomial distribution for the cardinality of the set of random 'inliers', the probability of obtaining a random inlier set with size i is given by Eq.3.9.

$$P_n^R(i) = \beta^{i-m}(1-\beta)^{n-i+m}\begin{pmatrix} n-m \\ i-m \end{pmatrix} \tag{3.9}$$

For each sample generation set with size n, the minimum number of inliers ${I_n}^{min}$ is calculated so that the probability of obtaining a random support of that size falls below a threshold, ψ. This is given by Eq.3.10.

$$I_n^{min} = min\{j : \sum_{i=j}^{n} P_n^R(i) < \psi\} \tag{3.10}$$

Then, only a model which provides the number of inliers I_n greater than this minimum support size is selected as a solution, given by Eq.3.11.

$$I_n > I_n^{min} \tag{3.11}$$

The iterations are repeated until the two conditions given by Eqs.3.8 and 3.11 are satisfied. Once the final best set of inliers is obtained, the final transformation can be estimated from it using one of the closed form solutions (Umeyama, 1991).

Algorithm 3 Fine Alignment

▷ %comment $T_0 = Transformation_{coarse}$, $i = 0$ %comment

1. Form a kd-tree from the target cloud points.
2. Apply the transformation T_i to the source point cloud to align with the target point cloud.
3. For each of the source point, estimate the nearest neighbour in the target kd-tree and form the set of correspondences.
4. Perform correspondence refinement.
5. Estimate rigid body transformation matrix (δT) using closed form solutions.
6. Update the transformation $T_i = \delta T * T_{i-1}$
7. Repeat steps 2 to 6 until $i < max\_iterations$ or distance between corresponding points < threshold

3.3.7 Fine alignment

Once the final transformation is obtained from the coarse registration stage, it can be further refined by fine alignment methods. Some of the existing approaches for fine alignment are discussed in section 2.1.2. The accurate coarse alignment obtained by the proposed method helps in reducing the time spent in the fine alignment stage

and also prevents local minima. The fine alignment method (similar to ICP) employed can be summarised in Algorithm 3. The correspondence refinement strategy by (Buch et al., 2013) and the closed form transformation estimation by (Umeyama, 1991) are employed in this work.

3.4 Experiments

To evaluate the proposed algorithm, pairwise registration of 3D scans of different models are performed and the performance is compared with existing algorithms. The partially overlapping point clouds of various objects acquired for 3D modelling task were used for testing. Scans from two different datasets, as given in Table 3.1, were considered for the study.

Table 3.1: The datasets used in this study.

S.No	Model	Dataset	Scanner	Ground Truth
1	Buddha	Stanford	Cyberware 3030 MS	Available
2	Bunny			
3	Dragon			
4	Chef	UWA	Minolta Scanner	Not available
5	Chicken			
6	Parasaurolophus			

To compare the performance of the proposed algorithm, 3D pairwise coarse registration of partially overlapping point clouds was implemented using the three existing algorithms RANSAC, PROSAC (Chum and Matas, 2005) and LoSAC (Chum et al., 2003), in addition to the proposed algorithm. The registration was performed using 3D feature matching. The reference scan and the scan to be registered are called *target* scan and *source* scan respectively in this work.

The set of corresponding point pairs as well as the distance between corresponding feature vectors was given as the input to the algorithm. Coarse registration is then performed by the compared methods by estimating the best transformation matrix from the set of correspondences. The pairwise transformations computed by all the four

algorithms were then compared. In the case of Stanford models, the actual transformations were provided in terms of 3D translation units and 3D rotations represented as unit quaternions. After registration, the estimated 3D transformation matrix was converted into x, y, z translation units and the rotation unit quaternion. The actual and estimated translation units and quaternions were compared. The norm of the distance between translation units and the dot product between the rotation unit quaternions were computed. Lesser the norm of translation difference, better the computation of 3D translation. The rotations were compared by taking the dot product of the estimated and actual quaternions. In the case of normalized quaternions, the dot product of value of 1 means that the 3D rotations are approximately equal. The dot product value indicates how close the estimated rotation is to the actual rotation. Higher the dot product value, better the estimated value. The entire processing was performed on an Intel I3 3.10 GHz processor and the time taken by the algorithms (in milliseconds) in the main loop were computed.

For evaluation of UWA models, an error measure called transformation validation score was used instead of translational distance norm and dot product between quaternions, due to the absence of ground truth. This score is a measure of the distance between corresponding points of the scan pairs after applying transformation. To deal with partial overlap, i.e. to account for the points with no actual corresponding points, a distance threshold was applied so that the points at distances greater than this threshold were filtered from the calculation. The score was normalized with respect to the number of correspondences found. The lower the transformation validation score, the better is the alignment and the registered clouds were additionally verified manually for ensuring this. The number of inliers to the computed transformation model was also computed. The number of hypotheses or models evaluated was also computed and this is an indication of the number of samples tested before finding the final transformation. The experiments were repeated 50 times and the average values are reported.

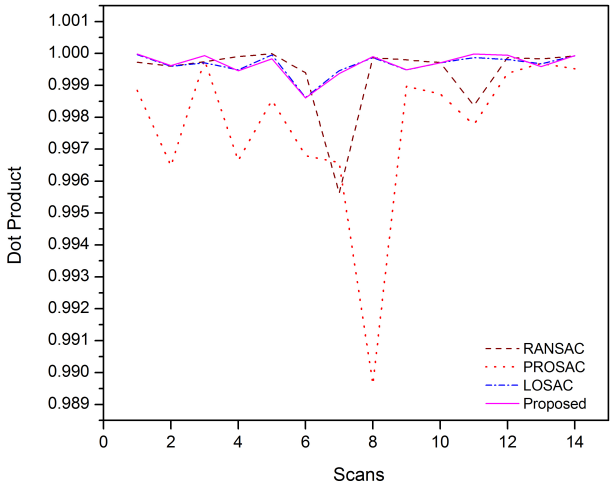
3.5 Results

The partially overlapping scans of the Stanford model Buddha were registered and the various evaluation metrics were compared across the different algorithms. The x-axis indicates various scan pairs of the considered model in all the figures.

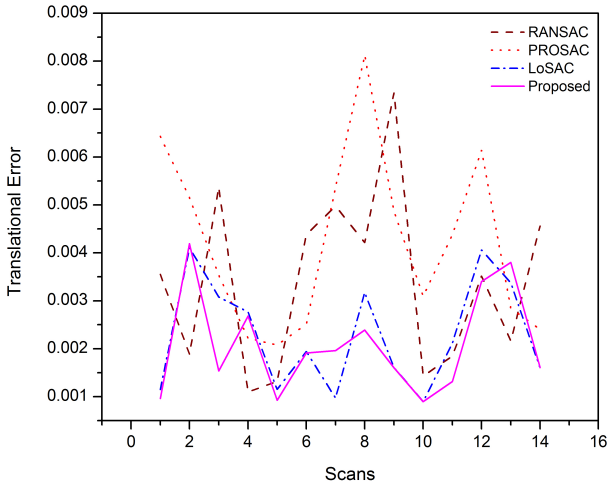
Fig.3.2(a) shows the comparison of dot product of the actual rotation unit quaternion and the one obtained from the registration for different set of scan-pairs, for the tested algorithms. From the figure, it is observed that the proposed algorithm and LoSAC obtained consistently good rotation estimates across the different scan pairs. PROSAC incurs more rotation errors and RANSAC results are not consistently good. The results show that the proposed algorithm obtains good and consistent results for rotation estimation. Fig.3.2(b) shows the error in computation of 3D translation for the same set of scans. It indicates that the proposed algorithm achieves comparatively smaller errors in the estimation of 3D translation. RANSAC and PROSAC converged in large errors. LoSAC and the proposed algorithm achieve comparable results.

The time taken in the main loop (mentioned in graphs as RANSAC iterations) is also plotted for different algorithms in Fig.3.2(c). From the figure, it can be clearly seen that the time taken by the proposed algorithm is much lower compared to LoSAC or RANSAC for all the scans. Only PROSAC runs faster than the proposed algorithm and it should be noted that the rotational and translational accuracy of the PROSAC is much less compared to the proposed algorithm.

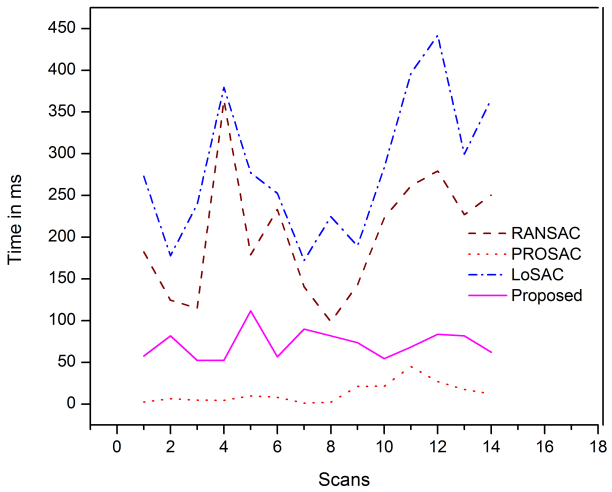
Another set of experiments were conducted using the overlapping scans of the Stanford Bunny dataset and the results are given in Figs.3.3(a) - 3.3(c). The percentage of inliers in the Bunny dataset is much less as compared to the Buddha dataset. However, the trend in the results is similar to that of the Buddha dataset. The proposed algorithm attains good 3D transformation estimates compared to the other algorithms in lesser time. In this dataset, the time taken by the proposed algorithm is even lesser than that of PROSAC. This may be attributed to the lower percentage of inliers in this dataset. When the inlier percentage is very low, the proposed algorithm even outperforms PROSAC in execution time. It shows better results than RANSAC and PROSAC in terms of accuracy.



(a) Dot product between actual and estimated rotation quaternions

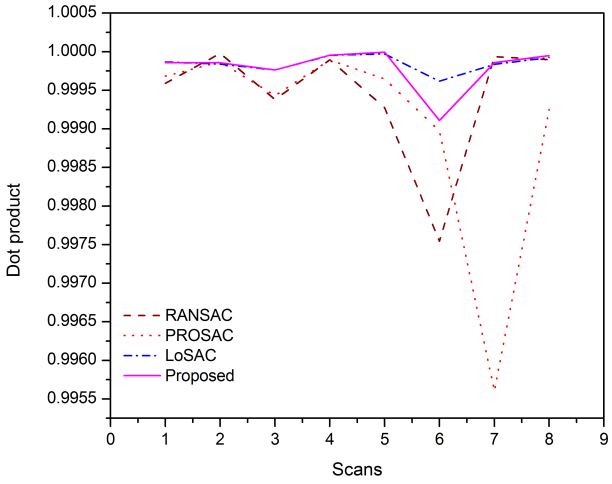


(b) Error in translation estimate

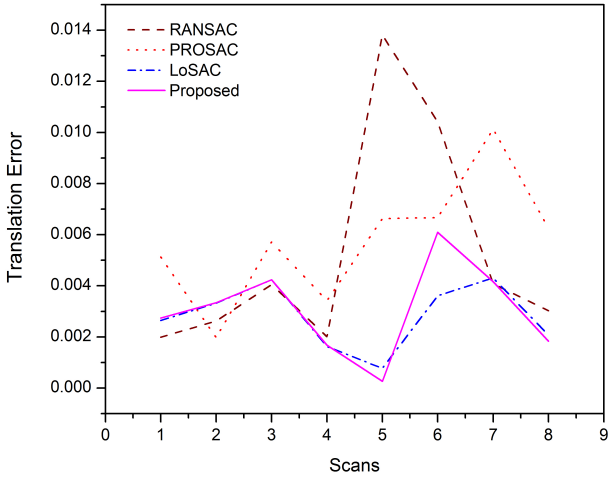


(c) Time taken for RANSAC iterations in ms

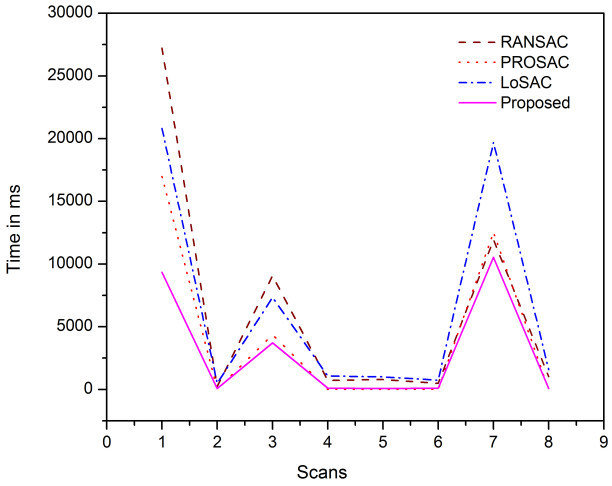
Figure 3.2: Stanford Buddha dataset coarse registration results.



(a) Dot product between actual and computed rotation quaternions

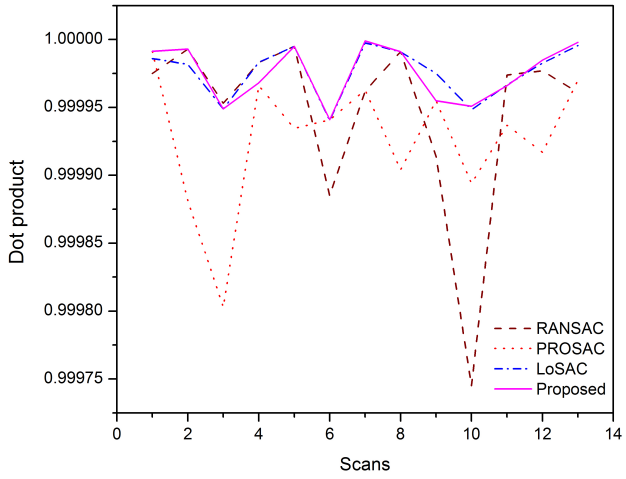


(b) Error in translation estimate

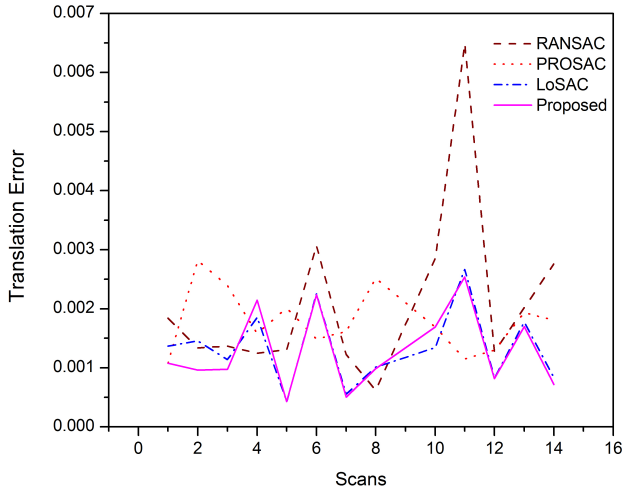


(c) Time taken for RANSAC iterations in ms

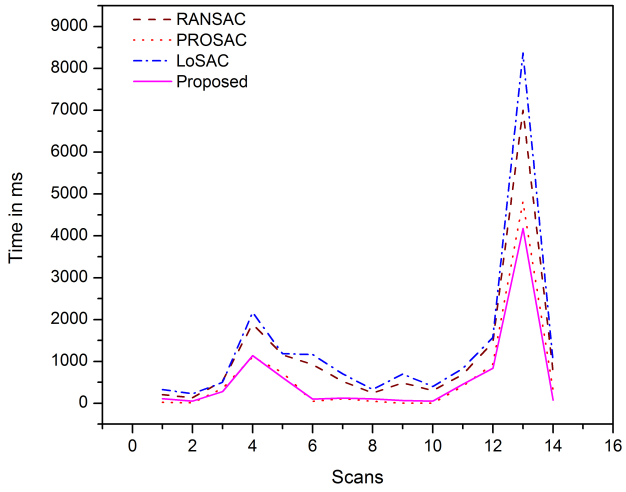
Figure 3.3: Stanford Bunny dataset coarse registration results.



(a) Dot product between actual and computed rotation quaternions



(b) Error in translation estimate

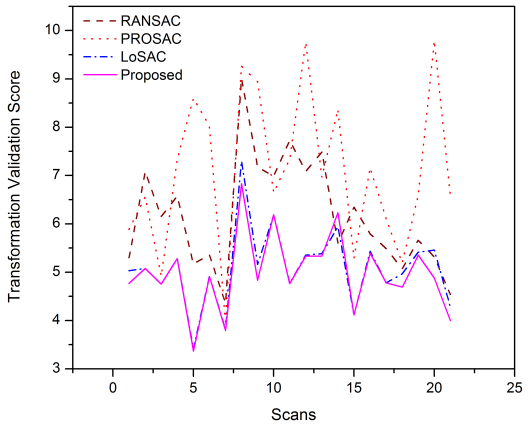


(c) Time taken for RANSAC iterations in ms

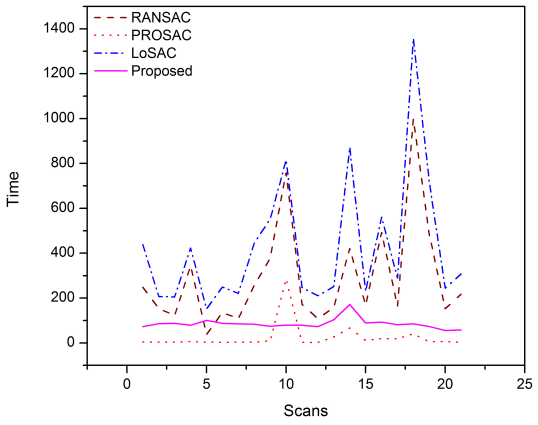
Figure 3.4: Stanford Dragon dataset coarse registration results.

The same set of experiments was also conducted using the scans of the Dragon dataset. The results are given in Figs.3.4(a) - 3.4(c). The inlier percentage of correspondences in this dataset is also low for the selected feature. The results are similar to that of Bunny dataset. Here also the proposed algorithm outperforms the PROSAC and RANSAC algorithms in terms of accuracy of the estimates. It is comparable to LoSAC in the case of accuracy but performs better in terms of execution time. The execution time of the proposed algorithm is also comparable to that of PROSAC and even outperforms it in some instances.

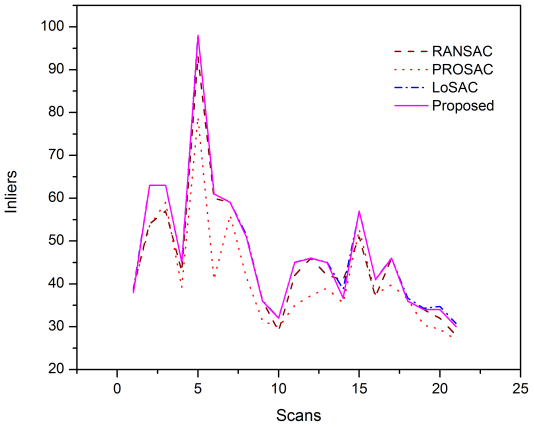
Scans of 3D models from the UWA Dataset were also used for evaluating the proposed algorithm. The transformation validation score which indicates the average error per corresponding pair is reported. The unit of this score is in accordance with the unit of the real world coordinates of the point cloud used. Since this is not mentioned in the dataset, it is not indicated in the results.



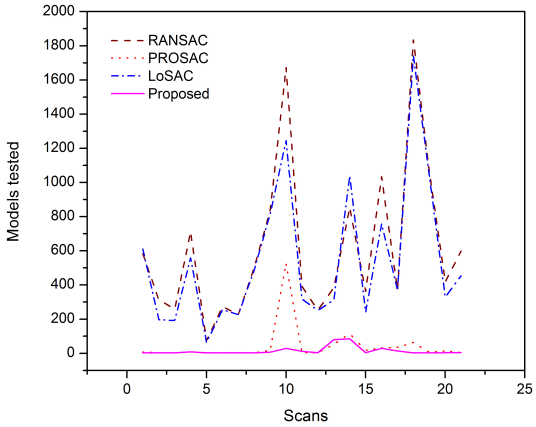
(a) Transformation validation score



(b) Time taken for RANSAC iterations in ms



(c) Number of inliers

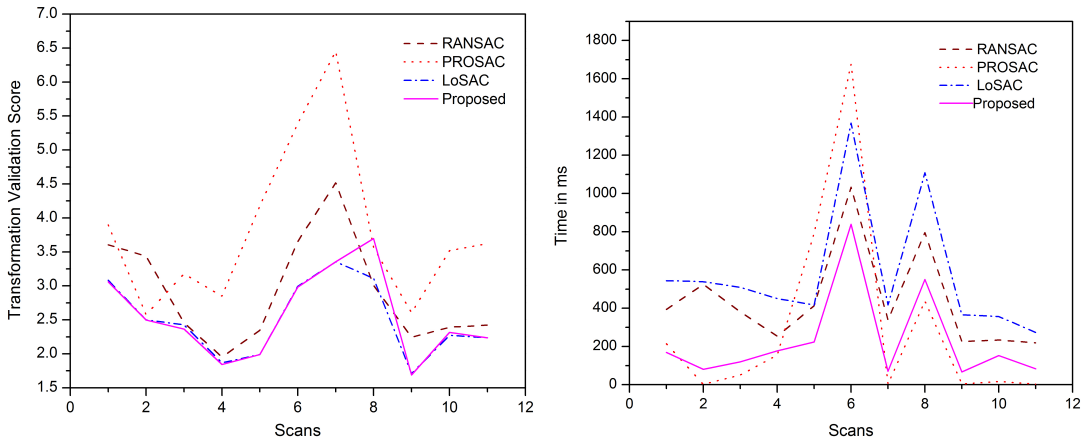


(d) Number of models tested

Figure 3.5: UWA Chef dataset coarse registration results.

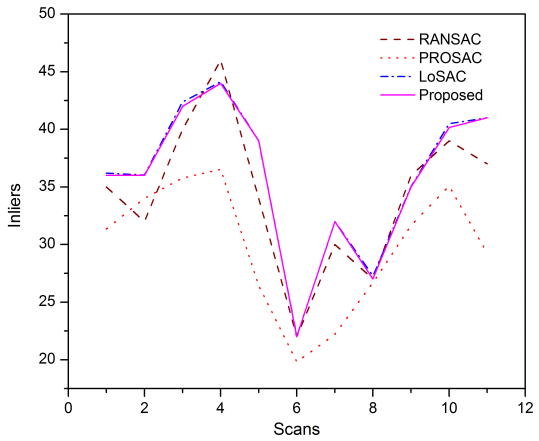
Fig.3.5(a) shows the transformation validation scores of the different algorithms on the Chef dataset scan pairs. It is observed that the proposed algorithm obtains the minimum error compared to other algorithms. The time taken by the proposed algorithm is much less than LoSAC and RANSAC and slightly more than PROSAC, in most cases, as in Fig.3.5(b). For one instance, RANSAC is faster than the proposed algorithm, but the estimation error for RANSAC is higher. The number of inliers found out by the proposed algorithm is the highest and is comparable to that of LoSAC, as given by Fig.3.5(c). The number of model hypotheses tested by the algorithms is also reported in Fig.3.5(d) and the results indicate that this is lowest for the proposed algorithm. The models tested inside the local optimization stage are not included. However, the numbers of models tested by LoSAC, without including the inner models are also much higher than that the proposed algorithm.

Similar experiments were conducted on the scan pairs from two other models of the UWA dataset - Chicken and Parasaurolophus models. The results obtained for Chicken dataset are shown in Figs.3.6(a) - 3.6(d) and for Parasaurolophus model are shown in Figs.3.7(a) - 3.7(d). The results on these datasets also support the reasoning in the previous cases. There is a case in the Chicken dataset where the accuracy of the proposed approach is lower than that of other algorithms and similar to that of PROSAC. This is attributed to the lower percentage of inliers for the scan pair and the small subset of correspondences selected by guided sampling. In other cases, the proposed algorithm performs consistently well.

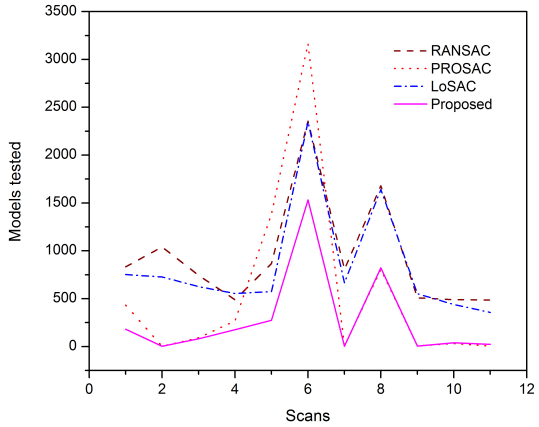


(a) Transformation validation score

(b) Time taken for RANSAC iterations in ms



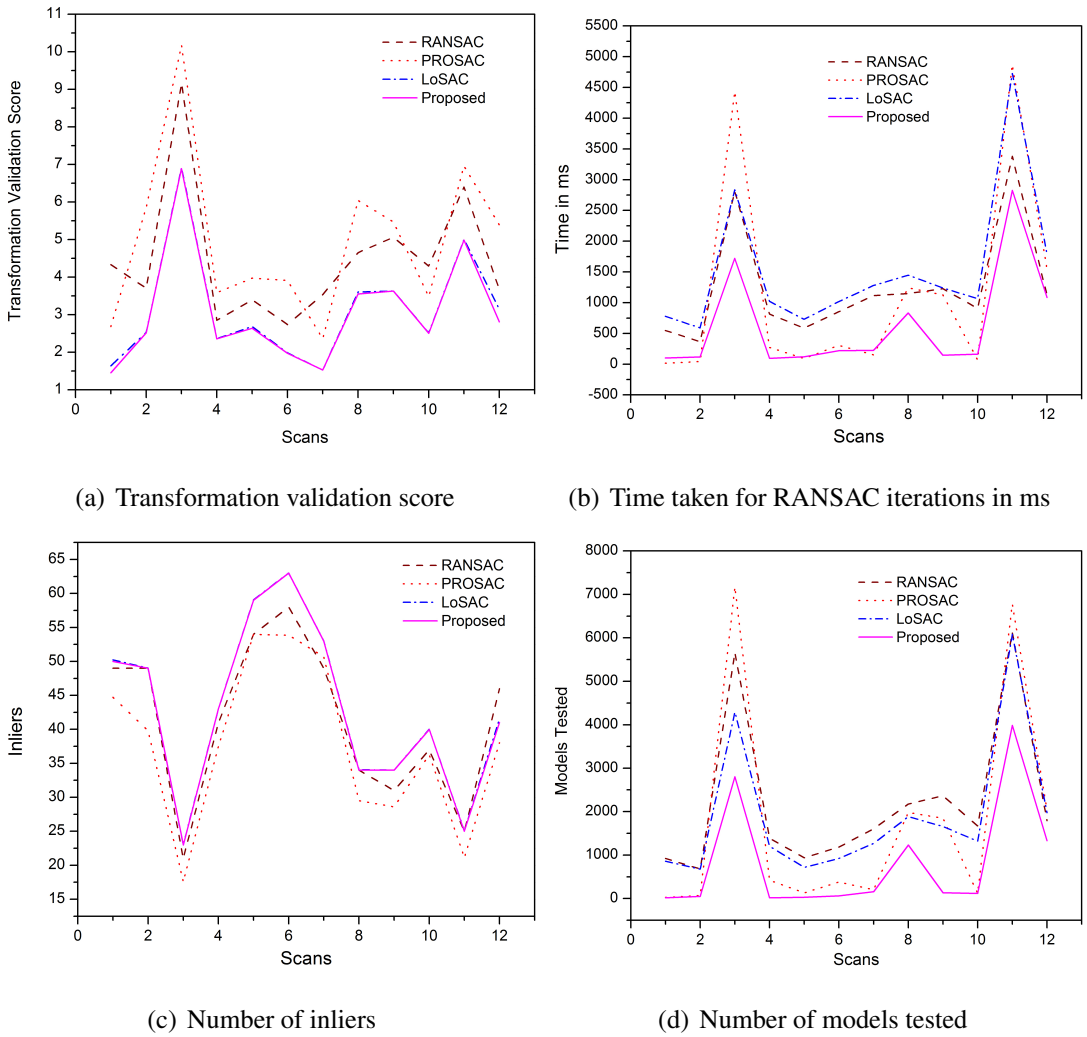
(c) Number of inliers



(d) Number of models tested

Figure 3.6: UWA Chicken dataset coarse registration results.

The results of registration of two sample scan pairs are shown in Figs.3.8 - 3.9. The results obtained by RANSAC and the proposed algorithm are shown for comparison. The figures show that the registration obtained by the proposed algorithm is very accurate compared to RANSAC and its variants. This helps in reducing the time taken by the fine registration stage to achieve convergence. Accurate results help the subsequent stages to converge to a precise solution.



(a) Transformation validation score

(b) Time taken for RANSAC iterations in ms

(c) Number of inliers

(d) Number of models tested

Figure 3.7: UWA Parasaurolophus dataset coarse registration results.

The unregistered point clouds and the clouds after coarse registration by the proposed method are shown in figures Figs.3.10 - 3.15. From this one can see that the proposed algorithm is able to achieve a coarse registration of scan pairs at arbitrary initial positions. The results are not limited by the extent of initial transformation since the correspondences are estimated by feature matching where the features are rigid body motion invariant. The subsequent costly fine registration stage can be accelerated by the accurate results achieved in the coarse registration stage.

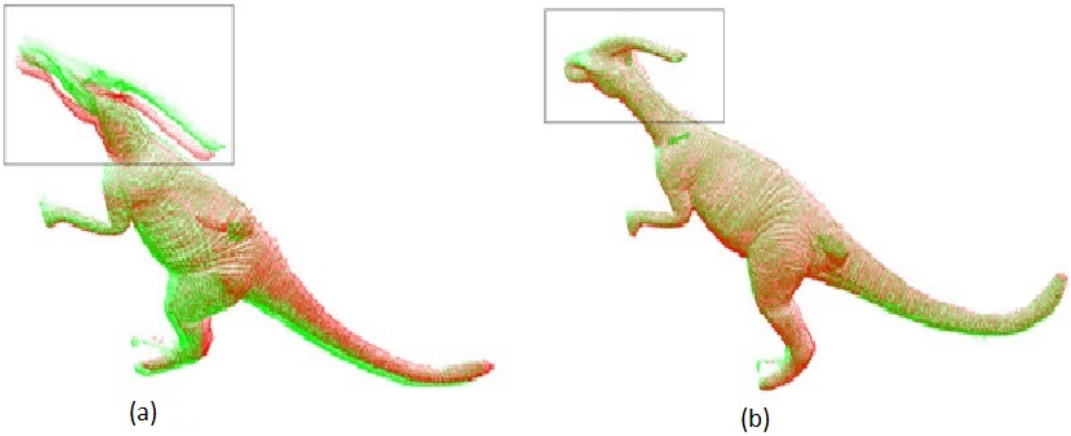


Figure 3.8: Registration results of Parasaurolophus model scan pair by a) RANSAC, and b) proposed method.

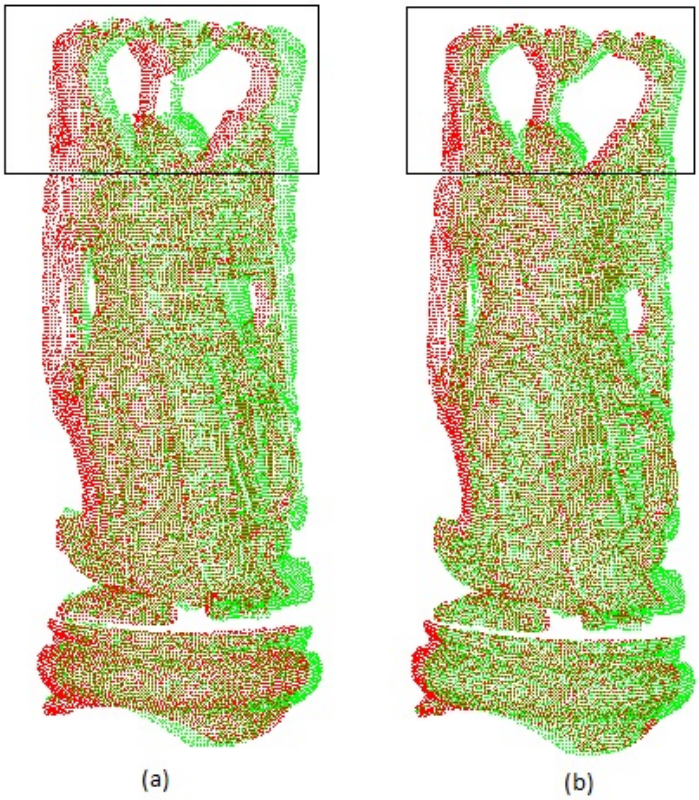


Figure 3.9: Registration results of Buddha model scan pair by a) RANSAC, and b) proposed method.

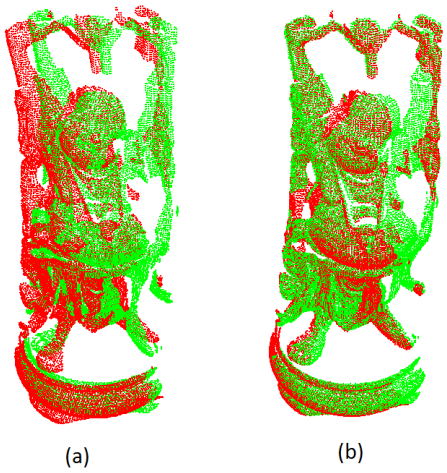


Figure 3.10: Buddha dataset scan pair a) before registration, and b) after initial registration.

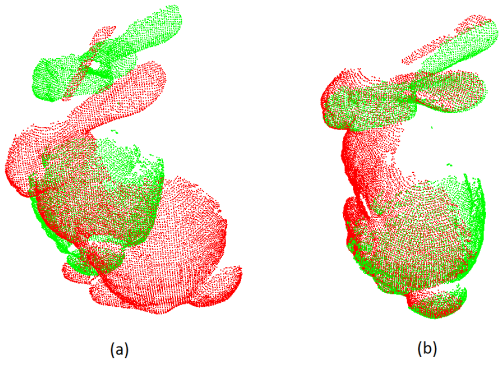


Figure 3.11: Bunny dataset scan pair a) before registration, and b) after initial registration.

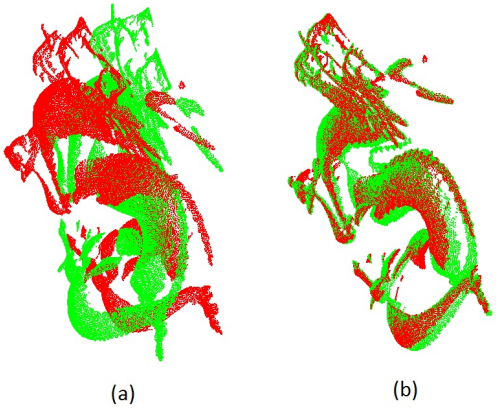


Figure 3.12: Dragon dataset scan pair a) before registration, and b) after initial registration.

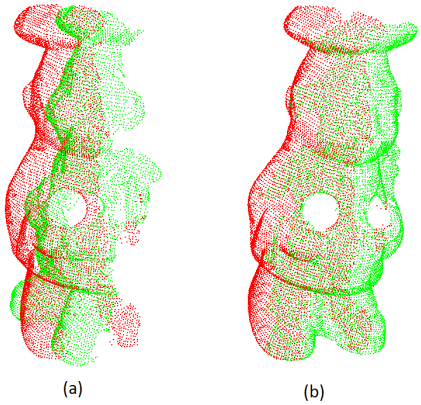


Figure 3.13: Chef dataset scan pair a) before registration, and b) after initial registration.

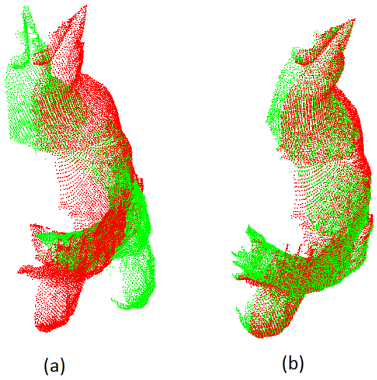


Figure 3.14: Chicken dataset scan pair a) before registration, and b) after initial registration.

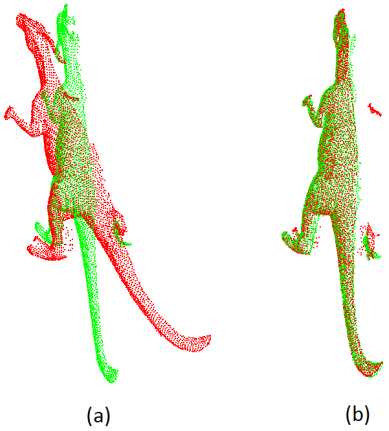


Figure 3.15: Parasaurolophus dataset scan pair a) before registration, and b) after initial registration.

3.6 Discussion

In this work, a novel robust transformation estimation algorithm is developed and tested on different models. From the results obtained, it is observed that the proposed algorithm (Progressive sampled Locally Optimised RANSAC -ProLoSAC) outperforms the existing approach RANSAC algorithm in terms of accuracy as well as computational time. The various observations from the tests conducted are detailed here.

The guided sampling effectively utilizes the correspondence match scores from the 3D feature matching in the proposed approach. Utilising the probably best correspondence pairs in early iterations helps in reducing the number of models tested and thus results in less computational time. This is evident from the various results discussed. The algorithms which utilize guided sampling, like PROSAC (Chum et al., 2003) and the proposed algorithm, achieves the best running time. However the results obtained by PROSAC are less accurate compared to the other algorithms. This can be attributed to the fact that many possibly good correspondence pairs may not be considered by PROSAC for model generation due to the early termination. This drawback is rectified with the help of a local optimization step in the proposed algorithm.

In the local optimization step, the model generation samples are formed from the set of inliers. This sample is also a non-minimal sample, which helps in including many possibly good correspondences for computing the model. Thus the non-minimal samples formed from inliers results in refinement of the transformation. In addition to this, since varying thresholds are used for computing the model in various iterations, a local refinement of the computed transformation is carried out. This results in a more accurate estimation of the transformation model. This is evident from the results obtained. The proposed algorithm achieves the accuracy comparable to that of LoSAC (Chum et al., 2003). LoSAC, however is computationally more intensive which can be seen from the running time which is the highest among the algorithms considered. Random sampling results in a large number of models to be tested and this may lead to many local optimization steps. This extra computational burden is compensated by the guided sampling stage in the proposed method. The most likely correspondences evaluated early in the loop lead to consideration of good models initially. This helps in reducing the number of models tested and hence the running time. The stopping criterion ensures that a good model (i.e formed from inliers) is obtained by the algorithm

and it is not a model supported by random outliers. From the above results, one can see that the proposed algorithm performs the 3D coarse registration task with good accuracy as well as computational time.

The proposed method uses the correspondences found by 3D feature matching. The influence of selected features is in deciding the ratio of inliers to outliers in the corresponding pairs. So the influence of feature selection process can be considered equivalent to the influence of the percentage of inliers. The correctness of 3D feature matching depends upon the geometric distinctiveness of the point cloud surface. If the point cloud has less geometrical features, the ratio of inliers to outliers will be less in the found correspondences. A very low inlier percentage will lead to more computational time requirements by the proposed algorithm. In such cases, other 3D features which make use of the colour or texture information may be used. Also in the case of point cloud pairs where the inlier percentage is very high, the extra computational overhead by the local optimization may be made optional by including additional verification in the algorithm.

The proposed algorithm aims at the coarse registration of point clouds and hence can match partially overlapping point clouds at arbitrary locations. The convergence funnel (Mitra et al., 2004) using the Stanford Bunny model point cloud was examined. The model point cloud was rotated and translated and then registered with the original point cloud. Translations (radially) along the x-z plane up to a distance of 6 times the height of the Bunny and rotations about y axis with 30 degree increments up to 360 degrees were considered. The method achieved global convergence in all cases. Hence the proposed method can be considered as not limited by specific rotation or translation limits but only by the distinctiveness of the features selected for the input point clouds.

3.7 Conclusion

The pair-wise initial alignment of partially overlapping 3D point clouds generally employs 3D feature matching. On account of the possibility of high percentage of outliers in 3D correspondences, robust estimation techniques are often used. A novel approach for dealing with outliers in the set of correspondences for 3D transformation

estimation is presented. The proposed algorithm was implemented for the 3D registration of six different datasets. The performance of the algorithm was compared with the popular algorithm RANSAC and its variants and the results indicate that the proposed algorithm outperforms them in relative domains of time or accuracy. Since fine alignment stage like ICP requires good initial alignment, the results of the proposed algorithm lead to faster and accurate ICP convergence compared to other methods. The proposed method thus finds a balance between accuracy in the estimate and execution time.

3.8 Chapter Summary

In this chapter, pairwise registration of point clouds is discussed. A novel algorithm for coarse registration is proposed. The proposed algorithm finds transformation estimation from a set of corresponding points corrupted with outliers. In comparison with existing methods, the performance of the proposed approach is superior in terms of accuracy and computational time. The next chapter deals with the multiview registration of point clouds to align all the scans of an object into a common coordinate system.

CHAPTER 4

Robust Multiview Registration of Point Clouds

Prelude: This chapter discusses the multiview alignment of point clouds for 3D modelling. A multiview alignment stage is required after pairwise registration stage to minimise the accumulation of errors by sequential registration. A novel robust algorithm for multiview registration by motion averaging which makes use of the Lie group structure of rigid body transformation matrices is presented. The outliers in the set of relative transformations from pairwise registration stage are filtered out by the proposed approach before performing averaging .[1]

4.1 Introduction

3D modelling of an object from 3D scans requires all the partial scans to be brought into a common coordinate system. The transformations between partially overlapping pairs of scans are estimated by feature matching as detailed in the previous chapter. In this chapter, multiview registration, which utilizes the pairwise transformations, to align all scans to a single coordinate system, is explained. This thesis presents a novel robust algorithm for multiview registration utilising the Lie group structure of rigid body transformations.

Global transformation indicates the transformation from each of the scans to a common coordinate system. Without loss of generality, the coordinate system of the initial scan view can be considered as the common coordinate system. The relative transformations between the overlapping pairs of scans can be estimated by pairwise registration methods. The multiview registration is often preceded by a pairwise registration stage. In some cases, the pairwise registration stage is embedded within multiview registration.

Some multiview approaches identify the overlapping pairs from a set of scans be-

[1]Major contents of this chapter are published in *IET Computer Vision* Vol: 11 (1), 2017. Authors: Dhanya S Pankaj, Rama Rao Nidamanuri

fore performing registration. However, most of the approaches assume that the overlapping pairs are known as prior information. Once the pairwise transformations are obtained, this information can be used for performing multiview (global) registration. A straight forward approach to achieve global registration is to register the scans pairwise and propagate the relative transformations sequentially. However, this is a local approach and results in accumulated errors (Benjemaa and Schmitt, 1998). Another approach is to make use of an intermediate shape model. Chen and Medioni (1992) used a growing model to which all subsequent views were merged. However, in this case, the information in a subsequent view cannot improve the registration of a previously registered view (Bergevin et al., 1992). To deal with this, simultaneous registration methods which consider all the scan views simultaneously were proposed (Neugebauer, 1997b; Benjemaa and Schmitt, 1998; Pulli, 1999). Some approaches modelled the problem using dynamic systems where imaginary springs were considered between corresponding points (Eggert et al., 1996). Multiview registration could be modelled as non-linear optimization problem and various approaches were proposed to solve this (Williams and Bennamoun, 2001; Neugebauer, 1997a; Lawson and Hanson, 1995; Benjemaa and Schmitt, 1998). View graph based approaches are also a popular method of choice for global registration (Huber and Hebert, 2003; Neugebauer, 1997b; Mian et al., 2004; Kümmerle et al., 2011). The manifold structure of the 3D rotation and 3D rigid body transformation matrices were also utilised to solve the global registration (Krishnan et al., 2005; Mateo et al., 2014; Shih et al., 2008). A detailed review of the multiview registration techniques is available in section 2.2.

Cycles are often observed in turn-table sequences and robotic paths where identity transformation is expected between last and first scans. However this rarely occurs due to the accumulation of errors. This loop closing constraint was utilized to distribute the errors smoothly among the views (Borrmann et al., 2008b; Sprickerhof et al., 2009; Sharp et al., 2004; Shiratori et al., 2015; Lu and Milios, 1997; Matabosch et al., 2008; Weise et al., 2011). The cycle minimization approaches were categorized into analytical and statistical approaches by Batlle et al. (2007). Shih et al. (2008) broadly classified the multiview registration approaches into three viz. direct multiview, intermediate shape reconstruction, and minimising error between coordinate frames. Among these, the methods which minimize the error between coordinate frames (Krishnan et al., 2005; Govindu, 2004) have the best time and space complex-

ity though the goal of error spreading may not match that of registration (Shih et al., 2008). Govindu and Pooja (2014) iterated the registration and motion averaging approaches to deal with this issue by presenting an extension of ICP to multiple views. They made use of the Lie group structure of 3D rigid body transformations and the availability of redundant overlaps (not just the adjacent scans overlap). Their algorithm alternately iterated the correspondence estimation and motion averaging steps. The additional constraints helped to average out the errors than to simply distribute them. However, they assumed the input relative transformations to be correct and the results were affected in presence of outliers, since averaging could not manage outliers. The attempts in literature to deal with these outliers can be classified into two categories. The first approach is to identify the outliers beforehand and exclude them from the averaging step (Govindu, 2006). The second approach is to use a robust averaging technique (Chatterjee et al., 2013). The first approach was followed by Govindu (2006) and in the present study.

A graph-based sampling scheme using RANSAC was presented by Govindu (2006) to identify the inliers. The view-graph was sampled to obtain the minimum spanning trees (MST) and global motions were computed. The MST which provided the maximum consistency between the relative motions and global motions was selected. All the relative motions which agreed to the computed global motion were selected as inliers. The limitation of this approach is the exhaustive search of the view graph required for obtaining a consistent MST. Also, in the case of a large graph, the convergence can be slow since the number of RANSAC iterations required depends on the percentage of inliers, sampling set size etc.

The motion averaging algorithm assumes a good initialization of global motions and it can be argued that the relative motions which are consistent with it can be estimated in a single step. This thesis proposes a graph-based approach to identify such inlier relative motions. The proposed solution is better than the existing approach (Govindu, 2006) because the inliers can be obtained in a single step instead of performing an exhaustive search of the view graph. Since the inliers can be found in advance, the need for a costly L1 averaging step proposed by Chatterjee et al. (2013) can be avoided.

Algorithm 4 Proposed multiview registration algorithm.

▷ Input: Set of Relative Motions, Output: Set of Global Motion Estimates

1. Form the graph $G = (V, E)$ with $V = \{\text{scan views}, S_1 : S_N\}$ and $E = \{\text{relative motions}, M_{ji}\_k, k = 1 : n\}$

2. Find an initial estimate of the global motions M_i, $i = 1 : N$

3. Find the edge weights for the relative motion estimates

4. Find the shortest path from each of the scans to the initial scan S_1 by Dijkstra's Single Source Shortest Path (SSSP) algorithm.

5. Find the global motion estimate of each of the scans by sequentially registering along the shortest path found.

6. Identify the inliers among the relative motion estimates

7. Perform motion averaging to refine the global motion estimates using the selected relative motion estimates (Algorithm 5). The edge weights are updated in each iteration of the averaging and the global motions are estimated by the Dijkstra's SSSP algorithm

4.2 Method

In this work, global 'motion' refers to the global transformation and relative 'motion' refers to the relative transformation. The global estimates M_i represent the global transformation from each of the scans $S_i, i = 1 : N$ to the initial scan S_1. The relative motion estimate M_{ji} represents the 3D transformation from scan j to scan i. Assume that there are n relative motions. Let $G = (V, E)$ be the view graph where the vertices V represent the N partial 3D views. An edge in E represents the availability of a relative motion estimate between the vertex pair. The goal is to identify the correct relative motion estimates. The global motions are initialized according to the un-weighted view graph (section 4.2.1). Weights are assigned to the relative motion edges by checking their compatibility with initial global motion estimates and other relative motion estimates with common scan views (section 4.2.3). Once the edge weights are assigned, the shortest paths from each of the scans to scan S_1 are found using Dijkstra's single source shortest path algorithm (SSSP) (Joyner et al., 2010). Then, the inlier edges are identified (section 4.2.4) and motion averaging is performed using the inliers. The steps involved in the proposed algorithm are detailed in Algorithm 4.

4.2.1 Initialization of global motion estimates

The global motions can be initialized in a number of ways. The straight forward method is to sequentially register the scans from S_1. However, the errors will get accumulated towards the end. In order to minimize the propagation of errors at least by half, this work proposes to register the scan pairs in two directions if the relative motion from scan N to the first scan is available. The scan pairs from S_1 to $S_{N/2}$ are registered sequentially from the first scan. The scans from S_N to $S_{N/2+1}$ are registered in the reverse direction starting from S_1 through S_N (Eq.4.1). Here accumulation of errors towards the end of the scans is reduced. However, there will be a shift near to $S_{N/2}$, which will be usually less compared to the error towards S_N in sequential registration. M represents the transformation matrix $M = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$ where R is the 3D rotation matrix and t, the 3D translation vector. M_1 is the identity transformation and the global motion from second scan view to the first (M_2) is same as the direct relative transformation M_{21}.

$$\begin{aligned} M_{N-1} &= M_{NN-1}^{-1} * M_{N1} \\ M_j = M_i * M_{ji}; \quad i < j;\, j = 2 : N/2;\, i = 1 : (N/2-1) \\ M_j = M_i * M_{ij}^{-1}; \quad j = N-1 : (N/2+1);\, i = N : (N/2+2) \end{aligned} \tag{4.1}$$

A graph-based initialization can also be done when an error measure which indicates the quality of pairwise alignment is available. A weighted view graph can be formed with relative registration errors as weights. Then Dijkstra's SSSP algorithm can be used to find the shortest paths and global motions can be initialized by sequentially registering along the path. This will result in a good initialization when the error measure correctly points to the quality of registration. This method of registration is particularly useful when scans, which are not part of a turn table sequence, nevertheless have overlap with some of the scans along that sequence, are present. In this case, an un-weighted graph can also be used for initialization.

4.2.2 Mathematical background

Before discussing the proposed weight calculation method for relative motion edges, an elementary discussion on Lie Groups is presented here (Blanco, 2010; Kanatani, 2012; Fletcher et al., 2003).

Lie Group

A Lie group is a non-empty subset G of $\mathbb{R}^N$ that satisfies the conditions (Blanco, 2010)

- G is a group (Kanatani, 2012)
- G is a manifold in $\mathbb{R}^N$
- Both the group operation ($\cdot : G \rightarrow G$) and its inverse ($^{-1}$: $G \rightarrow G$) are smooth functions.

Manifold

A $D-$dimensional manifold M is a topological space where every point $p \in M$ is endowed with local Euclidean structure. A $D-$dimensional manifold embedded in $\mathbb{R}^N (N >= D)$ has an associated $N-$dimensional tangent space for every point $p \in M$. This space can be denoted by $T_x M$ and in non-singular points has a dimensionality of D (identical to that of the manifold).

Lie Algebra

A Lie algebra is an algebra $\mathfrak{m}$ together with a Lie bracket operator $[\cdot, \cdot] : \mathfrak{m} \times \mathfrak{m} \rightarrow \mathfrak{m}$ such as for any elements $a, b, c \in M$ it holds Eq.4.2. The Lie algebra $\mathfrak{m}$ associated to a Lie Group M happens to be the tangent space at the identity element I, $\mathfrak{m} = T_I M$.

$$\begin{aligned} [a, b] &= -[b, a] \qquad \text{(Anti - commutativity)} \\ [c, [a, b]] &= [[c, a], b] + [a, [c, b]] \qquad \text{(Jacobi identity)} \end{aligned} \tag{4.2}$$

Exponential and Logarithmic Maps

These maps provide a way to map the elements of the Lie group and the corresponding Lie algebra elements. The exponential map maps elements from the Lie algebra to the manifold $\exp : \mathfrak{m} \rightarrow M$ and the logarithmic map maps the elements from the manifold to the Lie algebra. $\ln : M \rightarrow \mathfrak{m}$

Linear Lie Groups or Matrix Groups

The Lie groups that are most commonly used in computer vision are the matrix groups (Fletcher et al., 2003). The matrix groups are the subgroups of the general linear group $\mathbf{GL}(n, \mathbb{R})$, the group of non-singular $n \times n$ real matrices. The Lie algebra associated with $\mathbf{GL}(n, \mathbb{R})$ is $\mathbf{L}(\mathbb{R}^n, \mathbb{R}^n)$, the set of all $n \times n$ real matrices. The Lie bracket is the commutator, $[X, Y] = XY - YX, \quad (X, Y \in \mathbf{L}(\mathbb{R}^n, \mathbb{R}^n))$

The exponential map of a matrix $X \in L$ is the standard matrix exponent $\exp(X) = \sum_{k=0}^{\infty} \frac{1}{k!} X^k$. The group of proper rotations $\mathbf{SO}(3)$ in $\mathbb{R}^3$ and the group of rigid transformations $\mathbf{SE}(3)$ are both Linear Lie groups (Blanco, 2010). Thus both have associated Lie algebra. The matrix Lie group also has the structure of a Riemannian manifold (Duan et al., 2013). In this work, the group of rigid transformations $\mathbf{SE}(3)$ are considered.

$\mathbf{SE}(3)$ as a Lie Group

The group of rigid transformations in $\mathbb{R}^3$, denoted as $\mathbf{SE}(3)$ are the set of 4×4 matrices of the form in Eq.4.3 with standard matrix product as the group operation.

$$\mathbf{SE}(3) = \left\{ \begin{bmatrix} A & b \\ 0 & 1 \end{bmatrix} \mid A \in \mathbf{SO}(3),\, b \in \mathbb{R}^3 \right\} \tag{4.3}$$

The Lie algebra $\mathfrak{se}(3)$ of $\mathbf{SE}(3)$ can be denoted by Eq.4.4.

$$\mathfrak{se}(3) = \left\{ \begin{bmatrix} \Omega & v \\ 0 & 0 \end{bmatrix} \mid \Omega^T = -\Omega,\, v \in \mathbb{R}^3 \right\} \tag{4.4}$$

The skew-symmetric matrix Ω can be uniquely expressed as $\begin{pmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{pmatrix}$, where $\omega = (\omega_x, \omega_y, \omega_z) \in \mathbb{R}^3$. $\|\omega\|_F$ gives the magnitude of rotation with respect to the unit vector along ω where $\|\cdot\|_F$ represents the Frobenius norm. ω represents the angular velocity and v represents the linear velocity of the object (Duan et al., 2013). The exponential map from $\mathfrak{se}(3)$ to $\mathbf{SE}(3)$ is given by Eq.4.5 where $V \in \mathfrak{se}(3)$ and $\theta^2 = \omega_x^2 + \omega_y^2 + \omega_z^2$.

$$\exp(V) = I_4 + V + \frac{1 - cos(\theta)}{\theta^2}V^2 + \frac{\theta - sin(\theta)}{\theta^3}V^3 \tag{4.5}$$

The logarithmic map from $\mathbf{SE}(3)$ to $\mathfrak{se}(3)$ is given by Eq.4.6 where $Q \in \mathbf{SE}(3)$.

$$\log(Q) = q_1(q_2 I_4 - q_3 Q + q_4 Q^2 - q_5 Q^3), \tag{4.6}$$

where,

$$q_1 = \frac{1}{8}\csc^3\left(\frac{\theta}{2}\right)\sec\left(\frac{\theta}{2}\right)$$
$$q_2 = \theta\cos(2\theta) - \sin(\theta),$$
$$q_3 = \theta\cos(\theta) + 2\theta\cos(2\theta) - \sin(\theta) - \sin(2\theta),$$
$$q_4 = 2\theta\cos(\theta) + \theta\cos(2\theta) - \sin(\theta) - \sin(2\theta),$$
$$q_5 = \theta\cos(\theta) - \sin(\theta),$$
$$tr(Q) = 2 + 2\cos(\theta), for - \pi < \theta < \pi.$$

BCH formula

For non-commutative Lie groups, the exponential relation $\exp^x \exp^y = \exp^{x+y}$ does not hold and an equivalent definition is given by Eq.4.7.

$$\exp^x \exp^y = \exp^{BCH(x,y)} \tag{4.7}$$

where BCH is the Baker-Campbell-Haursdorff (BCH) formula given by Eq.4.8

$$BCH(x,y) = x + y + \frac{1}{2}[x,y] + \frac{1}{12}[x-y,[x,y]] + \mathcal{O}(|(x,y)|^4) \tag{4.8}$$

Distances in Lie Group

The distance between two elements of a Lie group G can be either intrinsic or extrinsic (Fletcher et al., 2003). The extrinsic distance is obtained by embedding it in a Euclidean space and using the induced Euclidean distance. The natural choice of distance is the intrinsic distance, the Riemannian distance on G. The Riemannian distance between two points is the length of the shortest geodesic curve between the points (Fletcher et al., 2003). The Riemannian distance between two elements of a matrix Lie Group G is defined by the matrix logarithm operation. For any two $\mathbf{SE}(3)$ elements X and Y, the distance is defined by Eq.4.9 (Govindu, 2006) where $\log$ refers to the logarithmic map that maps Lie group to the corresponding Lie algebra.

$$d(X,Y) = \| \log(YX^{-1}) \| \tag{4.9}$$

By using the BCH formula Eq.4.8, this distance can be approximated as in Eq.4.10 where x and y are the corresponding Lie algebra elements of X and Y. $\|.\|$ is the Frobenius norm.

$$d(X,Y) = \|log(YX^{-1})\| \approx \|log(Y) - log(X)\| = \|y - x\| \tag{4.10}$$

Thus by tangent space approximation, the Riemannian distance between two elements of $\mathbf{SE}(3)$ group is now given by the 'Euclidean distance' in its Lie algebra (Govindu, 2006). The motion averaging algorithm in (Govindu and Pooja, 2014) utilizes this definition of distance between $\mathbf{SE}(3)$ motion matrices to outline an algorithm for averaging absolute motions as well as relative motions.

4.2.3 Weight calculation

The relative motion estimates are weighted based on their agreement with respect to the global motion estimates as well as with other relative motion estimates. Con-

sider another graph $G_1 = (V_1, E_1)$ where the vertices represent the different relative motion estimates (M_{ji}) and the edges represent the connection between vertices if there is a common scan view in both the relative motion estimates. i.e. if (M_{ji}) and (M_{lm}) are the two relative motion estimates, then there is an edge connecting them if $j = m$ or l or $i = m$ or l.

The distance between the relative motions computed from the global motions (Eq.4.11) and the actual relative motion estimates M_{ji} is computed and assigned as vertex weight of G_1 (v_i). This distance is given by Eq.4.12. The vertex weights indicate the agreement of the calculated relative motion with the initial global estimates. Lesser weights indicate better agreement as the weight is a measure of distance or error.

$$\tilde{M_{ji}} = M_i^{-1} * M_j \tag{4.11}$$

$$d(\tilde{M_{ji}}, M_{ji}) = \|log(\tilde{M_{ji}}) - log(M_{ji})\| \tag{4.12}$$

This is the first order linear approximation of the Riemannian distance between the motion matrices. The correctness of a relative motion estimate is further affirmed by checking its compatibility with other relative motion estimates. Two relative motion estimates can be compared if there is an edge between them, i.e. if they involve a common scan view. The global motion estimate of the common scan view is computed from both of the relative motion estimates using one of the equations in Eq.4.13. Then the two computed global estimates of the common scan view are compared by finding the distance between them using Eq.4.10. The computed distance is assigned as the edge weight e_j.

$$\begin{aligned} M_j &= M_i * M_{ji} \\ M_i &= M_j * M_{ji}^{-1} \end{aligned} \tag{4.13}$$

Once the edge weights (e_j) are computed, the vertex weights (v_i) are updated to v_i' as in Eq.4.14 where n_i is the number of edges at vertex i.

$$v_i' = \frac{1}{n_i} * \sum_{j=1}^{n_i} (v_i * e_j) \tag{4.14}$$

The average of the edge weights at each vertex is calculated as all vertices may not be having the same number of edges and this depends on the input (edges in the view graph) data. The average edge weights are multiplied by the initial vertex weights as the compatibility with global motion estimates is more indicative than the compatibility with other relative motion estimates which may contain outliers as well.

4.2.4 Calculation of inliers

Once the weights for the relative motion estimates are obtained, this can be assigned to the view graph $G = (V, E)$ where V represents the scan views and E, the edges. The weights computed in section 4.2.3 act as the edge weights. The weights assigned to the outlier edges will be high because of their disagreement to the global motion estimates as well as to the other relative motion estimates whereas the weights assigned to the inlier edges will be less. Now from the view graph, the shortest path to the scan S_1 is computed from all of the scan views using Dijkstra's SSSP algorithm. Since the shortest paths are computed, the outlier relative motions which have higher assigned weights will be excluded from the paths. Once the shortest paths are obtained, the global motion estimates can be calculated by registering sequentially along the computed path using the relative motion estimates.

When the global motions are obtained, the inliers to the global motions are found out. The distance between the actual relative motions and the relative motions obtained from the updated global motions (as in Eq.4.11) are computed using the Riemannian distance measure in Eq.4.12. This distance can be compared against a threshold (computed empirically) to identify the inliers.

4.2.5 Motion Averaging

Once the inlier relative motions are identified, the global motions are estimated from them using motion averaging. The estimation of global motions is presented as a multiview extension to the ICP algorithm by Govindu and Pooja (2014). The correspondence and relative motion estimation stages of ICP are combined with motion averaging in an iterative algorithm. The steps involved (reproduced from Govindu and Pooja (2014); Li et al. (2014)) are briefly outlined in Algorithm 5.

Algorithm 5 Motion Averaging

1. Compute the incremental update to the relative motions using ICP and update the relative motions
2. Initialize global motions from the relative motions (recalculate registration path)
3. Motion averaging

   $\Delta M_{ji} = M_i^{-1} * M_{ji} * M_j$

   $\Delta m_{ji} = \log(\Delta M_{ji})$

   $\Delta v_{ji} = vec(\Delta m_{ji})$

   $\Delta \mathfrak{V} = D^{\dagger} \Delta \mathcal{V}_{ji}$

   $\forall k \in [2, N],\ M_k = \exp(\Delta m_k) * M_k$

   $Repeat\ until \|\Delta \mathfrak{V}\| < \varepsilon$
4. Update relative motions from global motions $M_{ji} = M_i^{-1} * M_j$
5. Repeat steps 1 to 4 until convergence

Δm_{ji} indicate the corresponding Lie algebra element of ΔM_{ji}, and since Δm_{ji} is a skew-symmetric matrix, it can be uniquely expressed by a 3-element vector given by Δv_{ji} which is obtained by the operation $vec(.)$. The relation in Eq.4.15 is derived from Eq.4.10, where $\mathfrak{V}$ is obtained by stacking the vectors v_i in a column. $D_{ji} = [\ldots, -\mathbf{I}_{6\times6}, \ldots, \mathbf{I}_{6\times6}]$ is a matrix with 6 rows and $6N$ columns, where $-\mathbf{I}_{6\times6}$ appears at $6i^{th}$ column and $\mathbf{I}_{6\times6}$ appears at the $6j^{th}$ column. $\Delta \mathcal{V}_{ji} = [\Delta v_{ji1}, \Delta v_{ji2}, \ldots, \Delta v_{jin}]^T$, $D = [D_{ji1}, D_{ji2}, \ldots, D_{jin}]^T$ and $D^{\dagger}$ is the pseudo inverse of D. The $\log()$ and $\exp()$ operations are as defined in Eqs.4.5 and 4.6.

$$\begin{aligned} m_{ji} &\approx m_j - m_i \\ \implies v_{ji} = v_j - v_i &= [\ldots -\mathbf{I} \ldots \mathbf{I} \ldots] \mathfrak{V} \end{aligned} \tag{4.15}$$

The global motion estimates in step 2 can be estimated by finding out the minimum spanning tree of the view graph (Prim's minimum spanning tree algorithm) (Joyner et al., 2010). The registration path is recalculated at each step to keep it updated. The proposed weight calculation algorithm and Dijkstra's SSSP algorithm is employed for estimating global motion. This is introduced to keep the registration path up to date according to the modified relative and global motion estimates.

4.3 Experiments

Partially overlapping point clouds of 3D models from two different datasets were used for testing the proposed algorithm. The scans of two models from the Stanford dataset (Levoy et al., 2005), acquired using Cyberware 3030 MS scanner - Buddha and Dragon models - and scans of Chef, Chicken and Parasaurolophus models, acquired by Minolta scanner from the University of Western Australia (UWA) dataset (Mian et al., 2010) were used for evaluation. Actual transformations were available for the Stanford models and hence this information was used for evaluating the performance of the algorithms. A set of pairs of overlapping scan views was considered for registration, which may contain outliers. The relative motion estimates of these pairs were obtained by performing the coarse and fine pairwise registration. The coarse alignment of the scans was done using the correspondences obtained by feature matching, as explained in Chapter 3. Keypoint detection algorithm from Zhong (2009) and 3D Shape Context feature extraction algorithm from Frome et al. (2004) were employed for feature matching to obtain correspondences. The proposed algorithm takes the set of pairs and the relative motion of the pairs as input for the inlier identification step. Once the inliers were identified, the motion averaging was performed on the identified inlier set of relative motions.

The registration results of the proposed algorithm are compared with the results obtained with a) sequential registration by ICP, and b) motion averaging without the removal of outliers. The cross sections of the merged clouds are also presented for comparison. The outliers from the input set of relative motions detected by the proposed algorithm were also evaluated for correctness. In the case of Stanford models, the actual transformations were represented as 3D translation units and 3D rotation (as unit quaternions). After global registration, the obtained 3D transformation matrices were converted into x, y, z translation units and the rotation unit quaternions (Diebel, 2006). The norm of the distance between the actual and computed translation units and the dot product between the actual and computed rotation unit quaternions were computed. Lesser the norm of the translation difference, better the estimate of 3D translation. The dot product was used as a measure of similarity between the rotation unit quaternions. The greater the dot product value, the better is the similarity. The difference between the estimated and actual global motions was calculated using the

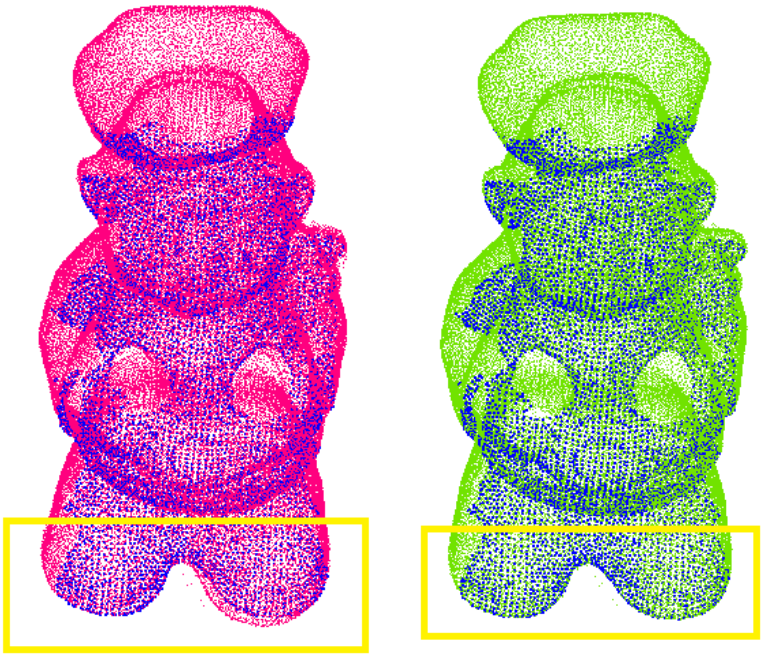
Riemannian distance measure given by Eq.4.12.

The quality of the registration was identified visually from the registered model when actual motion estimate was not available. The cross sections of the models also provide evidence on the quality of registration. The proposed algorithm was tested on different sets of input relative motion pairs. The result for only a single set for each model is shown here due to paucity of space; although the results obtained supported the proposed method. The algorithm was implemented in C++ with the help of PCL library (Rusu and Cousins, 2011).

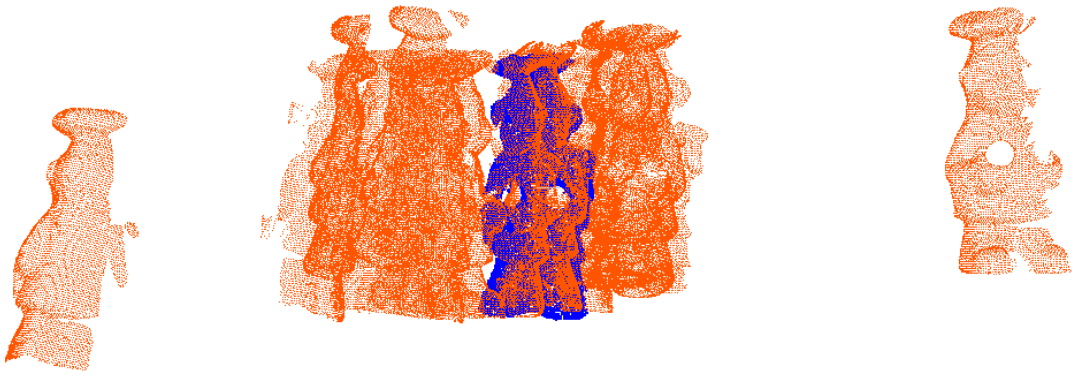
4.4 Results

For each of the 3D models, the scan views that form a turn-table sequence were considered. The set of scan views along with overlapping pairs information and the relative motions of the overlapping pairs were passed as inputs to the multiview algorithm. The global motions were initialized according to the second method (called as Initial II hereafter) in section 4.2.1. The method of sequential multiview registration, based on the pairwise registration results, is referred to as Sequential ICP hereafter in this chapter. The method of registration of the input scan pairs by motion averaging using the entire set of input relative motions is referred to as motion averaging (Govindu and Pooja, 2014) in this chapter.

The registration results of the UWA dataset models Chef, Chicken and Parasaurolophus using a) Sequential ICP, b) proposed method, and c) motion averaging method are shown in Fig.4.1 to 4.3. For clarity, the first scan of the model (in blue) is overlapped with the results. From the result of Sequential ICP method, one can see that the errors had accumulated towards the end and this is highlighted for clarity. The areas affected are marked inside yellow boxes for easy reference. The initial scan is not matching with the registered final scan due to the accumulated error. The motion averaged results show that the outliers corrupted the results to a great extent. The results by the proposed method indicate that the outliers were removed successfully and averaging by the inliers resulted in a properly merged point cloud.

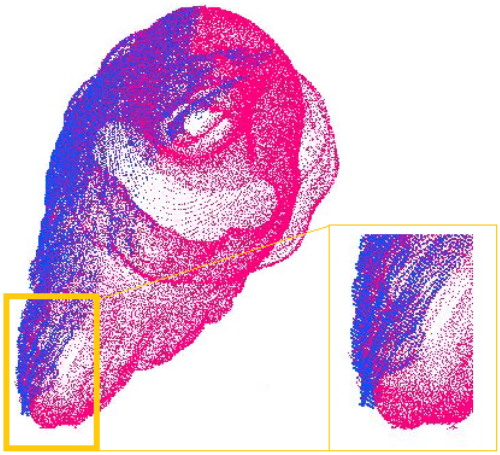


(a) Sequential ICP (b) Proposed method

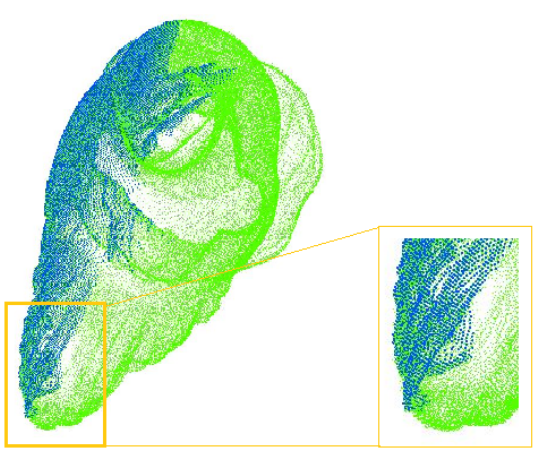


(c) Motion averaging

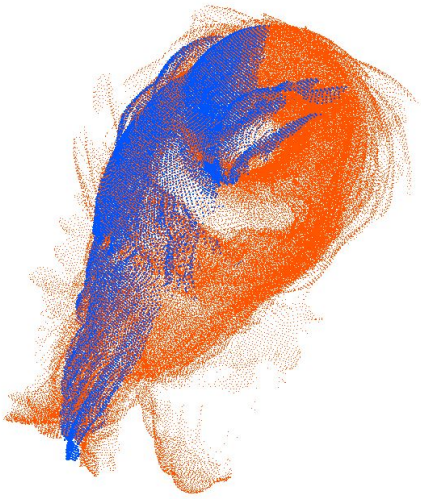
Figure 4.1: Registered models obtained by multiview registration for the UWA Chef model using a) Sequential ICP, b) proposed method, and c) motion averaging. The accumulated error areas in Sequential ICP and corresponding smoothed areas by proposed method are highlighted in yellow boxes. The initial scan (seen in blue colour) is overlapped with the results for clarity.



(a) Sequential ICP

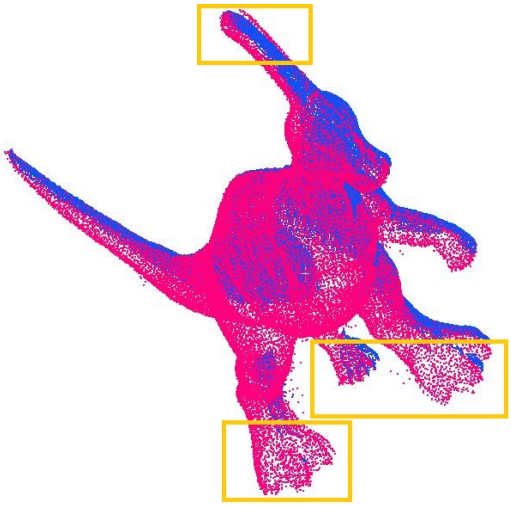


(b) Proposed method

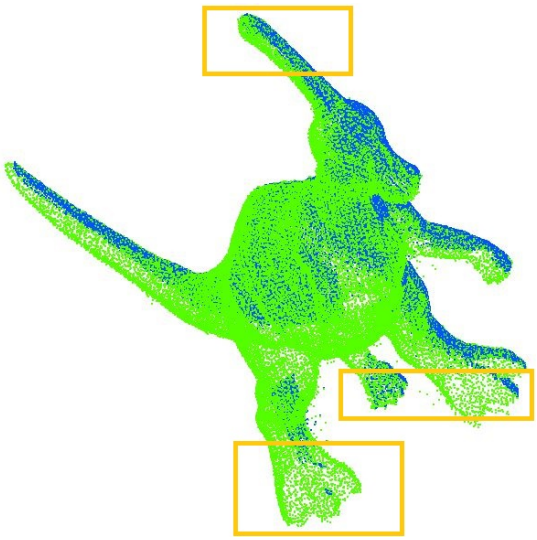


(c) Motion averaging

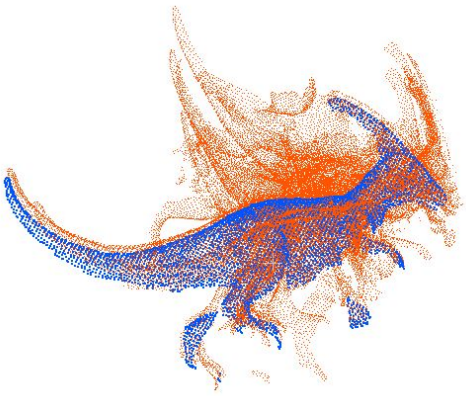
Figure 4.2: Registered models obtained by multiview registration for the UWA Chicken model using a) Sequential ICP, b) proposed method, and c) motion averaging. The accumulated error areas in Sequential ICP and corresponding smoothed areas by proposed method are highlighted in yellow boxes and enlarged for perception. The initial scan (seen in blue colour) is overlapped with the results for clarity.



(a) Sequential ICP



(b) Proposed method

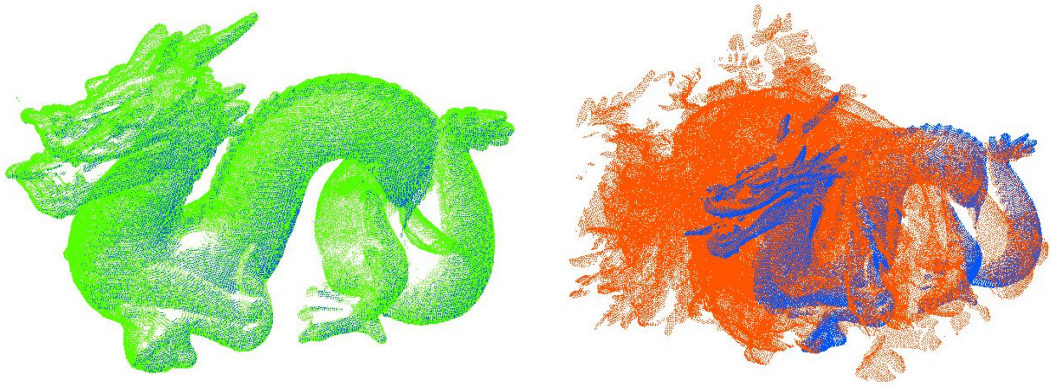


(c) Motion averaging

Figure 4.3: Registered models obtained by multiview registration for the UWA Parasaurolophus model using a) Sequential ICP, b) proposed method, and c) motion averaging. The accumulated error areas in Sequential ICP and corresponding smoothed areas by proposed method are highlighted in yellow boxes. The initial scan (seen in blue colour) is overlapped with the results for clarity.

Multiview alignment was performed on the scan views of the two models of the Stanford dataset. The registration results of the Dragon model obtained by the proposed method and by the motion averaging method without filtering are shown in Fig.4.4. The results by Sequential ICP are not shown as the input relative motions used for Sequential ICP were mostly accurate and the accumulated error was not vi-

sually significant. The registered Buddha model obtained by a) Sequential ICP, b) proposed method, and c) motion averaging are shown in Fig.4.5. The results indicate that the proposed method outperformed Sequential ICP in accuracy by motion averaging the inlier relative motions. Proposed method identifies and removes outliers before averaging thus making the motion averaging robust. The outcomes from registering this model also support the previous argument that the proposed method correctly identifies the inliers in the input relative motions. The initial scan is shown in blue, overlapped with the results, for reference. The motion averaging with outliers failed to produce a good registration.

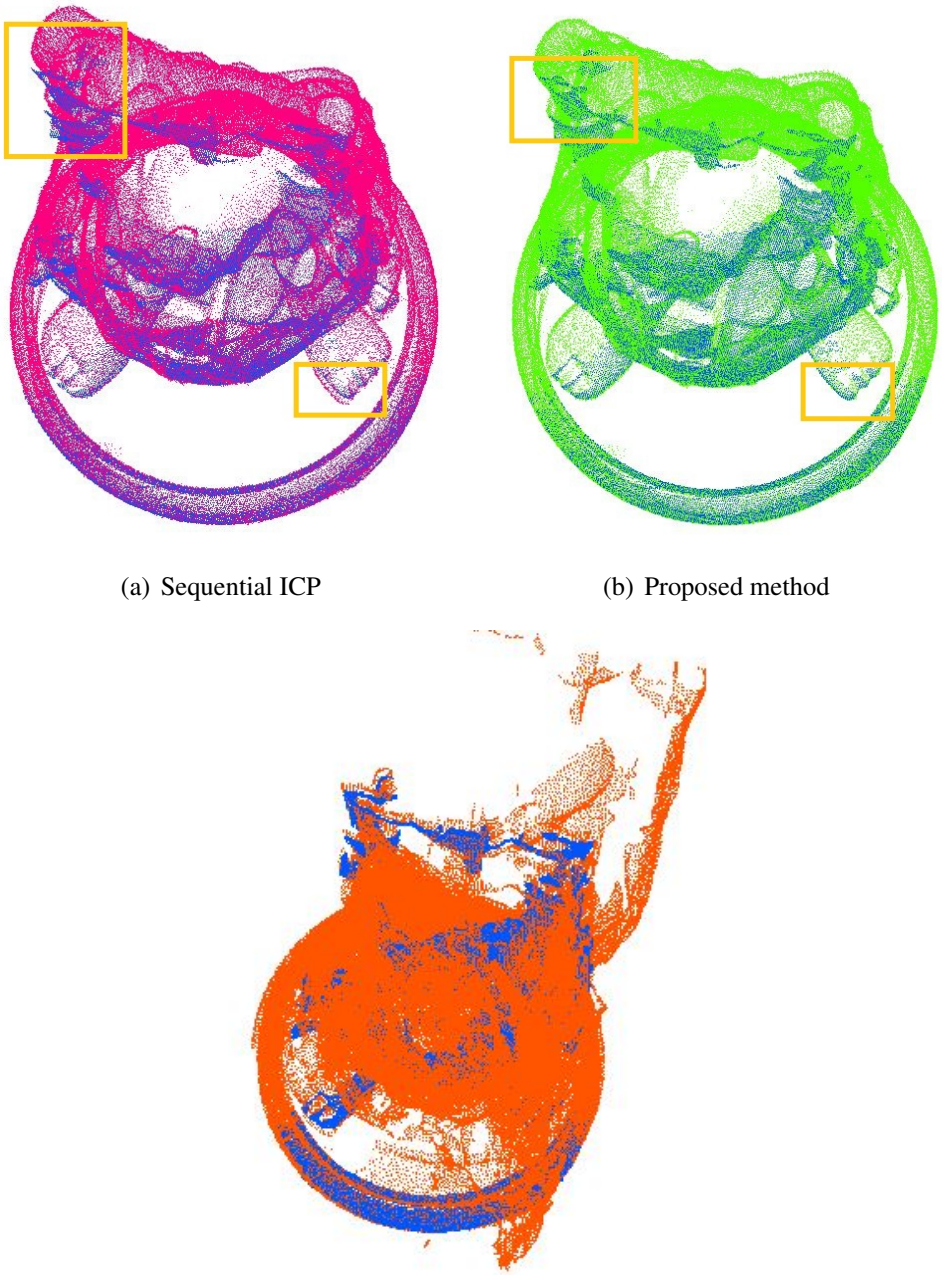


(a) Proposed method (b) Motion averaging

Figure 4.4: Registered models obtained by multiview registration for the Stanford Dragon model using a) proposed method, and b) motion averaging. The accumulated error areas in Sequential ICP and corresponding smoothed areas by proposed method are highlighted in yellow boxes. The initial scan (seen in blue colour) is overlapped with the results for clarity.

The cross sections of the registered models were extracted to highlight the loop closing error by Sequential ICP. The cross sections of the Chef, Chicken, and Buddha models obtained by Sequential ICP and proposed methods are shown in Fig.4.6. It is clear that the loop closing error in Sequential ICP was rectified by the proposed method to result in a smooth cross section. For the Chicken model, the number of inlier redundant scan pairs available for averaging was less and hence the smoothing achieved was also less compared to other models. The smoothing of the loop closing error was possible due to the effect of motion averaging. The proposed algorithm

identified the best relative motions to make the results of motion averaging free from the effect of outliers.

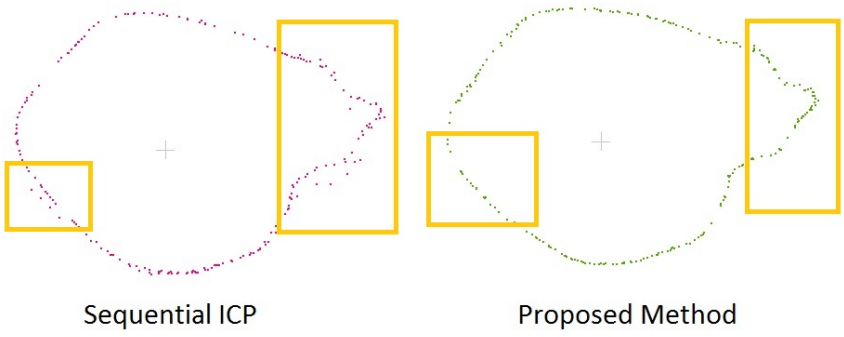


(a) Sequential ICP

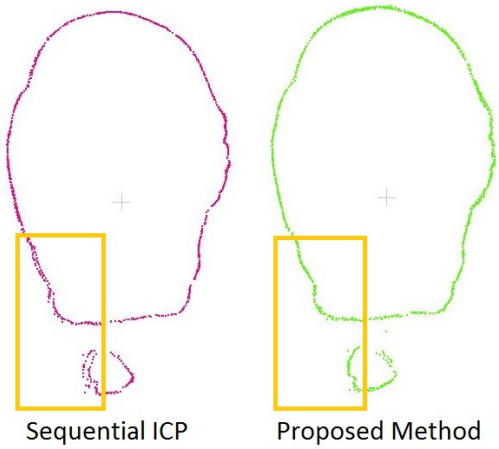
(b) Proposed method

(c) Motion averaging

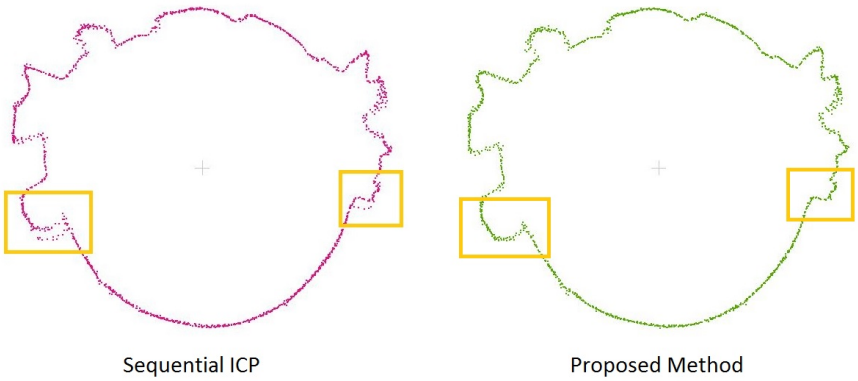
Figure 4.5: Registered models obtained by multiview registration for the Stanford Buddha model using a) Sequential ICP, b) proposed method, and c) motion averaging. The accumulated error areas in Sequential ICP and corresponding smoothed areas by proposed method are highlighted in yellow boxes. The initial scan (seen in blue colour) is overlapped with the results for clarity.



(a) Chef



(b) Chicken

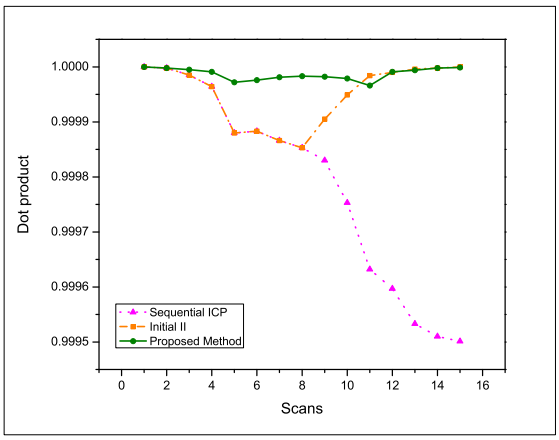


(c) Buddha

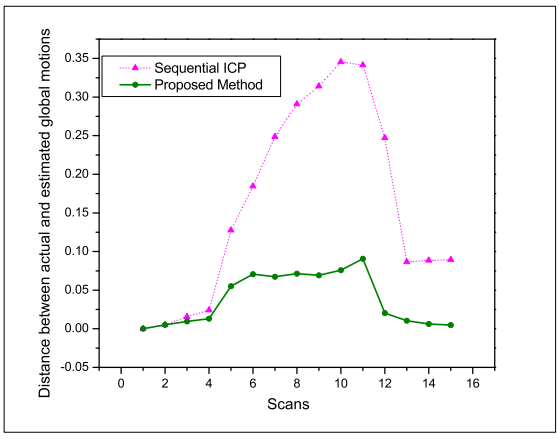
Figure 4.6: Cross sections of the registered Chef, Chicken and Buddha models. The global registration errors in Sequential ICP are eliminated successfully by outlier removed motion averaging by the proposed method. The highlighted areas clearly show the smoothed results obtained by the proposed method compared to Sequential ICP.

For the Stanford models, the actual global motions were available in the form of

unit quaternions and 3D translation vectors. The estimated transformation matrices from the multiview registration algorithm were converted to unit quaternions and translation vectors for comparison. The rotations were compared by taking the dot product of the actual and estimated unit quaternions. Ideally the dot product value should be close to 1.



(a) Dot product of actual and estimated global rotations



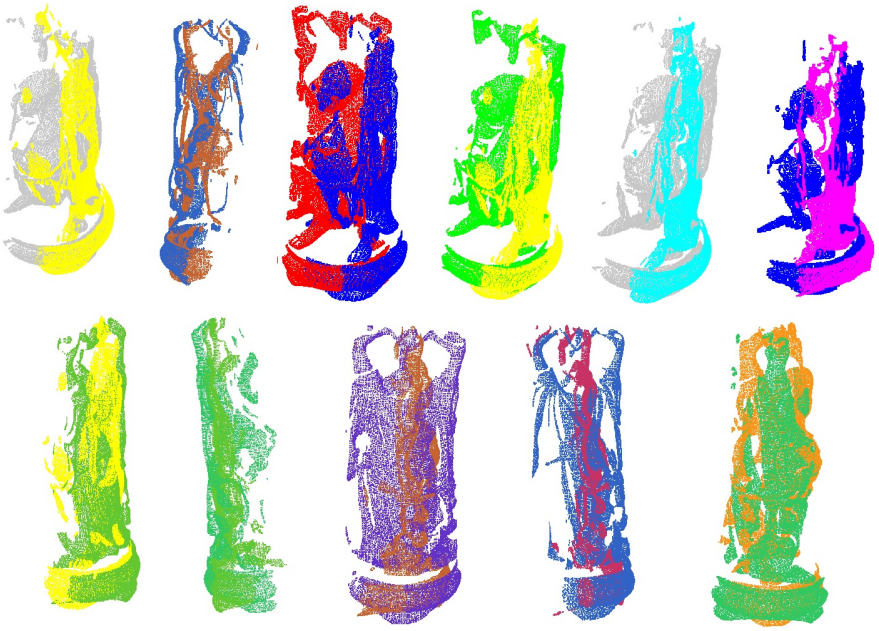
(b) Distance between estimated and actual global motions

Figure 4.7: Comparison of actual and estimated global motions for Buddha model. Fig.(a) shows the dot product between the ground truth motions and global motions estimated by Sequential ICP, initialisation method used, and the proposed method for different scan views. Fig.(b) shows the distance between actual and global motions estimated by Sequential ICP and the proposed method for the scan views.

The dot product comparison for different scan pairs of the Buddha model is provided in Fig.4.7(a). The dot product comparison is shown for the results obtained by Sequential ICP, Initial II and the proposed method. One can observe that the accuracy of rotation estimates obtained by Sequential ICP deteriorates towards the right due to the accumulation of registration error. From the result of the proposed initialization method (Eqn.4.1), one can see that by registering around $S_{N/2}$, the accumulation of errors can be reduced all the way to the last scan. The dip in the middle portion is due to the accumulated error from both directions. The proposed method is initialized using the result of Initial II.

From the Fig.4.7, one can see that the proposed method is able to achieve reasonably accurate global rotation estimates by averaging the relative motions after performing outlier rejection. The results by motion averaging are not included here as the merged cloud is largely corrupted by outliers. The distance between actual and estimated global motions were computed using Riemannian distance measure (Eq.4.10). Ideally the distance should be close to 0. Large distance values indicate that the estimated global motions vary greatly from the actual global motions. From Fig.4.7(b), one can see that the proposed method was able to achieve a good global motion estimate by selecting the valid inliers and averaging them.

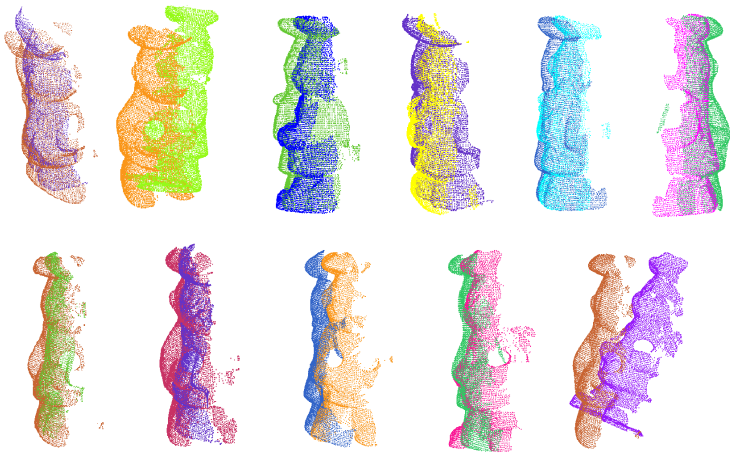
The outliers present in the set of relative motions identified by the proposed approach for the Stanford models are shown in Fig.4.8. Fig.4.9 shows the outliers in the relative motion set for the UWA models. The results show that the proposed method filters out the true outliers. The scan pairs shown in the figures are aligned according to the input relative motion estimates.



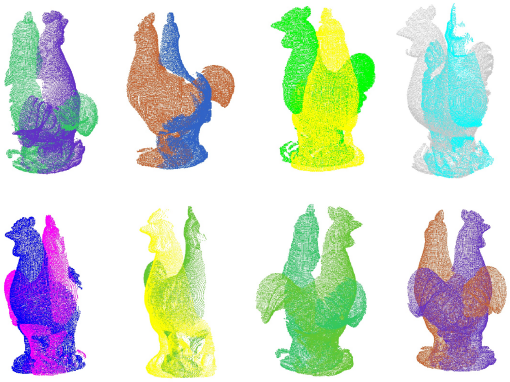
(a) Stanford Buddha model outliers

(b) Stanford Dragon model outliers

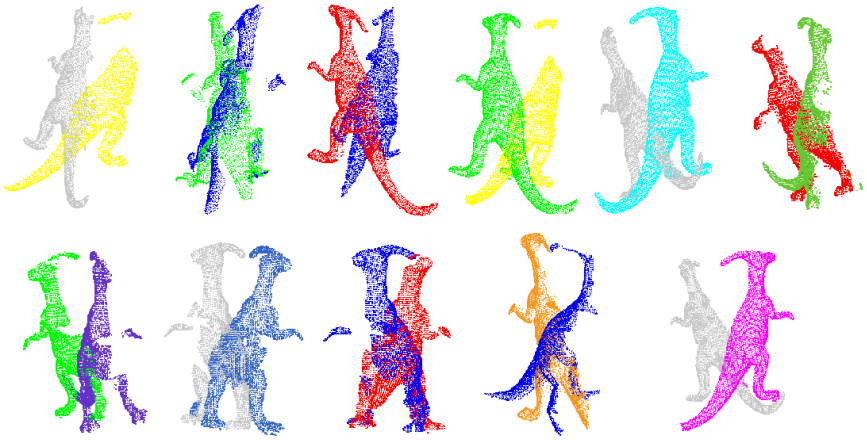
Figure 4.8: Outliers in the relative motions identified by the proposed algorithm for the Stanford models. The scans are aligned according to the pairwise registration results. The identified set of outlier relative motions are shown for Stanford Buddha model is shown in Fig.(a). Fig.(b) shows the identified outliers in the input set of relative motions for the Stanford Dragon model.



(a) UWA Chef Model



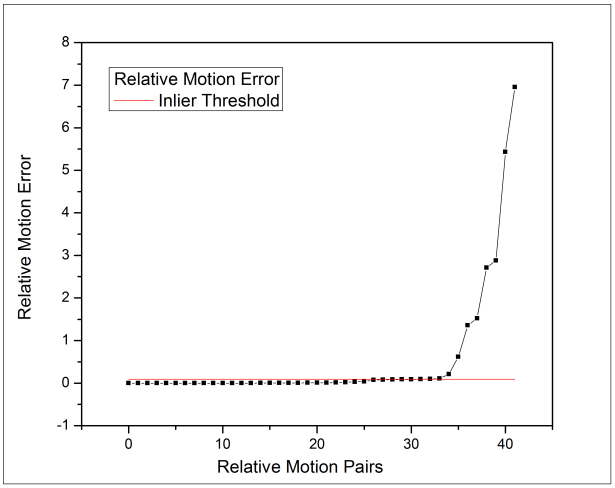
(b) UWA Chicken Model



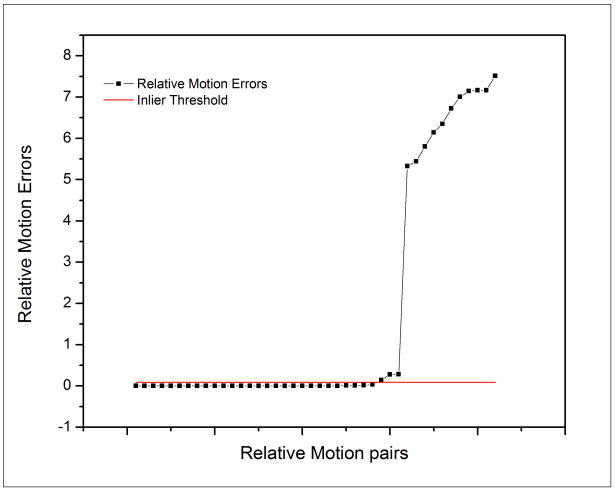
(c) UWA Parasaurolophus Model

Figure 4.9: Outliers in the relative motions identified by the proposed algorithm for the UWA models. The scans are aligned according to the pairwise registration results. The identified set of outlier relative motions are shown for (a) UWA Chef model, (b) UWA Chicken model and (c) UWA Parasaurolophus model.

Fig.4.10 and 4.11 show the inlier selection by the proposed method for the various models of Stanford and UWA datasets. The Riemannian distance between input and estimated relative motions (section 4.2.4) was compared against a threshold value (which was calculated empirically, guided by point cloud resolution). The relative motion was considered an inlier if this distance was below the threshold value.

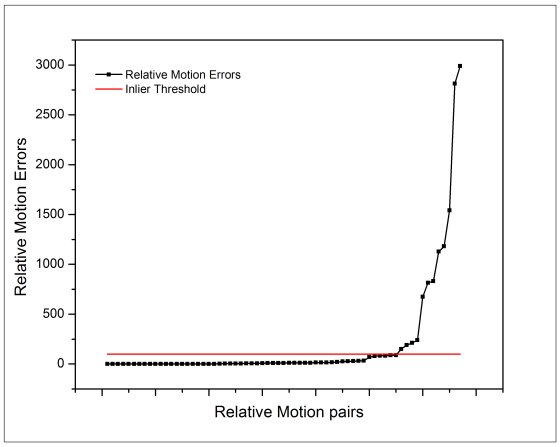


(a) Stanford Buddha model

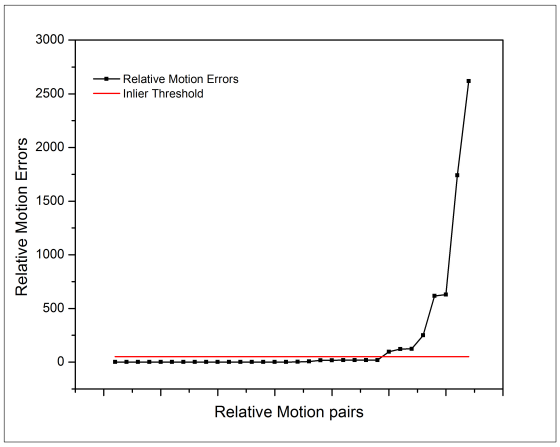


(b) Stanford Dragon model

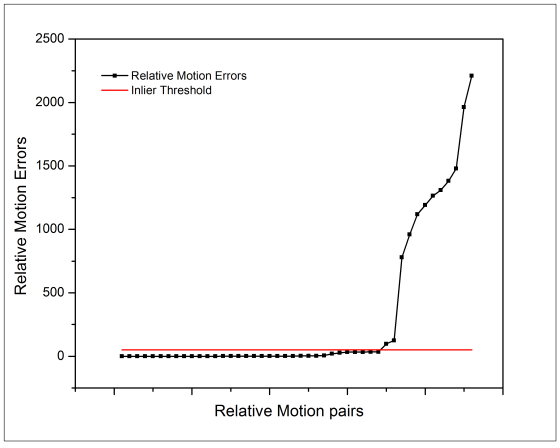
Figure 4.10: Inlier calculation by the proposed algorithm for the Stanford models. The input set of relative motions is indicated in the horizontal axis and the distances between the estimated and original set of relative motions are indicated in the vertical axis. The threshold selected is indicated by the red line.



(a) UWA Chef model



(b) UWA Chicken model



(c) UWA Parasaurolophus model

Figure 4.11: Inlier calculation by the proposed algorithm for the UWA models. The input set of relative motions is indicated in the horizontal axis and the distances between the estimated and original set of relative motions are indicated in the vertical axis. The threshold selected is indicated by the red line.

In Figs.4.10 and 4.11, the red line shows the threshold selected. One can see that the inliers and outliers are separated by a relatively large margin which makes the selection of threshold easier. The relative motions are sorted according to distance for clarity. Slight refinement errors in the relative motions can be handled by motion averaging and threshold may be selected in such a way as to avoid only large outliers.

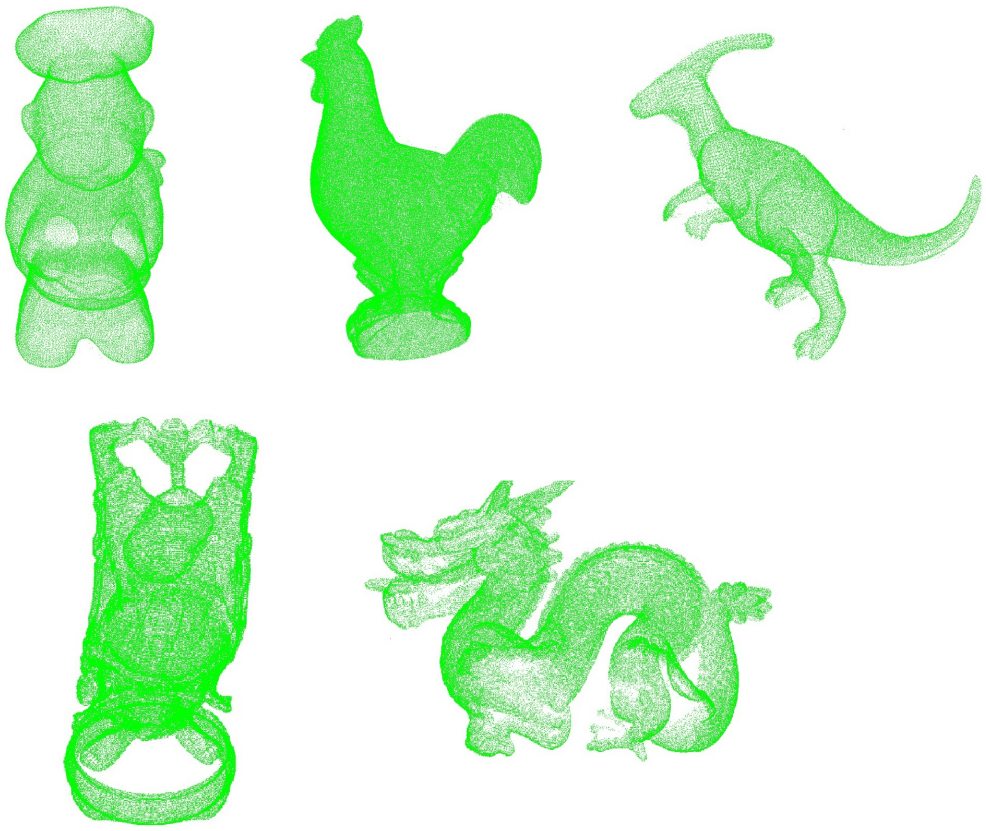


Figure 4.12: Merged point clouds by proposed method - Chef, Chicken, Parasaurolophus, Buddha, and Dragon models.

The final merged point clouds of the models achieved by the proposed algorithm are shown in Fig.4.12. The quality of registration shows that the proposed method was able to successfully achieve multiview registration in the presence of outlier relative motions.

4.5 Discussion

This work mainly considered the case of registration of 3D models from scans acquired in a turn-table kind of set up. The global motion initialization in section 4.2.1 based on the first and second method assumes that the adjacent scans have a good overlap for correct calculation of relative motions and it is true in most of the cases. If this is not the case, graph based initialization method can be adopted based on the assumption that a quality score is available from the relative motion calculation.

The vertex weight of graph G_1 in section 4.2.3 is a measure of agreement of a particular relative motion with the associated relative motions and initial global motion estimates. By considering the agreement with other relative motions, it is ensured that outlier relative motions get a large vertex weight compared to the correct relative motions and hence get excluded from the succeeding global motion estimation algorithm (SSSP). The calculation of global motions using the SSSP algorithm helps in considering the edges which may not be included in the initial global estimation. The calculation of inliers is straight forward and the threshold can be calculated empirically. As is evident from the graphs in Fig.4.7, the inliers and outliers are mostly separated by a large margin. The motion averaging step iteratively refines the relative and global motion estimates and hence slight refinement errors in relative motion estimates can be handled. Hence threshold may be selected in such a way as to avoid only large outliers. The recalculation of the registration path within the motion averaging iteration helps in updating the path according to the latest estimates.

The proposed algorithm for robust motion averaging performs better compared to the existing approaches (Govindu, 2006; Chatterjee et al., 2013) from a computational point of view. The runtime of RANSAC loop in Govindu (2006) is less predictable and data dependent compared to the single step calculation of inliers in the proposed algorithm. The L1 averaging step in Chatterjee et al. (2013) and the robust averaging are not required in the proposed approach as the inliers are identified before the motion averaging step.

4.6 Conclusion

A novel algorithm for introducing robustness in the multiview registration of 3D point clouds using motion averaging is proposed. The proposed algorithm is simple and efficiently filters outliers using a graph-based approach. The rigid body motion has a Lie Group structure. By making use of the Riemannian distance measure to form the edge weights in the view graph, the algorithm is able to identify the outliers among relative motions. The results are compared with the existing approaches and are shown to be robust and accurate.

4.7 Chapter Summary

This chapter presented a novel robust algorithm for multiview alignment of point clouds. The pairwise results used in this chapter was obtained using the method discussed in Chapter 3. The pairwise and multiview alignment methods are combined to form a framework for 3D registration in the next chapter.

CHAPTER 5

End-to-end registration framework for 3D modelling

Prelude: An automatic registration framework for 3D modelling from scan sequences is presented in this chapter. The framework is proposed for 3D object reconstruction as well as scene reconstruction. The pairwise and multiview alignment algorithms explored in the previous chapters are utilised in the framework. The various stages involved in the registration of point clouds are detailed.

A complete framework for 3D point cloud registration for 3D modelling is proposed in this chapter. Starting from the acquisition stage, each stage of the registration framework is detailed. Given a scan sequence for constructing the 3D model of an object, this framework will provide a complete registered object point cloud. A number of 3D models acquired by different acquisition techniques were registered and the results indicate the effectiveness of the proposed framework. The framework was also tested in the case of scene reconstruction. Scan sequences of real world scenes were acquired using Kinect sensor as well as a terrestrial laser scanner and the scenes were registered using the framework. The resulting registered scene reaffirms the efficiency of the framework for scene reconstruction case as well.

5.1 Method

The novel framework for point cloud registration from a sequence of scans for 3D modelling is presented in this section. The pictorial representation of the pipeline is given in Fig.5.1. Each of the stages is detailed below.

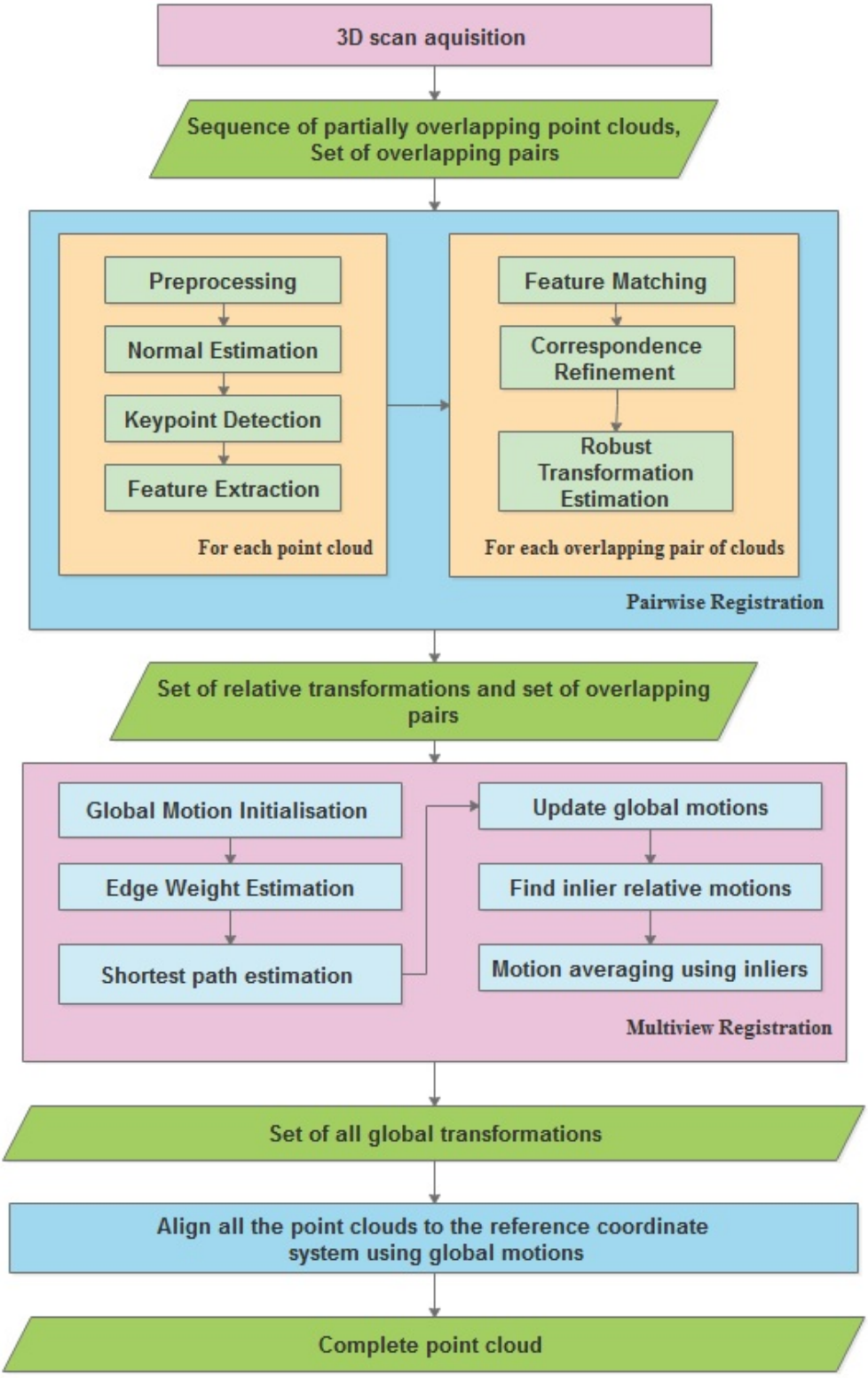


Figure 5.1: Framework for automatic 3D registration.

5.1.1 3D data acquisition

For the case of 3D modelling, multiple scans of an object can be acquired from different viewpoints by keeping the 3D camera stationary and by rotating the object in a turn-table set up to capture the entire object surface. Alternatively, the scans can also be acquired by placing the 3D camera spanned in locations around the object. Here rigid body transformation is assumed between the scans, which is a combination of 3D rotation and 3D translation, which aligns the two camera coordinate systems (or object coordinate systems). The adjacent scans are to be acquired ensuring that there is sufficient overlap between the scans. In this framework, the registration of a sequence of partially overlapping scans of the object surface in a cycle or a sequence of scene frames is considered.

5.1.2 Registration

In order to attain a single point cloud from the multiple partially overlapping point clouds captured in a sequence, a global registration method termed 'robust motion averaging' which make use of the redundancy in overlap is employed in this framework. It is assumed that not just the adjacent scans overlap. Additional pair of scans like alternate scans in the sequence, which may also overlap enough to produce a good pairwise estimate, are also considered. In the case of cycles, the first and last scans can be matched and registered using direct pairwise registration. The input to the algorithm is the set of scan pairs which are expected to overlap. The immediately adjacent scans are assumed to be having good overlap so that their pairwise registration results are correct to some extend. This is required since an initial global motion estimate is calculated based on the pairwise registration estimates of immediately adjacent scans. This forms the minimum requirement for finding a correct global registration. This rough estimate of global motions is refined with the use of additional constraints provided by the other pairwise estimates. If some of these pairwise registration results are wrong, then simple motion averaging will fail to find a correct global motion estimate. In order to deal with the wrong pairwise estimates (outliers), a novel robust motion averaging approach is proposed in Chapter 4. The pairs of scans from the input set are registered using pairwise registration algorithm. The results of the pairwise

registration stage form the input to the global registration stage. The robust motion averaging stage identifies the correct pairwise transformations and performs robust motion averaging to find the final global motion estimates.

5.1.3 Pairwise registration

The pairwise registration stage is employed to register all the scan pairs present in the input set of scan pairs. All scan pairs in the input set are registered separately to find the pairwise transformation matrices. The transformation which aligns the source scan with the target scan is identified for each pair. The following stages are involved in the pairwise registration of a source and target scan. This work employs the pairwise registration method detailed in Chapter 3.

Preprocessing

The acquired scans are stored in the formed of 3D point clouds. The preprocessing techniques are aimed at removing noise in the point clouds as well as at down-sampling high density point clouds. The preprocessing techniques can be adopted based on nature of the clouds to be registered. The down-sampling stage may not be required for point clouds with low density or less number of points. Similarly, a noise reduction stage like statistical filtering may be avoided in the case of noise-less good quality point clouds.

To filter out the noise, statistical filtering technique is used. This filtering helps to remove sparse isolated points from the point clouds. Voxel grid filtering technique is employed to reduce the density of points without affecting the discriminative power of the cloud. Reducing the number of points helps in slashing the high computational and memory requirements in subsequent processing stages. These techniques are detailed in Chapter 2.

Normal estimation

This stage is optional and is performed if any of the further stages make use of the normal or curvature information. The normals are estimated at all points in the

source and target point clouds. The normal estimation may be performed before down-sampling to obtain smother distribution of normals.

To estimate the normals at each point, a local neighbourhood around the point is considered. The neighbourhood at a point is considered to be the points falling within a sphere of fixed radius around the point. Alternatively, a fixed k nearest neighbours of the point can be considered. A plane is then fitted to the neighbourhood points using PCA (Jolliffe, 2014). The eigen vector corresponding to the smallest eigen value, of the covariance matrix computed from the points, is approximated as the surface normal at the point. Since the direction of normal is ambiguous, all normals are oriented consistently towards the view point (Rusu, 2010).

Keypoint detection

This work employs the coarse registration method based on feature matching. 3D features are often extracted only at salient points called keypoints or feature points instead of extracting them at all the points in a cloud. This is primarily used to obtain good localisation of features and to avoid multiple matches in a neighbouring region. This also aids in reducing the computational complexity of the subsequent stages. Keypoints are extracted from source and target point clouds.

Based on the nature of the point clouds, a keypoint detection stage which detects salient points in the cloud is employed. A number of different keypoint detection methods available in literature are discussed in section 2.1.1. Based on the previous evaluations available in literature (Tombari et al., 2013), this work employed the ISS keypoint detection method discussed by Zhong (2009) for extracting the keypoints. The ISS keypoint detection considers the geometry of the points in the local neighbourhood for finding the salient points. The third eigen value of the scatter matrix of the neighbourhood points is considered as the saliency measure. A pruning stage which discards the points which does not satisfy constraints on the ratio of eigen values is employed. The idea is to avoid keypoints at locations where the spread of keypoints is not similar. i.e. non-corner points. Non maximal suppression is done on a fixed neighbourhood to improve the localisation of the keypoints. For coloured point clouds with less geometric information, keypoint detection methods which make use of the intensity information or a combination of both, can be employed.

Feature extraction

The 3D features describe the local neighbourhood at a salient point. Feature descriptors are extracted at the keypoints of both the source and target point clouds. Based on the nature of the point clouds, efficient 3D features can be extracted for matching. For scans with rich geometry information, geometric features can be employed. For 3D scans with additional intensity information, a combination of both intensity and geometric information can be employed for feature extraction. For scans with less distinctive geometric features, features which make use of just the intensity information may be employed. If the normal estimation method applied is accurate, features which rely largely on the normal information, like SHOT (Tombari et al., 2010b) can be employed. The 3D Shape Context feature (Frome et al., 2004), which rely on the distribution of points in the neighbourhood for feature extraction, is employed in this work in the case of point clouds with no intensity information. The normal estimation method employed is sub-optimal and hence the 3D Shape context feature which relies less on the normal information is employed to reduce its effect on the resultant feature. For point clouds with intensity information, the colour SHOT feature (Tombari et al., 2011), which makes use of both geometry and colour information, is used as the feature descriptor.

3D Shape Context is a shape histogram based descriptor which bins the number of points in the spatial local neighbourhood. The local spherical neighbourhood, with north pole aligned with surface normal is subdivided along radial, azimuthal and elevation dimensions. The descriptor is calculated as described in section 3.3.4.

SHOT Colour (CSHOT) depends on both the geometric and colour information of the spherical neighbourhood at a point to form a descriptor. A repeatable local reference frame (LRF) is defined at each keypoint based on the Eigen decomposition of the scatter matrix of the neighbourhood. An isotropic spherical grid is defined based on the LRF to define a signature structure to encode spatially localised information. For each sector of the grid, a histogram of normals is defined and the descriptor is formed by concatenating the histograms. The colour information is similarly encoded by using the colour representation in CIELab space. Both geometric and colour histograms are concatenated to form the final descriptor. The structure of the colour SHOT descriptor is illustrated in Fig.5.2 (Tombari et al., 2011).

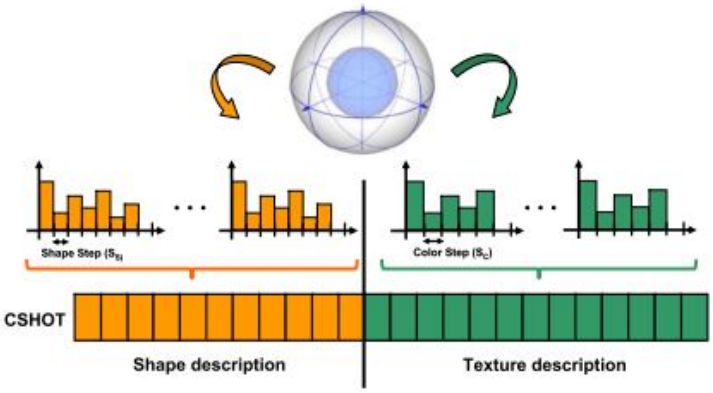


Figure 5.2: CSHOT descriptor structure.

Correspondence Estimation and Refinement

To estimate the rigid body transformation between the source and target point clouds, the feature vectors extracted from both the scans are matched. For each feature vector in the source point cloud, the nearest feature vector in feature space from the target point cloud forms the match and the respective keypoints form the corresponding pair. The closest match in feature space is computed with the help of approximate nearest neighbour estimation (FLANN) technique and the space partitioning representation of clouds called kd-trees.

The set of correspondence pairs thus formed may consist of a lot of false matches due to partial overlap of clouds, surface symmetry, noise etc. A number of refinement techniques to filter out the outliers can be used as discussed in section 2.1.1. In this work, the correspondence refinement techniques based on distance and a polygon based approach (Buch et al., 2013) are used. A correspondence refinement technique based on the similarity of curvatures of the match points is also employed.

Initial Transformation Estimation

Given a minimum set of three corresponding point pairs, rigid body transformation can be estimated by closed form least square solutions (Arun et al., 1987; Umeyama, 1991). However, the presence of a high percentage of wrong correspondences may lead to incorrect estimates by the closed form methods. To deal with the presence of outliers in the set of correspondences, robust estimation techniques are employed. In

this work, the robust estimation technique detailed in Chapter 3 is used. The method outperforms the popular RANSAC in terms of accuracy and computational time. An initial estimate of the transformation is obtained by the proposed coarse registration method. This is then fine tuned by fine registration.

Pairwise Fine Registration

Once a rough initial estimate of transformation between the source and target point clouds are obtained, a refined estimate can be computed by fine tuning. One of the benchmark algorithms for fine registration is the Iterative Closest Point (ICP) algorithm proposed by Besl et al. (1992). This is a local optimisation based method and is described in section 2.1.2. In this work, a variation of ICP is used for fine registration. The correspondences are formed from the nearest neighbours calculated using point-to-plane or point-to-point distance measures. The correspondences are refined by a polygon based method (Buch et al., 2013). This refinement technique helps in establishing rigid body motion constraints. The transformation is calculated using a closed form solution by least squares technique from the refined correspondences.

5.1.4 Multiview Registration

The pairwise registration of each pair in the input set of scan pairs is performed. This set of relative transformations is provided as input to the multiview registration algorithm. The multiview registration is performed by a robust motion averaging technique detailed in Chapter 4. The transformations from each of the scans to a reference scan (usually the first scan) are called global transformations. From the input set of relative transformations, the incorrect ones or outliers are identified and removed by the algorithm. An initial estimate of global transformations is obtained by sequential registration. Using a graph based approach, the set of relative motions which agree with the global motions as well with other relative motions are identified. The Lie group structure of rigid body transformations helps in finding the distance between input relative motions and their estimates computed from global transformations. Once the inlier relative motions are identified, motion averaging is performed on them to give the final global transformations.

5.1.5 Post-processing

Once the global transformations from each scan to the reference scan are obtained, the 3D scans can be converted to the reference coordinate system by applying these transformations using Eq.5.1, where x', y', z' and x, y, z indicate the transformed and original coordinates of the scan points, R_g and T_g are the global rotation and translation.

$$\begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} R_g & T_g \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{5.1}$$

Once all the scans are thus converted to the common coordinate system, post processing steps like scan integration, hole filling etc. are performed for 3D modelling. In this work, voxel grid filtering is performed on the merged point cloud as part of scan integration. This helps in providing uniform density to the resultant point cloud and smooths out the duplicate points appearing in overlap regions.

5.2 Datasets Used

The proposed registration framework was evaluated for registration of different datasets. The 3D scans captured by scanners which work on different acquisition techniques like laser triangulation, structured light, time of flight etc. were used for testing. Two application scenarios considered include object reconstruction and scene reconstruction. In the case of object reconstruction, a sequence of scans of an object was acquired in overlapping fashion from multiple view points. These scans were then registered by the proposed framework to form a single point cloud of the object. The scans were acquired by placing the object in a rotating turn-table set up or by moving the camera around the object. In the case of scene reconstruction, a sequence of scans of a scene was considered. Scenes with and without cycles were registered. The datasets used are given in Table 5.1 and explained below.

Table 5.1: The datasets used in this study.

S.No	Model	Dataset	Scanner
1	Buddha	Stanford	Cyberware 3030 MS
2	Bunny		
3	Dragon		
4	Armadillo		
5	Chef	UWA	Minolta Scanner
6	Chicken		
7	Parasaurolophus		
8	TRex		
9	Duck	Kinect DB	Microsoft Kinect
10	Super Mario		
11	Pup	acquired	Microsoft Kinect
12	Room Scene1	acquired	Microsoft Kinect
13	Room Scene1	acquired	FARO FOCUS^{3D} X 330

Stanford Data set: Scans of four different models from the Stanford dataset (Levoy et al., 2005), acquired using Cyberware 3030 MS scanner - Buddha, Dragon, Bunny and Armadillo - were used for evaluation.

University of Western Australia dataset: The scans of Chef, Chicken, Parasaurolophus and TRex models, acquired by Minolta scanner from the University of Western Australia (UWA) dataset (Mian and Bennamoun, 2006; Mian et al., 2006) were used for evaluation.

Kinect dataset: The scans acquired by the Microsoft Kinect for Windows sensor were used for evaluation. The Kinect for windows sensor works based on the principle of structured light. The cases of both object and scene reconstruction were considered. The scans were acquired in our office space to obtain multiple views for registration. In addition to this, some of the models from a Kinect database (Petrelli and Di Stefano, 2016) were also used for registration. For scene reconstruction, a sequence of scans was acquired by Kinect sensor for registration.

Terrestrial Laser Scanner Dataset: A terrestrial laser scanner by FARO (FARO

FOCUS^{3D} X 330) was used to acquire scans of a room scene in our office space for registration. Unlike the previous datasets which capture data from a single view point, FARO scanner acquires data from multiple view points in a single scan. This is made possible by the specialised design of the scanner which can capture data by rotating a mirror in 360° while capturing. Multiple scans of a room were acquired by placing the scanner at multiple viewpoints across the room, where each scan consisted of data visible from around the scanner in 360°. Acquisition from multiple view points were required due to the occlusion caused by the objects in the scene. Multiple scenes were registered to form a complete point cloud of the room.

5.3 Results

Stanford and UWA Dataset Models: The Buddha, Dragon, Bunny and Armadillo were the models considered from the Stanford dataset. The Chef, Chicken, Parasaurolophus and TRex models of the UWA dataset were also tested using the proposed registration framework. Since colour information was not available for these models, 3D shape context feature was considered for registration. ISS keypoint detection method was used for identifying the keypoints. Fine registration was performed by a modified ICP registration, where the correspondences were refined by ensuring rigid body constraints. The results obtained for each of the models are detailed here.

Figure 5.3: Partially overlapping input scans of the Buddha model.

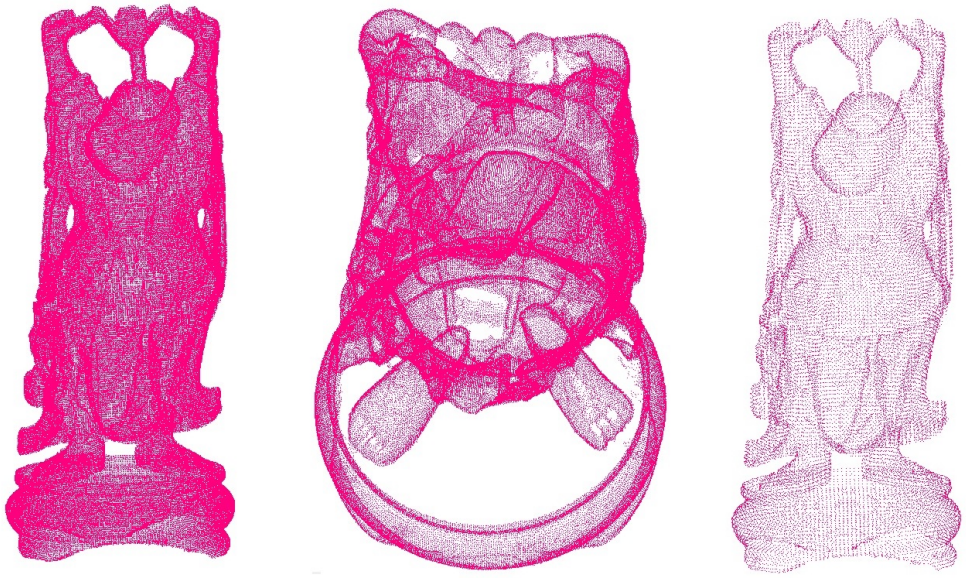


Figure 5.4: Merged Buddha model: two different views and a down-sampled view.

Buddha model: The snapshots of the input scans of Buddha dataset, captured from multiple view points around the object are shown in Fig.5.3. The scans were down-sampled and registered using the proposed framework. Snapshots of the registered model taken from different viewpoints and a down sampled registered version are shown in Fig.5.4. The cross sections of the registered model obtained after the initialisation using sequential method and after final motion averaging are shown in Fig.5.5. From the cross-section, it can be observed that a smooth cloud is obtained by the proposed method.

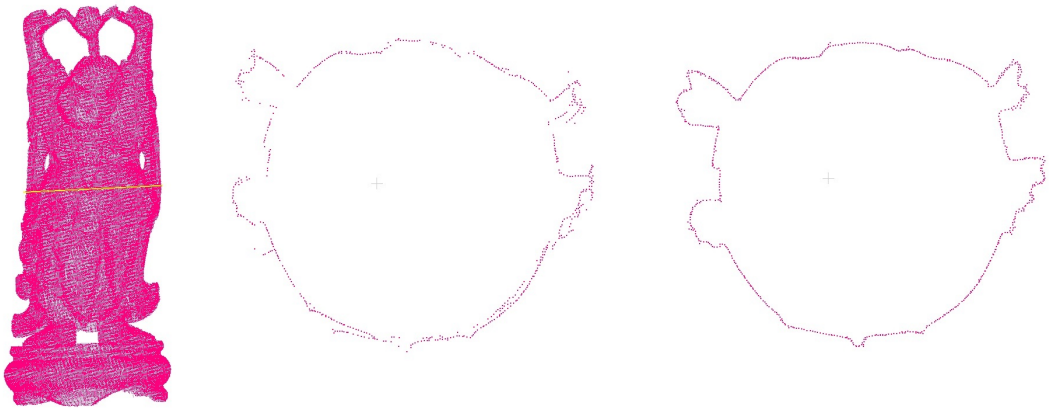


Figure 5.5: Cross section of registered Buddha model: a) after sequential initialisation, and b) after final processing.

Dragon model: The scans of the Stanford Dragon model (Fig.5.6) were registered by the proposed framework and the obtained registered cloud is shown in Fig.5.7. Two views of the registered cloud, along with a down-sampled version of the same are shown. For this model, since the relative estimates were very good, the global registration by simple sequential method was also successful, as shown by the cross sections of the registered model in Fig.5.8.

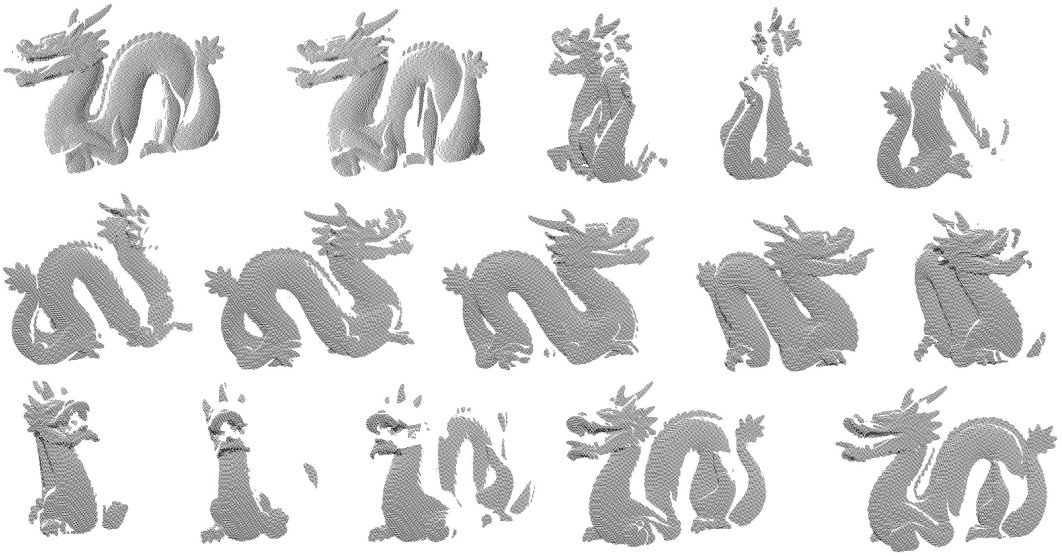


Figure 5.6: Partially overlapping input scans of the Dragon model.

Figure 5.7: Merged Dragon model: two different views and a down-sampled view.

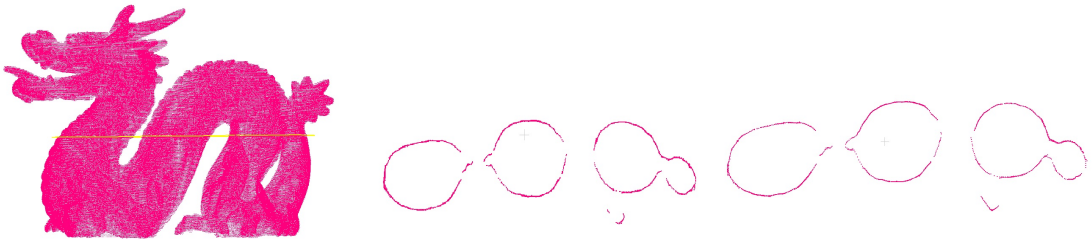


Figure 5.8: Cross section of registered Dragon model: a) after sequential initialisation, and b) after final processing.

Bunny model: Stanford Bunny model scans were registered using the proposed framework. The input scans considered are shown in Fig.5.9. Different views of the registered model and a down-sampled version are shown in Fig.5.10. The cross section of the registered model by sequential initialisation shows the loop closing error, which is rectified by the robust motion averaging, as shown in Fig.5.11. The Bunny dataset was challenging because there were only a few scans in the sequential cycle. For this dataset, additional scans which are not from a strict turn-table sequence, were also considered, as observed in the Fig.5.9. The overlap was identified manually for these additional scan views, unlike the other datasets where the order of the scan defined the sequence. Hence, one can see that once an approximate overlap information is available, these additional scans can also be registered using the proposed method.

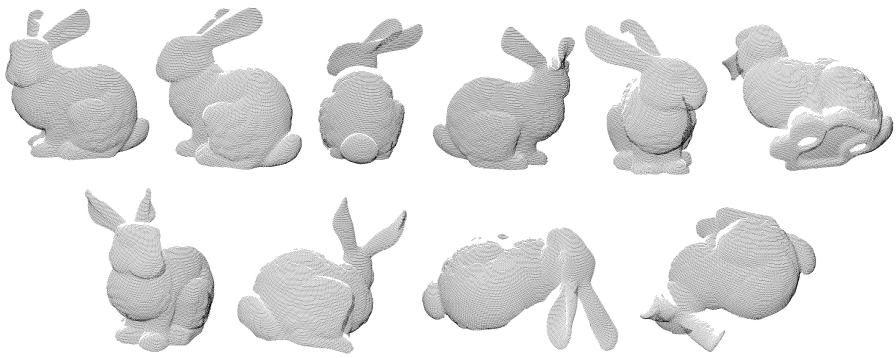


Figure 5.9: A set of input scan views of the Bunny Model.

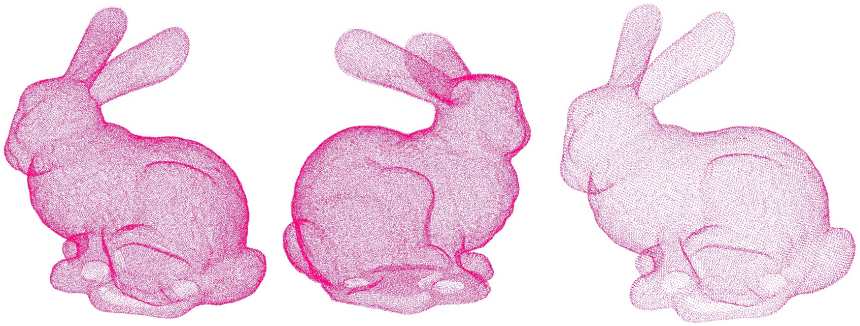


Figure 5.10: Merged Bunny model: two different views and a down-sampled view.

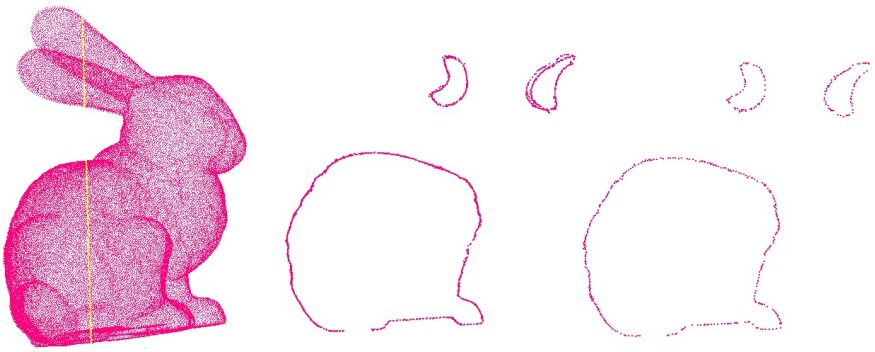


Figure 5.11: Cross section of registered Bunny model: a) after sequential initialisation, and b) after final processing.

Armadillo model: Armadillo model scans from the Stanford dataset were registered using the proposed framework. The input scans considered are shown in Fig.5.12. The registered cloud views and the cross section views are shown in Figs.5.13 - 5.14.

Chef model: The multiview scans of the UWA Chef model were registered by the proposed framework. Fig.5.15 shows some sample input clouds from the dataset. Two different views of the registered cloud and a down-sampled cloud are shown in Fig.5.16. The cross-sections of the registered cloud after initial registration by sequential method and after final alignment are shown in Fig.5.17.

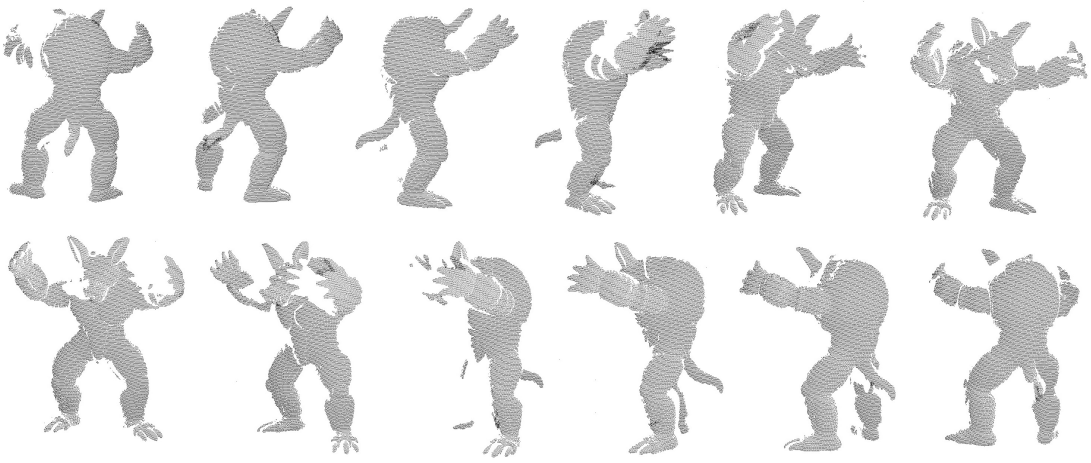


Figure 5.12: A set of input scan views of the Armadillo model.

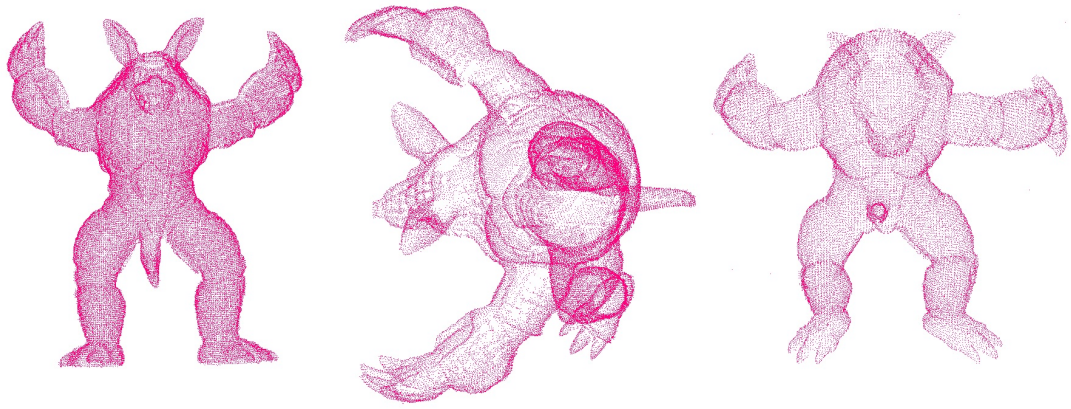


Figure 5.13: Merged Armadillo model: two different views and a down-sampled view.

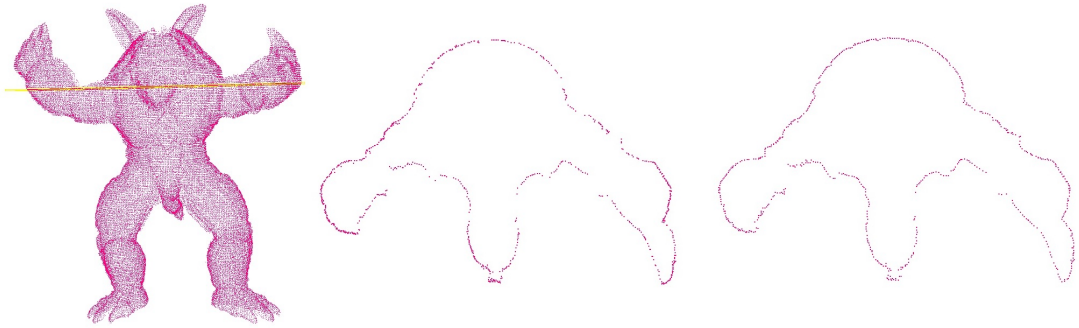


Figure 5.14: Cross section of registered Armadillo model: a) after sequential initialisation, and b)after final processing.

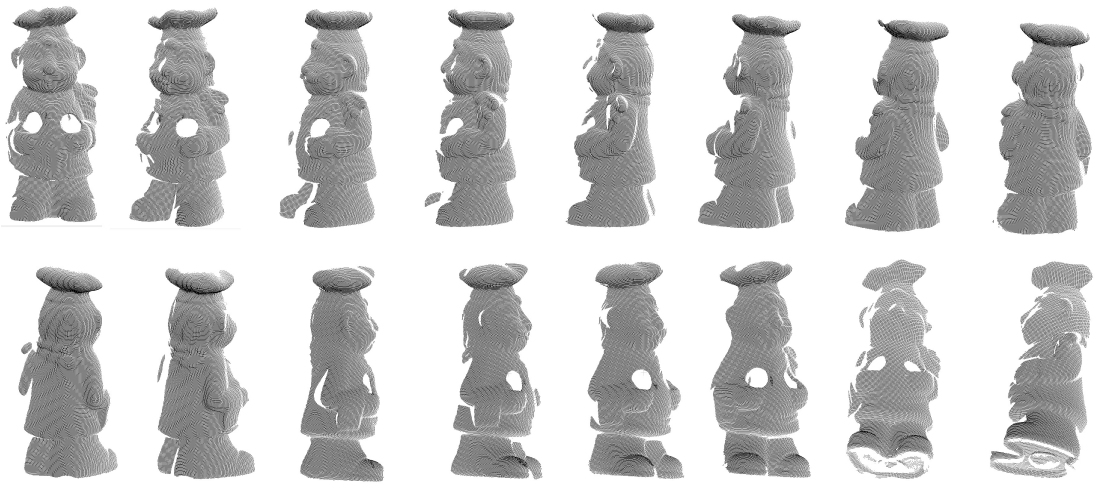


Figure 5.15: A set of input scan views of the Chef model.

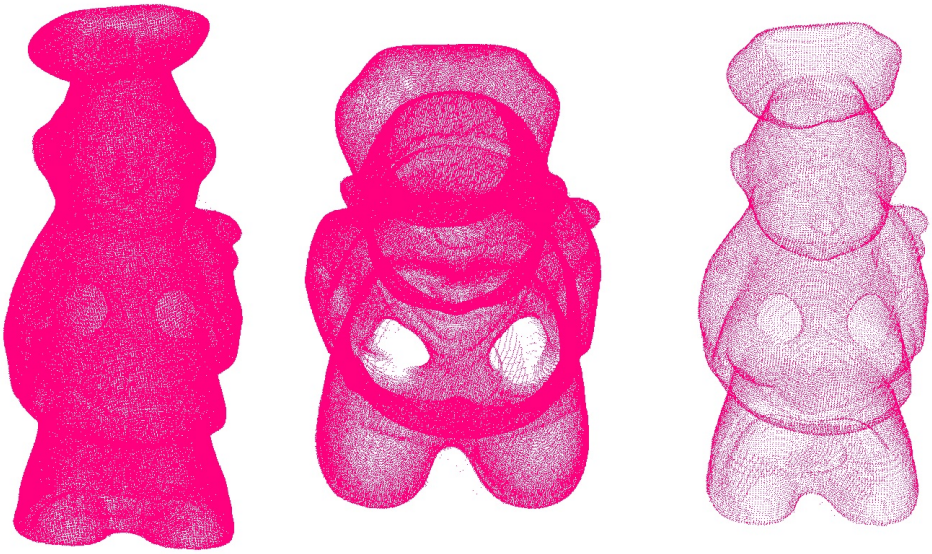


Figure 5.16: Merged Chef model: two different views and a down-sampled view.

Chicken model: Multiple scan views of the UWA Chicken model (sample input scans in Fig.5.18) were registered using the proposed framework to produce the registered model shown in Fig.5.19. The cross sections of the registered model after sequential registration and proposed motion averaging are shown in Fig.5.20.

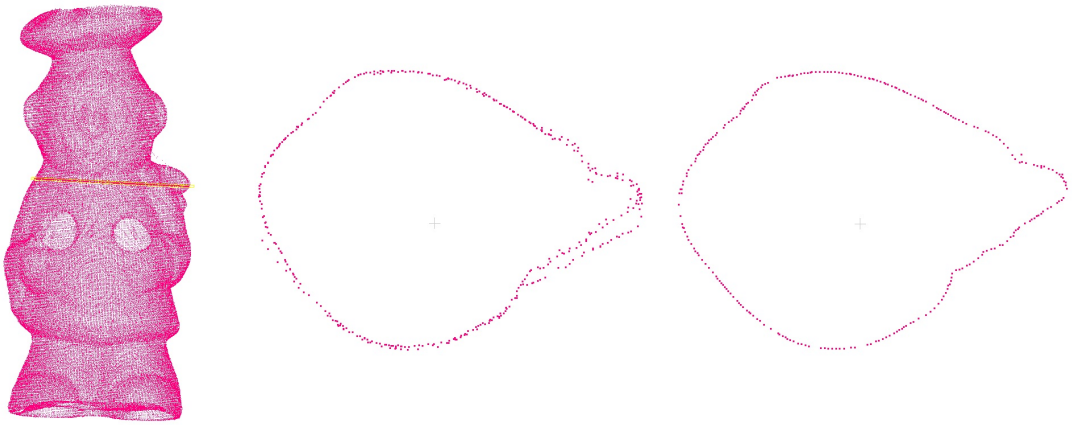


Figure 5.17: Cross section of registered Chef model: a) after sequential initialisation, and b) after final processing.

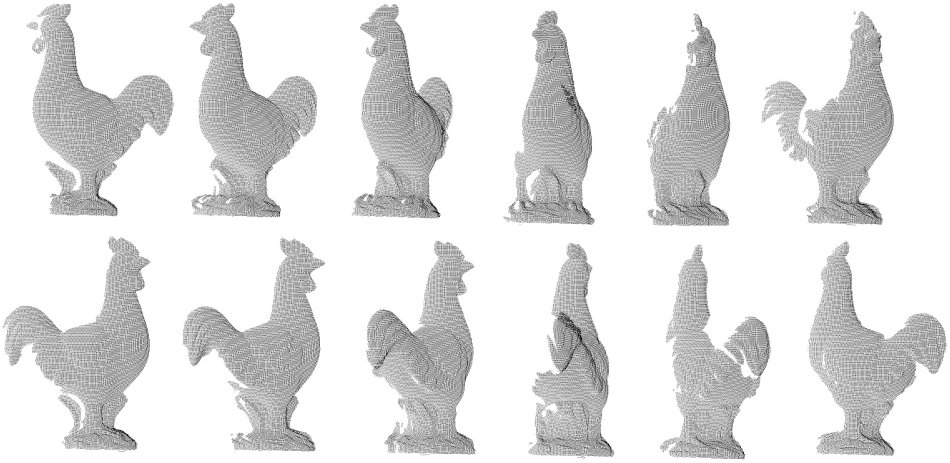


Figure 5.18: A set of input scan views of the Chicken model.

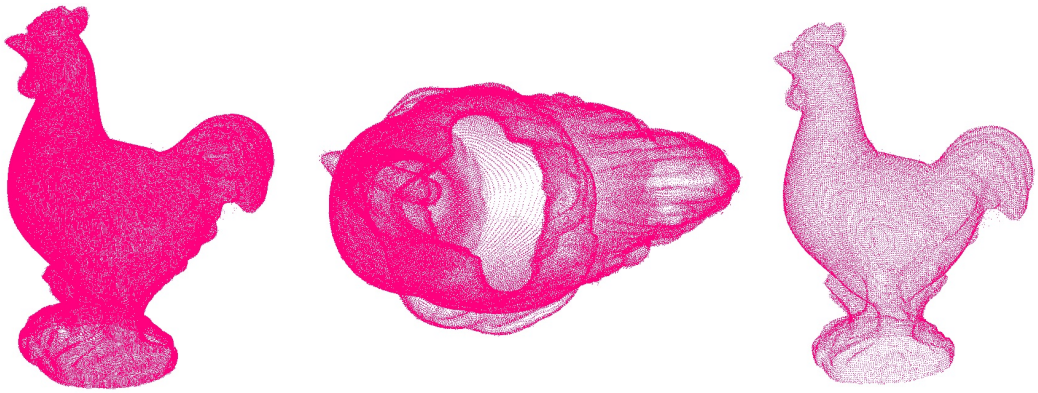


Figure 5.19: Merged Chicken model: two different views and a down-sampled view.

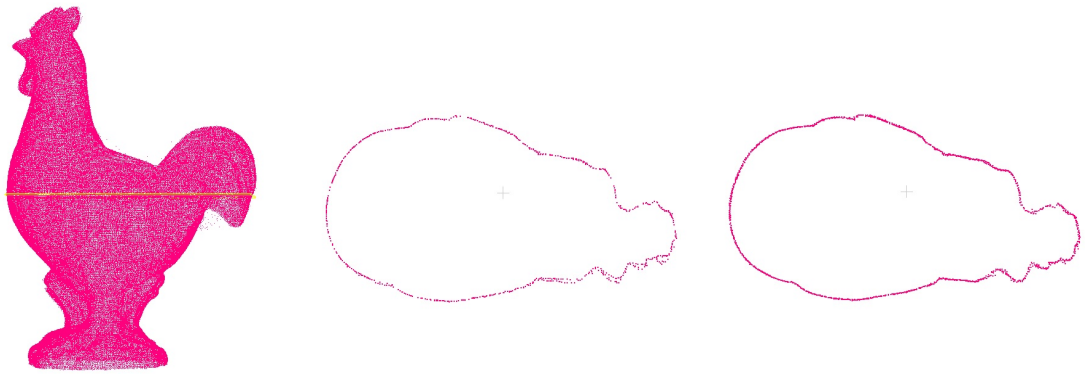


Figure 5.20: Cross section of registered Chicken model: a) after sequential initialisation, and b) after final processing.

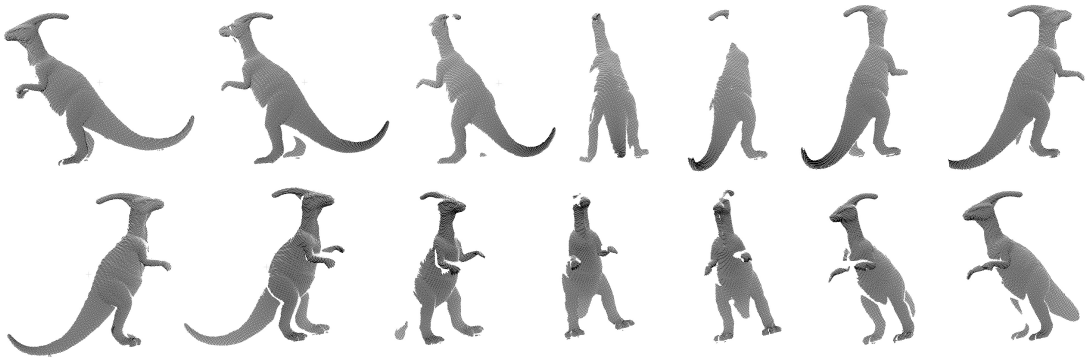


Figure 5.21: A set of input scan views of the Parasaurolophus model.

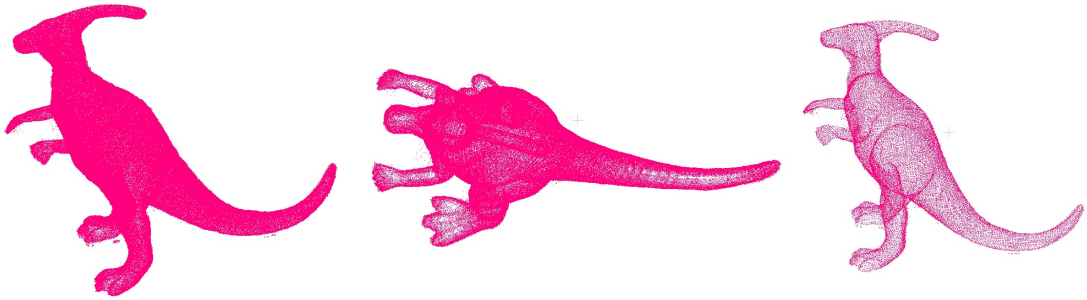


Figure 5.22: Merged Parasaurolophus model: two different views and a downsampled view.

Parasaurolophus model: The results of aligning the multiple scans (Fig.5.21) of the Parasaurolophus model from the UWA dataset using the proposed registration pipeline is shown in Fig.5.22. By comparing the cross-sections (Fig.5.23) of the registered scans of the global initialised cloud and the final cloud, it can be clearly observed that the loop closing error was minimised and distributed across the scans.

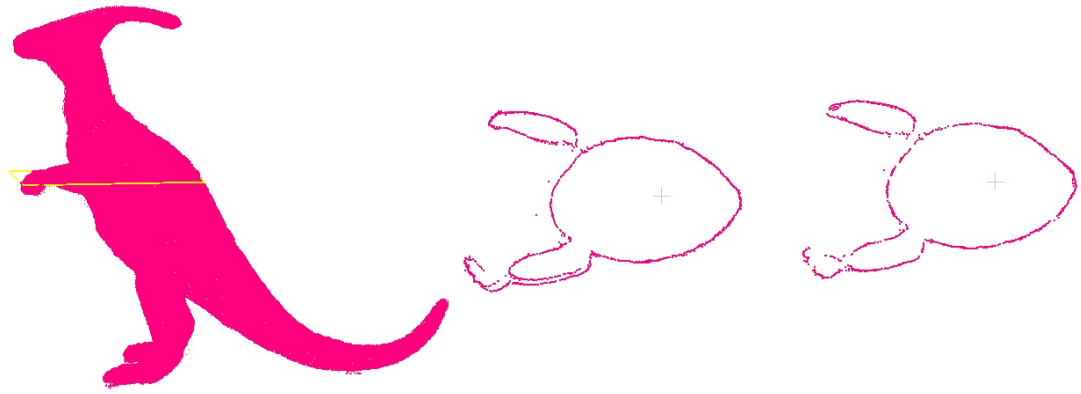


Figure 5.23: Cross section of registered Parasaurolophus model: a) after sequential initialisation, and b) after final processing.

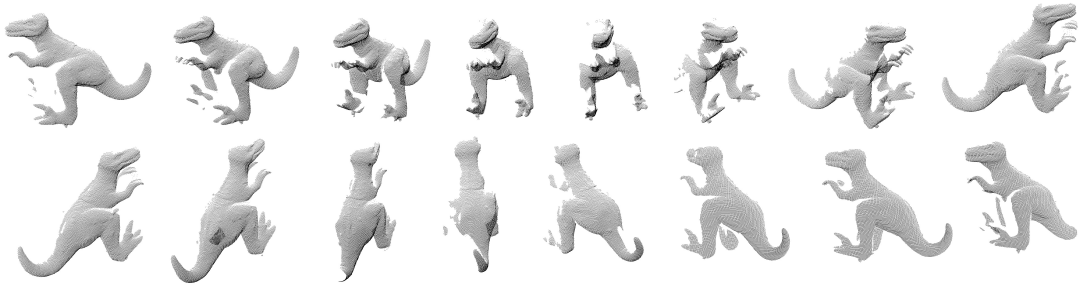


Figure 5.24: A set of input scan views of the TRex model.

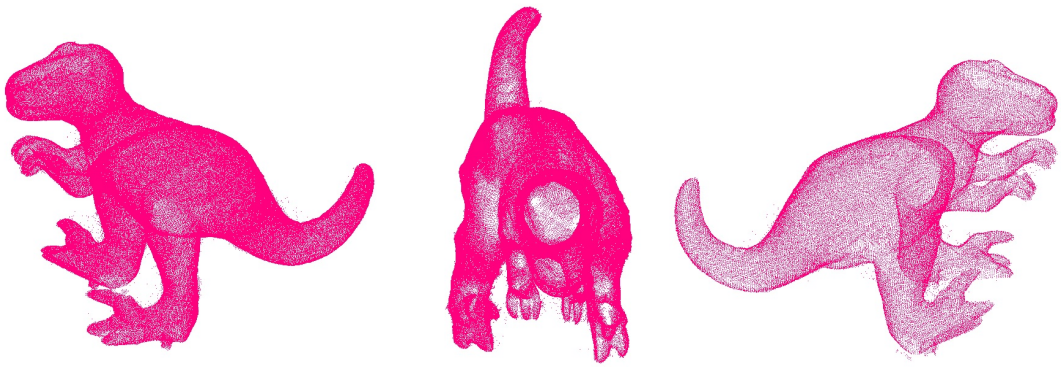


Figure 5.25: Merged TRex model: two different views and a down-sampled view.

TRex model: TRex was another model considered from the UWA dataset for registration. The input scans, the final cloud obtained and the cross sections of the registered cloud before and after robust averaging are shown in Figs.5.24 - 5.26.

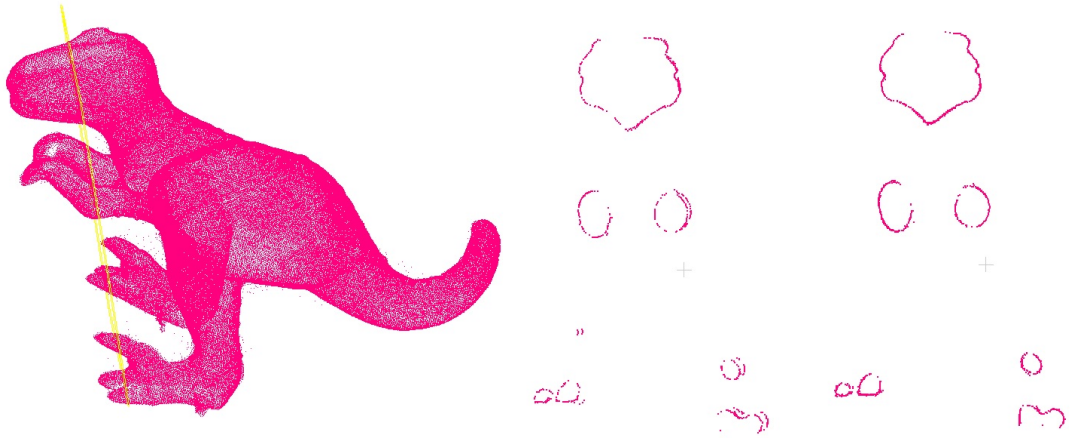


Figure 5.26: Cross section of registered TRex model: a) after sequential initialisation, and b) after final processing.

The results of the UWA models also reaffirm the effectiveness of the proposed pipeline for 3D scan registration.

Duck model: The multiple scans of the Duck model acquired by Kinect sensor were registered using the proposed framework. Since this model had colour information for the points, the coloured point cloud processing was used. SHOT Colour features were used for feature extraction. As Kinect sensor data is prone to noise, the registration task is more challenging compared to the laser data registration. The input scans considered are shown in Fig.5.27. The scans were also aligned using the ground-truth global transformations available in the database. The final point clouds obtained using the ground-truth transformation and the proposed framework are shown in Fig.5.28(a) and (b) respectively. Since the merged cloud contain some noisy points, a filtering was performed on the final cloud to obtained a cleaned point cloud of the model, which is shown in Fig.5.28(c). The cross-sections of the ground-truth model, merged model and cleaned model are shown in Figs.5.29(a) - 5.29(c). From the figures, it is observed that the aligned point cloud obtained with the proposed method is even better than the one obtained using ground-truth transformations.

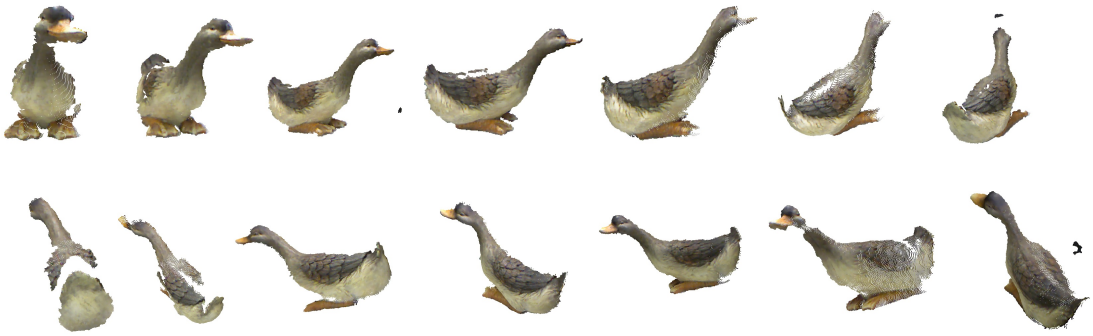


Figure 5.27: A set of input scan views of the Duck model.

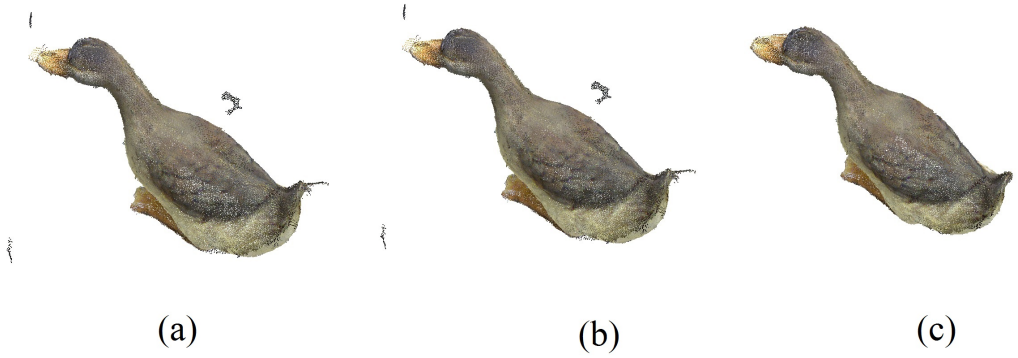


Figure 5.28: Final merged Duck model obtained from (a) ground-truth, (b) proposed method, and (c) after noise filtering.

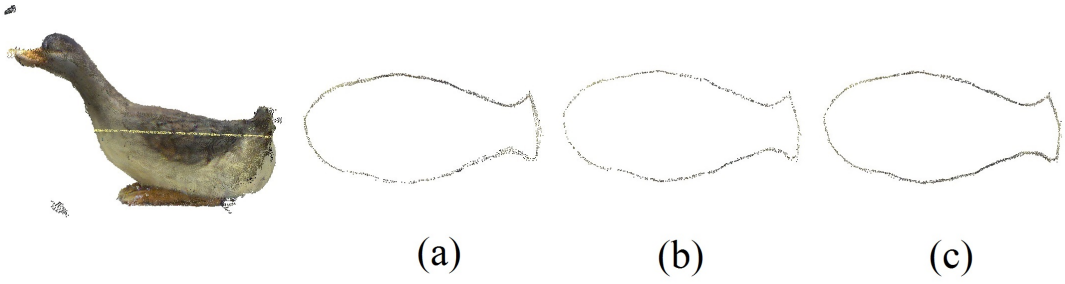


Figure 5.29: Cross section of registered Duck model: (a) ground-truth, (b) proposed method, and (c) after noise filtering.

Super Mario model: Another model considered was the Super Mario model from the Kinect database and the scan views were captured using Kinect sensor. The input scans are corrupted with noise typical to the Kinect scan data and hence the registration is challenging. The set of input scans considered are shown in Fig.5.30. The

registered cloud from the proposed framework was compared with the registered point cloud obtained using ground-truth transformations in Figs.5.31(a) - 5.31(b). A cleaned registered model obtained by statistical filtering is shown in Fig.5.31(c). Multiple views of the model are shown for clarity.

Figure 5.30: A set of input scan views of the Super Mario model.

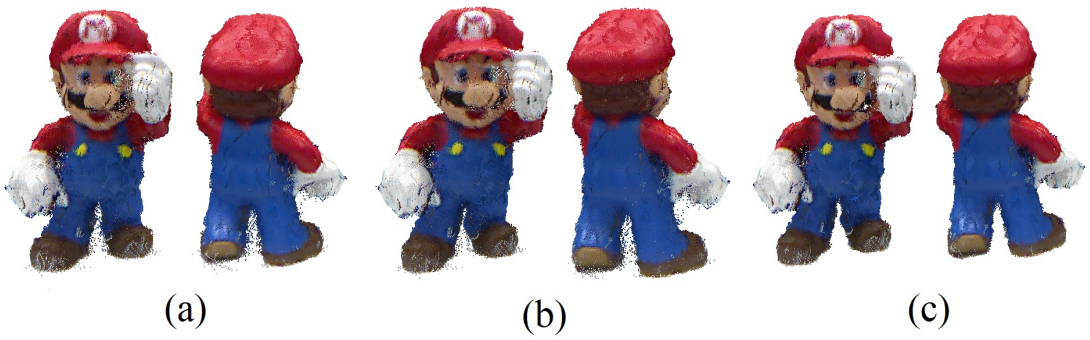


Figure 5.31: Final merged Super Mario model obtained from (a) ground-truth, (b) proposed method, and (c) after noise filtering.

Pup model: The 3D scan data of a Pup model was acquired in our office space using the Microsoft Kinect sensor. The views were registered using the proposed registration pipeline. Some of the input scans are depicted in Fig.5.32. The complete registered model after initialisation using sequential registration and the one after final processing are shown in Figs.5.33(a) and (b). The registered noisy cloud cleaned using statistical filtering technique is shown in Fig.5.33(c). From the figure, it is observed that the loop closing error in the sequential initialisation is rectified by robust motion

averaging.

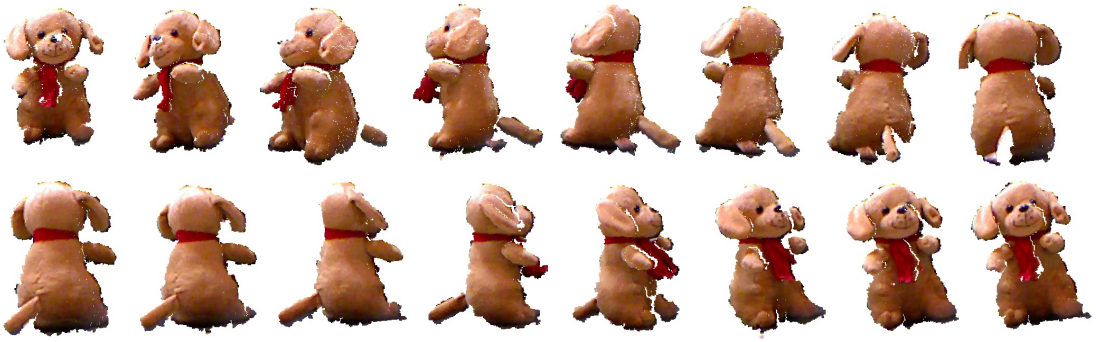


Figure 5.32: A set of input scan views of the Pup model.

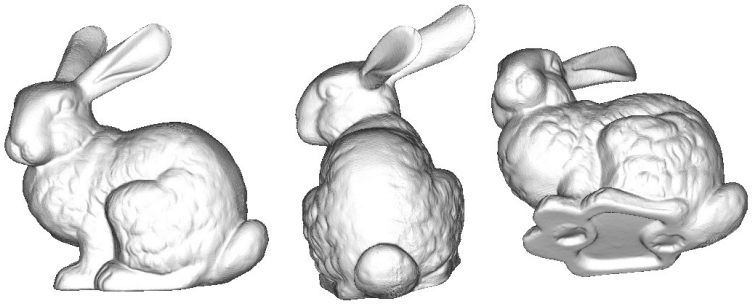
Figure 5.33: Final merged Pup model obtained by (a) sequential initialisation, (b) proposed method, and (c) after noise filtering.

Surface Reconstruction: The surface reconstruction of the registered models was performed using the open source software Cloud Compare (Girardeau-Montaut, 2015). Poisson surface reconstruction was performed on the registered clouds to obtain the complete reconstructed models. Figs.5.34 - 5.36 show the surface reconstructed models. The results signify the quality of registration as well-defined surfaces could be reconstructed from the registered clouds. The details on the reconstructed surface of models with scans captured from laser scanners show the quality of 3D scan data and the correctness of registration. One can observe that the clouds captured with Kinect sensor are of low quality and this has resulted in the registered clouds which are not smooth and hence has produced models with limited surface details. The clouds having less point density or holes in the data also can lead to incorrect

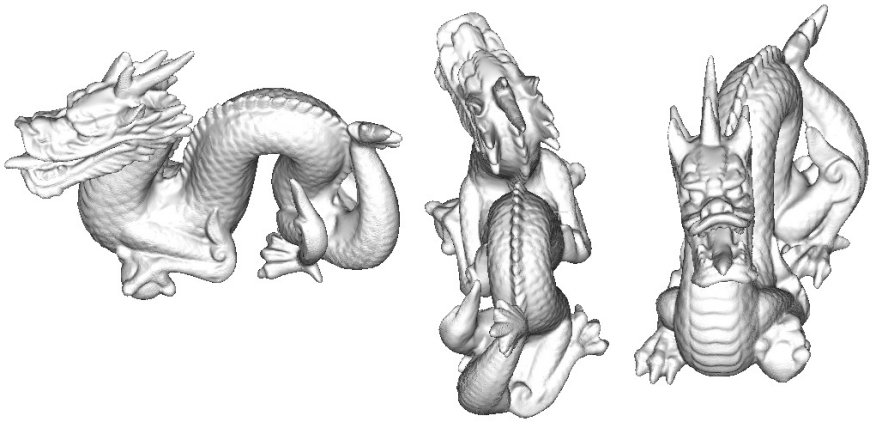
surface reconstructions. For example, in the case of the Pup model and Super Mario model, the point density is lower and hence the surface reconstruction is not rich in details as in the case of laser scanned dataset.

(a) Buddha

(b) Armadillo

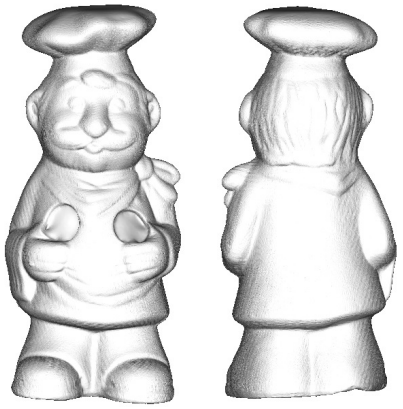


(c) Bunny

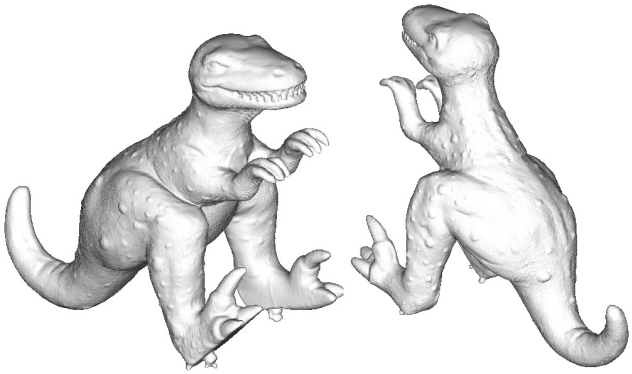


(d) Dragon

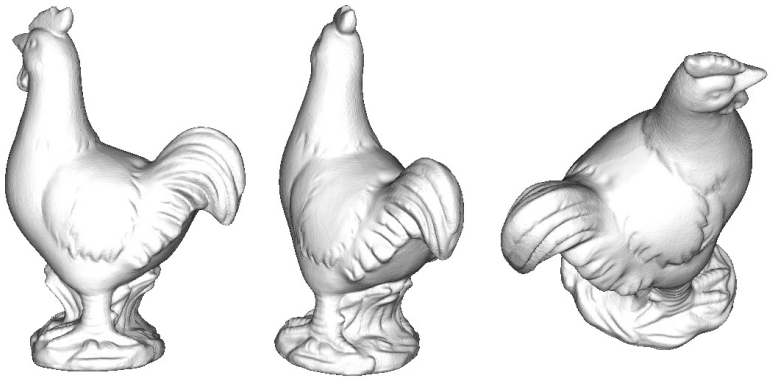
Figure 5.34: Stanford dataset: the surface reconstructed models created from the registered point clouds.



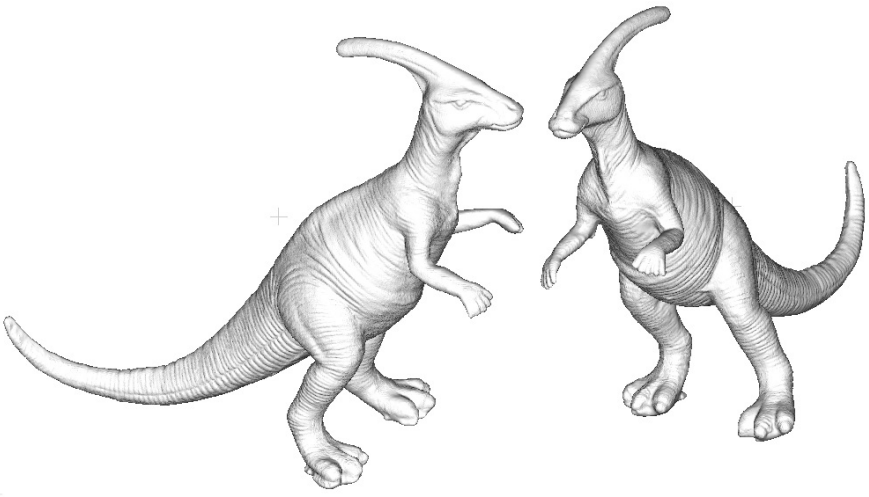
(a) Chef



(b) TRex



(c) Chicken



(d) Parasaurolophus

Figure 5.35: UWA dataset: the surface reconstructed models created from the registered point clouds.

(a) Super Mario

(b) Duck

(c) Pup

Figure 5.36: Kinect Models: the surface reconstructed models created from the registered point clouds.

5.3.1 Scene Reconstruction

Kinect sensor: The proposed framework was evaluated for the case of scene reconstruction as well. A sequence of scans of an indoor scene in our office space was acquired using the Kinect sensor. The sensor was placed on a table and scans were obtained by rotating the scanner. These scans were then registered using the proposed framework. A set of the input scans considered are shown in Fig.5.37. Since the data was captured without moving the sensor around in the room, the input scans contain holes and are not accounted for in the current study. The holes can be avoided by capturing scans by moving the sensor in a linear motion around the room walls. The limited field of view of the sensor allowed only a portion of the wall to be captured in a scene. This data can be captured by moving the sensor in a vertical translation mode. Two views of the registered room scene are shown in Figs.5.38 - 5.39.

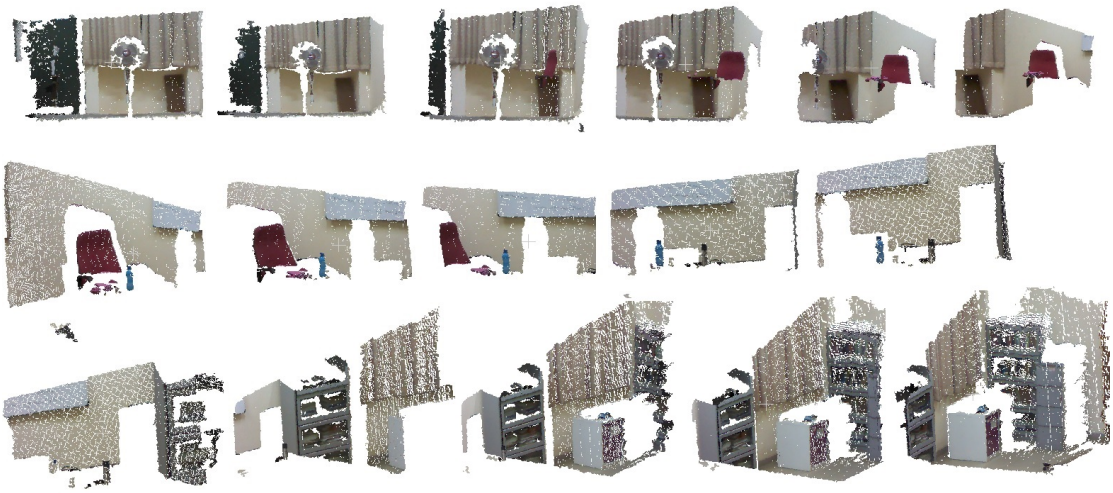


Figure 5.37: A set of input Kinect scans of the scene to be registered.

Terrestrial Laser Scanner: A terrestrial laser scanner by FARO - FARO FOCUS^{3D} X 330 - was used to capture the views of a room scene (our lab). Since the sensor was able to acquire data at a span of 360°around the scanner in a single scan, multiple views were acquired to account for the occlusion in the scene. Multiple views were registered using the proposed framework. The ceiling data is clipped for visual clarity in the results. The set of input scans captured are shown in Fig.5.40. Different views of the final registered scene are shown in Figs.5.41 - 5.42.

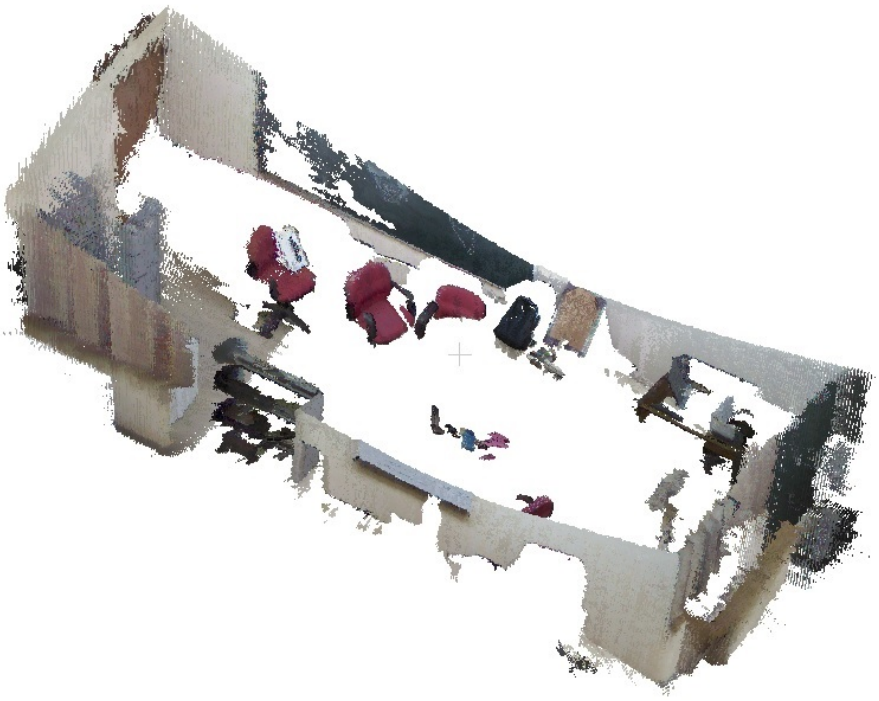


Figure 5.38: A view of the registered room scene.

Figure 5.39: A view of the registered room scene.

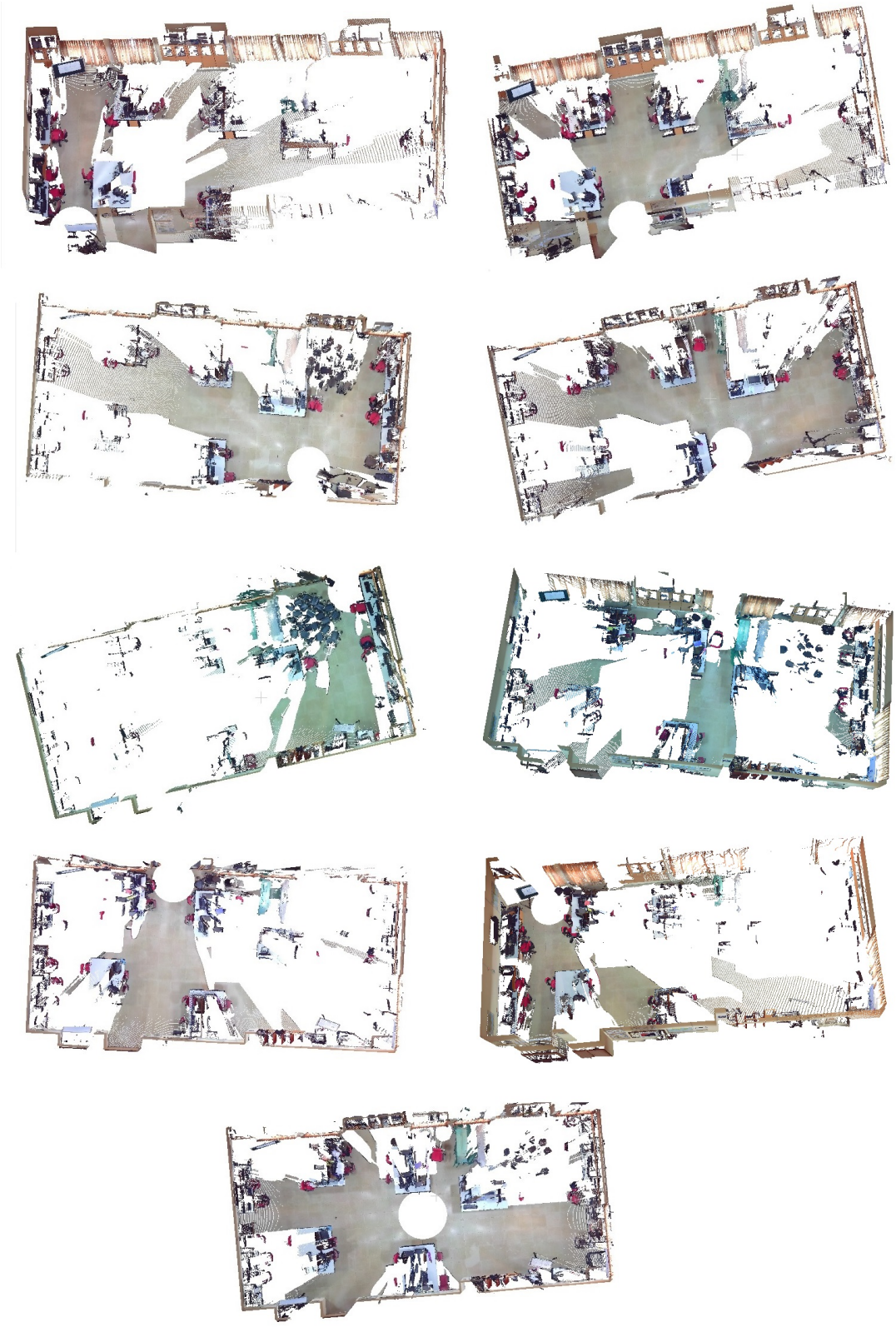


Figure 5.40: Input scans of the room scene acquired by FARO scanner.

Figure 5.41: A view of registered room scene.

Figure 5.42: A view of registered room scene.

From the results obtained on scene reconstruction and object reconstruction, it is observed that the proposed framework is successful in obtaining automatic global registration. The case of point clouds with and without colour or intensity data was evaluated. Data acquired from multiple sensors were considered for registration.

Table 5.2: The average registration error for the various models considered obtained by the proposed framework.

Model	**Average Registration Error**	**Resolution**
Buddha	1.01E-07	0.00035
Dragon	1.12E-07	0.0006
Bunny	1.15557E-07	0.0006
Armadillo	8.84626E-08	0.0006
Chef	0.098988306	0.55
Chicken	0.082089355	0.55
Parasaurolophus	0.106290889	0.55
TRex	0.098203306	0.55
Duck	3.10157E-06	0.0016
Super Mario	2.36346E-06	0.0016
Pup	1.05978E-06	0.0013

The average distances between the corresponding points of the adjacent overlapping scans of different models were calculated. The relative transformations considered were calculated from the final global transformation estimates obtained by robust motion averaging. Each pair of adjacent overlapping scans in the scan sequence was first aligned by the estimated relative transformations. Then the corresponding points were identified as the nearest neighbour points in the other point cloud within a predefined threshold. The distance between each adjacent pair was calculated as the average distance between the corresponding points. Then the mean distance for a model was computed as the average distance for all adjacent pairs. This mean distance points to the average registration error by the proposed framework and the average registration error obtained for different models is shown in Table 5.2. The unit is same as that of the unit of measurement of the point cloud. As the measurement unit is not provided for the models considered, the approximate resolution of the point clouds are listed for reference in Table 5.2.

In the case of cycle of scans for object reconstruction, the loop closure error was evaluated. The first and last scans have the maximum amount of overlap in a cycle. The initial and final scans were first aligned according to the relative transformation

calculated from the global transformation estimate. The corresponding points are the nearest neighbours between the two point clouds, which are at a distance less than a given threshold. Loop closure error was calculated as the average sum of distances between the corresponding points of the initial and final scans. The number of corresponding points within the threshold and the average distance between them were listed for a set of models registered. This is given in Table 5.3. The loop closure error was evaluated on the global registration obtained after initialisation by sequential alignment and after final alignment by the proposed framework. A threshold was applied to find the corresponding points in order to exclude the points which do not have corresponding points due to partial overlap. Large number of corresponding points represent a good registration as it indicates the number of nearest points in the two scan views within a given threshold after registration. In the case of a cycle, it indicates that a good loop closure is achieved since the first scan matches well with the final scan in the loop.

From the Table 5.3, it is observed that the number of corresponding points obtained has increased after final alignment when compared to that of the sequential alignment. This suggests a better alignment. The loop closure error which indicates the average distance between the corresponding points depicts the loop closure registration accuracy. From the results, it is clear that the proposed framework has greatly improved the registration accuracy. The alignment obtained by sequential registration is prone to accumulation of errors which leads to large loop closure errors. After the robust multiview alignment, this error is minimised. The accuracy obtained is comparable to the point resolution of the scans considered.

The average registration error in Table 5.2 for each model is comparable with the loop closure error listed in Table 5.3 which shows that the loop closure is achieved without compromising much on the relative transformation accuracy. This is because of the averaging of errors by the robust motion averaging method. The robust motion averaging averages out the accumulated error with the help of redundant pairs of relative motions. This is superior to the error distribution strategy of the existing methods, where the accumulated error is simply distributed across the different scans, thus lowering the relative registration accuracy of the different scans.

Table 5.3: Comparison of loop closure errors and number of corresponding points between first and last scans after initial sequential global registration and after the final registration by robust motion averaging.

	Sequential alignment		**Final alignment**	
Model	**Number of Corresponding Points**	**Loop Closure Error**	**Number of Corresponding Points**	**Loop Closure Error**
Buddha	5791	1.68E-06	11224	9.94E-08
Dragon	31661	1.24E-07	31790	9.98E-08
Bunny	9455	7.13E-07	9984	3.02E-07
Armadillo	16175	1.87E-07	16435	1.59E-07
Chef	11164	0.69483	13926	0.0976393
Chicken	15946	0.157223	19858	0.081595
Parasaurolophus	15164	0.180105	25601	0.107609
TRex	22968	0.482873	27139	0.0933125
Duck	787	7.38E-06	938	3.81E-06
Super Mario	4283	7.31E-06	6463	2.14E-06
Pup	1696	4.09E-06	5503	1.16E-06

In the proposed framework, a sequence of scans of a model or scene and the list of overlapping pairs are provided as input. The list of overlapping pairs is given assuming that the scans are taken in a sequence. The adjacent scan pairs and the expected overlapping pairs are listed. The adjacent scan pairs are assumed to result in a good pairwise registration. This constraint is due to the initialisation of global motion by sequential registration. The correct relative motions are identified based on the initial global motions.

5.4 Chapter Summary

In this chapter, the automatic registration framework for a sequence of partially overlapping 3D scans is detailed. The framework is fed with a sequence of scans of an object or a scene, along with the expected list of overlapping pairs. After processing, a complete point cloud of the object or scene is obtained as output. Pairwise registration is achieved using the pairwise registration algorithm detailed in Chapter 3. Multiview registration is obtained by robust motion averaging detailed in Chapter 4. The registration of a set of models and scenes from 3D scans using the proposed framework is performed and the results indicate the efficiency of the proposed approach.

CHAPTER 6

Automatic Registration of Remote Sensing Satellite Images

Prelude: Image registration is a crucial preprocessing task in many remote sensing image analysis scenarios. The robust estimation algorithm developed for coarse pairwise alignment of point clouds is adapted to perform the automatic registration of remote sensing satellite images. Registration of images of various spatial, spectral and radiometric resolutions was performed for evaluation.

6.1 Introduction

The algorithm for pairwise alignment of 3D point clouds presented in Chapter 3 represents a general robust estimation method which can be employed in similar problems where a guided sampling procedure can be devised. The case of automatic registration of remote sensing satellite images, where a measure of the quality of the feature matches is available for guided sampling, is considered in this chapter. The adaptability of the algorithm to the automatic image registration problem is examined.

Image registration is a crucial step in all image analysis tasks like image fusion, change detection, and multichannel image restoration where the final information is gained from the combination of various data sources. Typically, registration is required in remote sensing (multi-spectral classification, environmental monitoring, change detection, image mosaicking, creating super-resolution images, integrating information into geographic information systems (GIS), in cartography (map updating), and in automated image analysis systems) for detecting and monitoring dynamic events in civilian and military domain etc. During the last two decades, there has been tremendous development in the variety and extent of remote sensing image acquisition. A range of remote sensors are available offering multispectral images at very high spatial resolution and with temporal resolution of a few days. Furthermore, recent developments in ground and UAV-borne multispectral sensors offer image ac-

quisitions with 360° field of view and enable continuous acquisition of remote sensing images over the target area.

The image registration process requires handling of a range of images captured from different sensors and with different resolutions with high throughput. The automatic image registration process involves three main steps: (i) correspondence matching, (ii) transformation model estimation, and (iii) image resampling. Correspondence matching, popularly called as generation of control points, involves automatic detection and matching of corresponding points in the reference and target images (For the case of satellite image registration, the image to be aligned is termed as target and the image to which target is to be registered is called reference). The methods of correspondence matching can be classified into two categories: intensity based matching (e.g. normalized cross correlation (NCC) matching (Pratt, 2001; Rohr et al., 2001), and feature based matching (e.g. scale invariant feature transformation (SIFT) (Gonzalez and Woods, 1992). In the second step, transformation model estimation, parameters of mapping functions which align the reference image and target image are estimated. Once the appropriate mapping function parameters are estimated, in the third step, image resampling, the corrected coordinates of target image are computed by appropriate interpolation technique. Amongst the three steps, the second step, estimation of robust mapping functions, is critical to the success of automated satellite image registration (Kim and Im, 2003; Goncalves et al., 2011; Hu et al., 2014). For this, the robust estimation algorithm, random sample consensus (RANSAC) (Fischler and Bolles, 1981) and its variants (Chum et al., 2003; Chum and Matas, 2005) have been the primary methods in use. The estimation of robust mapping functions in the presence of numerous outliers is challenging for satellite image registration as the satellite images captured at different periods, view angles and resolutions can result in a lot of outliers in the matching points (Ton and Jain, 1989). Accuracy and speed are two important characteristics required for registration of high resolution large throughput remote sensing data. The RANSAC algorithm has been used for automated registration of medium resolution satellite images (Torr et al., 2001; Kim and Im, 2003; Fan et al., 2013; Wang et al., 2015).

The RANSAC is an iterative algorithm operating in a hypotheses setting - and - verification framework. A minimal subset of the matching points is randomly selected and model parameters are estimated. The number of points in the minimal subset is

the minimum required to estimate the model. This minimal subset is selected so as to minimise the influence of outliers in the model generation set. The model estimated is then evaluated on the entire dataset and the number of correct matching points (inliers) consistent with the model is determined. This hypotheses setting - and - verification loop is repeated until the probability of not finding an all-inlier sample (and hence a good model) falls below a threshold. RANSAC algorithm has been found to be robust to various levels of image deformations and could offer accurate image registration. However, as the number of iterations depends on factors like number of corresponding points and the percentage of the inliers, the number of samples required to estimate an optimal solution increases exponentially leading to slow convergence and substantial computational cost. Especially, the rich geometric details of surface objects in high resolution satellite images help to accumulate large number of potential correspondence matches in the reference image and target image. There have been several variants of the RANSAC algorithm (R-RANSAC, MAPSAC, LoSAC, NAPSAC, PROSAC - a review of these methods are available in Pankaj and Nidamanuri (2016)) which are aimed at improving the robustness, accuracy or speed of RANSAC algorithm for various computer vision applications. While some of these variants aim to optimize the model verification thereby improving the accuracy (e.g. LoSAC, MAPSAC), others seek to guide the sampling process to generate more useful hypotheses thereby increasing the speed (e.g. R-RANSAC, PROSAC). While these improved algorithms have shown considerable promise in typical computer vision based applications, they compromise on one of the two main characteristics, speed, and accuracy in situations where fast and accurate registration is essential.

An improved robust estimation algorithm, named "ProLoSAC" is presented in Chapter 3 for accurate registration of 3D point cloud in less computation time. Point cloud registration is implemented as a two step, initial and final alignment scheme, popularly called coarse-to-fine registration scheme. A global registration (coarse registration) is implemented as an initial solution which would be optimized in the fine-tuning process. Coupled with the advantages of a robust estimator, ProLoSAC can be used for matching model estimation with feature based or intensity based matching points generation methods. The performance of the ProLoSAC for various 3D point cloud registration tasks has been found to be relatively accurate and fast. The proposed algorithm in principle can be adapted for serving as an efficient robust estimator for

automatic registration of satellite images. Hence, the main goal of this study is the adaptation and implementation of the proposed robust estimator (ProLoSAC), for the registration of satellite images of various spatial, spectral and radiometric resolutions. Remote sensing satellites images from Pleiades-1A, GeoEye, and WorldView-2, and Landsat-8 OLI are registered automatically using the proposed method. The performance of the proposed method has been compared with that of the classical RANSAC algorithm for speed and accuracy.

6.2 Methodology

Pertaining to the key steps involved in the automatic image registration, Fig.6.1 outlines the methodological process flow evolved for registration of satellite images. The featured detector and descriptor, SURF (Speeded Up Robust features) is adopted for generating the corresponding points. Akin to the popular SIFT (scale-invariant feature transform) feature detector and descriptor, SURF is an improved method for detecting and describing matching points in reference and target images. The major advantage of using SURF as feature detector and descriptor, compared to SIFT, is the computation efficiency achieved by SURF due to its working on integral images for computing the convolution operations. More information on the SURF feature detector can be found on Bay et al. (2006). The ProLoSAC was applied on the corresponding points (obtained by matching SURF features) to eliminate outliers and for generating the optimal transformation.

Reference and target images are aligned by deriving the required geometric transformation parameters from the widely used affine transformation model in Eq.6.1, where (x, y) and (x', y') are the coordinates of the target and reference images respectively and $(a_{11}, a_{12}, a_{21}, a_{22})$ represent the scale, rotation and shearing parameters, and (a_x, a_y) are the translation parameters. The intensity values are re-sampled using the nearest neighbourhood interpolation technique.

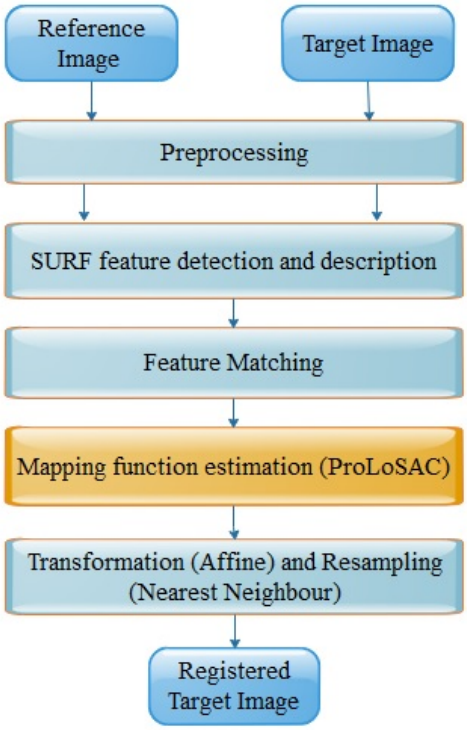


Figure 6.1: Key steps involved in automatic image registration.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_x \\ a_{21} & a_{22} & a_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{6.1}$$

Given the source and target images, the image registration is performed in the following steps.

- **Preprocessing**: preprocessing of images like noise filtering, multiple bands to grayscale conversion, intensity normalisation etc are carried out.
- **Keypoint detection and feature extraction**: SURF feature points are detected and SURF features are extracted from both the images
- **Feature Matching**: here the extracted target and reference features are matched by finding the nearest neighbour in the feature space (using Bruteforce L2 matcher in opencv). The extracted matches are refined by selecting only a fraction of the matches sorted according to their distance.

- **Transformation estimation**: the mapping function to align the source image with the target image is estimated using robust estimation technique (ProLoSAC - explained in Chapter 3).
- **Resampling**: the source image is aligned with the target image using the estimated affine transformation and the nearest neighbourhood resampling technique.

In the workflow devised, focus is on the applicability of the proposed algorithm for the satellite image registration task in the place of RANSAC. Hence the experiments are focussed on comparison of RANSAC and the proposed approach rather than optimisation of the image registration problem itself.

6.3 Experiments and Datasets Used

The proposed ProLoSAC robust estimator was applied for the registration of various remote sensing images encompassing a range of land cover settings and acquired at different periods with different spatial and radiometric resolutions. Table 6.1 shows the list of images used for registration experiments in this study. In order to assess the performance of the proposed methodology relative to the original RANSAC algorithm, image registration experiments were carried out using the similar methodology but replacing the ProLoSAC with RANSAC. The accuracy and computational parameters of both the algorithms were compared across the remote sensing images considered. All the image registration experiments were carried out on a desktop computer with Intel Core processor of 3.2GHz and 4GB memory.

6.3.1 Datasets

This study used five different remote sensing images for experimental implementation of the proposed automated image registration. The datasets used are summarised in Table 6.1.

Table 6.1: The datasets used in this study

S.No	Sensor	Spatial resolution	Location
1	Pleiades-1A	Multispectral: 2 m Panchromatic: 0.5 m Date of acquisition: 31 January, 2013	Melbourne, Australia
2	GeoEye-1	Multispectral: 2 m Panchromatic: 0.5 m Date of acquisition: 11 April, 2015	Trivandrum, India
3	Landsat-8 OLI	Multispectral: 30 m Panchromatic: 15 m Date of acquisition: 11 April, 2015	Kolleru lake, India
4	Resourcesat-1 IRS-AWiFS GeoEye-1	Multispectral: 56 m Multispectral: 2 m Date of acquisition: 14 October, 2013	Trivandrum, India
5	AISA Eagle (airborne)	Hyperspectral image: 4 m Date of acquisition: 2 April, 2011 and 9 August 2011	Isabena catchment, Spain

The first dataset was acquired over a part of Melbourne, Australia by the Pleiades-1A satellite on 31 January 2013. The dataset contained a multispectral image at 2 m spatial resolution and four spectral bands (near-infrared, red, green, and blue) and a panchromatic image acquired at 0.5 m spatial resolution. The multispectral image was considered as the target image for registration. Dominant urban land covers, building, roads etc are the major land covers in the study area. The second dataset was acquired over Veli region of Trivandrum, India by the GeoEye-1 satellite on 11 April 2015. Similar to the Pleiades-1A data, the GeoEye-1 dataset used in this study contained a multispectral spectral image of four spectral bands at 2 m spatial resolution and a

panchromatic band at 0.5 m spatial resolution. The major land cover setting is the extensive coverage of dense woody and shrub trees across the area and occupies more than 50% of the land area followed by a backwater lake. The panchromatic image was used as reference image.

The third image dataset consisted of a multispectral image acquired by Landst-8 OLI sensor at 30 m spatial resolution and a panchromatic image acquired at 15 m spatial resolution. The panchromatic image was used as the reference for image registration. The major land cover is the wetland consisting of shrimp farms in and around the freshwater lake Kolleru in Andhra Pradesh, India. The fourth dataset was acquired from two different satellites; a multispectral image acquired from the Indian Resourcesat-1 AWiFS sensor at 56 m spatial resolution and a multispectral image acquired by GeoEye-1 satellite at 2 m spatial resolution. The major land cover is an urban setting wherein woody vegetation such coconut palms and jackfruit tree form major land covers. Here the GeoEye-1 image was considered as the reference image. The fifth dataset is an airborne temporal hyperspectral imagery acquired over part of the Isabena river catchment area in Spain. The data was acquired at 4 m spatial resolution and in 450 spectral bands covering the entire optical wavelength region. Semi-agriculture and rural-urban built up form major land covers in the area covered by the images used. The images were acquired in April and August 2011. The image acquired on August 2011 was taken as the reference for registration.

6.3.2 Validation of the automatic image registration

The performance of the proposed algorithm on the image registration problem was compared with that of RANSAC. The registration accuracy was evaluated by computing the alignment error per correspondence match. The root mean square value of the Euclidean distances between the corresponding points was calculated after discarding the point pairs which have reprojection error greater than a specified threshold. This is termed as the RMSE metric. The lesser the RMSE value, the estimated transformation is expected to be more accurate. The execution time of the transformation estimation loop was estimated for RANSAC as well as the proposed approach. The number of iterations taken by both algorithms to converge to a solution was compared. The total number of inliers to the estimated transformation was also compared. The correspond-

ing points with reprojection error less than a specified threshold were considered as inliers.

Another method of assessment considered is the "inverse consistency" (Christensen et al, 2006). This is a non-parametric method for the assessing the consistency of image registration by changing the roles of images used for reference and target. The transformation model is assessed for 'invertibility' property between forward transformation and reverse transformation. In the forward transformation, target image is transformed to match to the reference image. Reverse transformation transforms the reference image to match with the target image. For a consistent matching between two images, in principle, forward transformation should be equal to the inverse of reverse transformation. However, as Christensen et al. (2006) observed, this theoretical consistency is not observed in practical image registration tasks. This may be due to inadequacy of the matching function algorithm criteria to uniquely describe the correspondence between the images to be registered. As a result, the quantification of the inverse consistency by the RMSE metric indicates the quality of image registration.

6.4 Results and Analysis

The registration performance by automatic image registration work flow is detailed in this section. The registered pair are shown as image swipes (overlays) to appreciate the geometric conformity. It can be easily observed that, in the registered image pair, the features overlay exactly and there is no artefact seen in the image swipe.

Pleiades-1A : The multispectral image (2 m resolution) of part of Melbourne captured using Pleiades-1A sensor was registered with the panchromatic image of 0.5 m resolution. Fig.6.2 shows the registered image shown by overlaying the images. Feature matches in the target and reference images, which were identified as inliers to the estimated transformation are shown in Fig.6.3. The matches were identified in a single band of the multispectral image.

Figure 6.2: The automatic registered false color composite of the Pleiades-1A multispectral image with 2 m spatial resolution overlayed with the panchromatic image with 0.5 m spatial resolution.

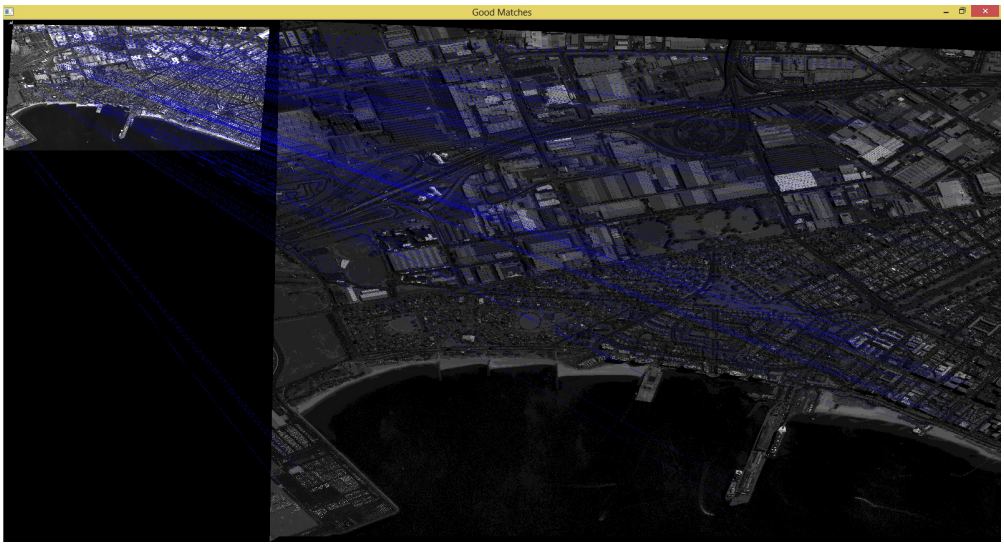


Figure 6.3: Pleiades dataset: the inlier feature matches in target and reference images.

GeoEye-1: A multispectral image with 2 m spatial resolution of Veli region of Trivandrum, India and a panchromatic image of 0.5 m spatial resolution were considered for registration. The transformation parameters were determined and applied to align the multispectral image with the panchromatic image. The transformed result with images overlayed is shown in Fig.6.4. The inlier feature match point pairs identified are shown in Fig.6.5.

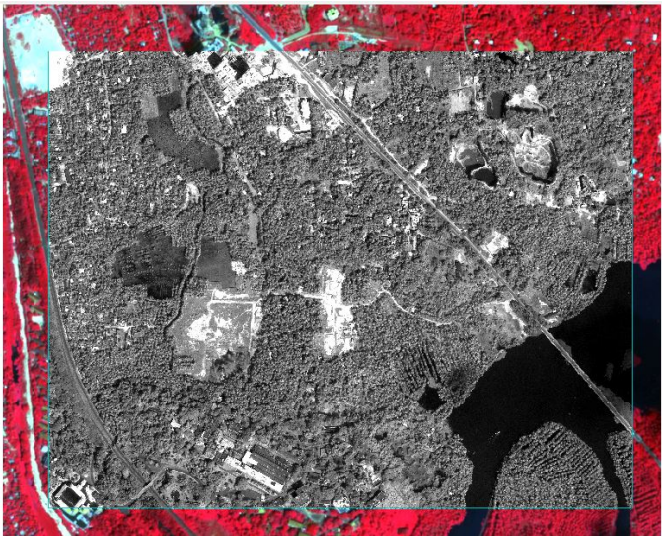


Figure 6.4: The automatic registered false colour composite of the GeoEye-1 multispectral image with 2 m spatial resolution overlayed with the panchromatic image of 0.5 m spatial resolution.

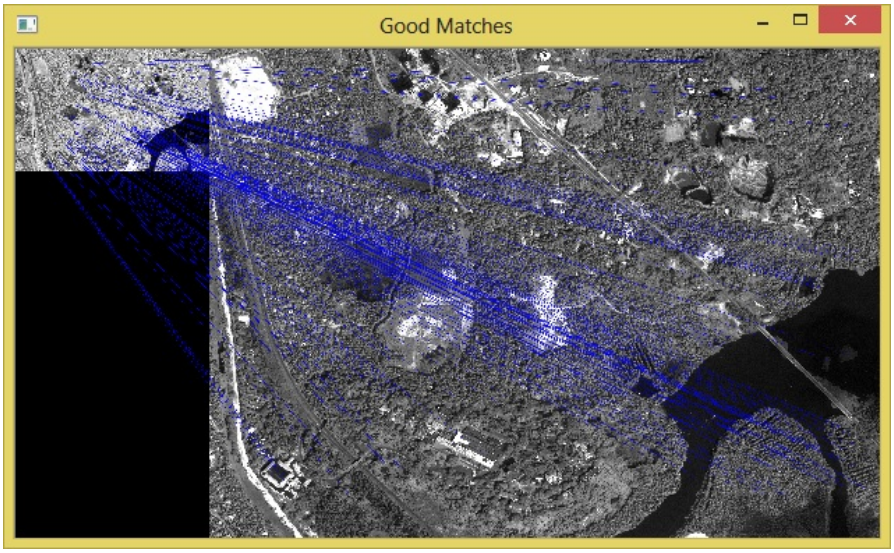
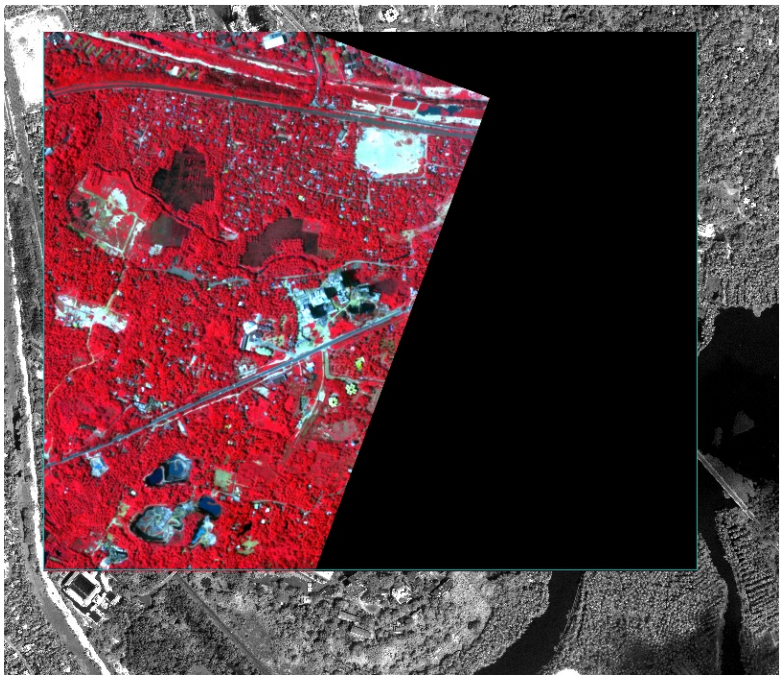
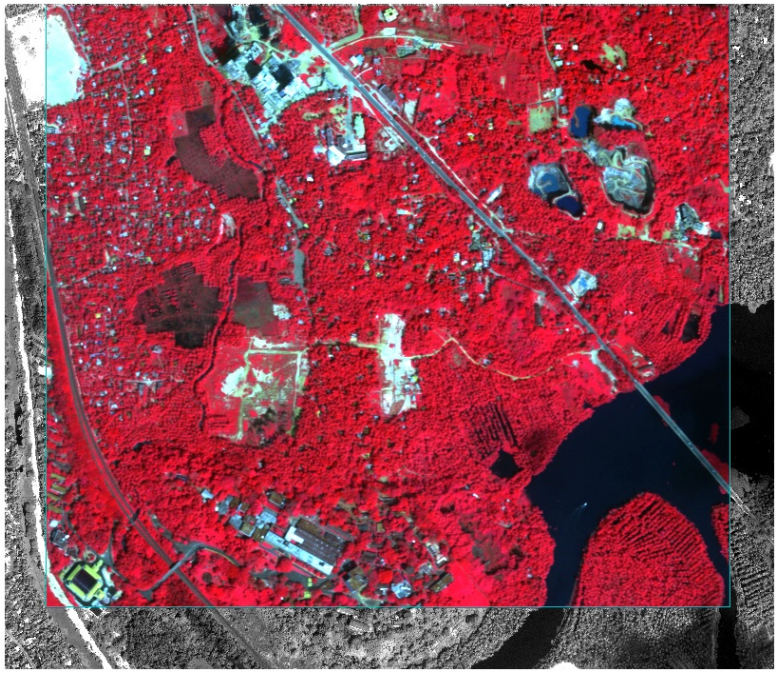


Figure 6.5: GeoEye-1 dataset: the inlier feature matches in target and reference images.



(a) RANSAC



(b) ProLoSAC

Figure 6.6: The automatic registered false color composite of the GeoEye -1 multispectral image obtained using a) RANSAC, and b) ProLoSAC using a limited set of keypoints, shown overlayed with the panchromatic image. RANSAC failed to converge to a global minimum whereas the proposed algorithm converged to a global minimum.

The keypoints extracted were pruned according to their response in order to limit the execution time in later stages. In the case of this dataset, experiments were conducted with the complete set as well as with pruned set of keypoints. When the number of keypoints was limited, RANSAC failed to converge to a global minimum. However, the proposed algorithm converged to a global minimum and obtained a reasonably good transformation matrix even when the number of keypoints was limited. The result of registration with RANSAC and the proposed algorithm with limited set of keypoints are shown in Fig.6.6. When the number of keypoints were limited, the execution time spent in further stages like feature extraction, matching and transformation estimation were reduced. Thus the result indicates that the proposed algorithm was able to find a good transformation matrix even when RANSAC failed to do so.

Landsat-8: A Landsat-8 OLI multispectral image with 30 m spatial resolution of part of Kolleru lake, India and Landsat-8 panchromatic image of 15 m spatial resolution were registered using the proposed approach. Compared to the previous dataset, images of lesser resolution were considered in this dataset. The transformation parameters were determined and applied to align the multispectral image with the panchromatic image. The transformed result with images overlayed is shown in Fig.6.7. The correct feature matches identified by the proposed approach are shown in Fig.6.8.

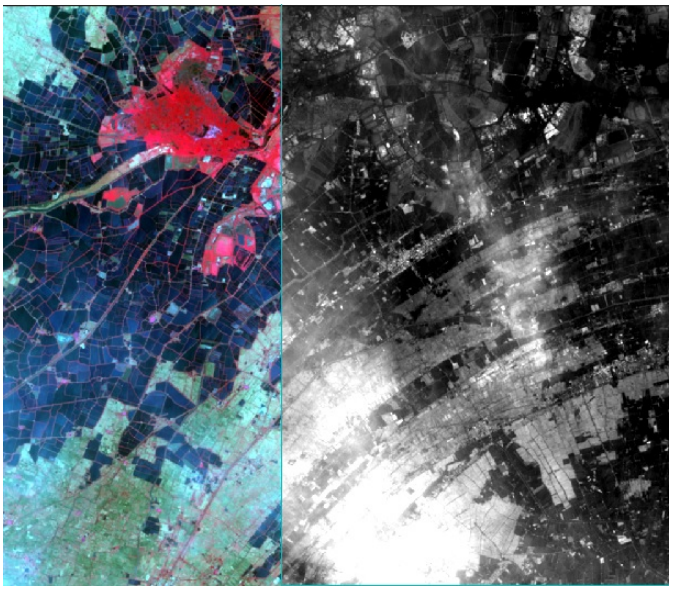


Figure 6.7: The automatic registered false colour composite of the Landsat-8 OLI multispectral image with 30 m spatial resolution overlayed with the Landsat-8 panchromatic image of 15 m resolution.

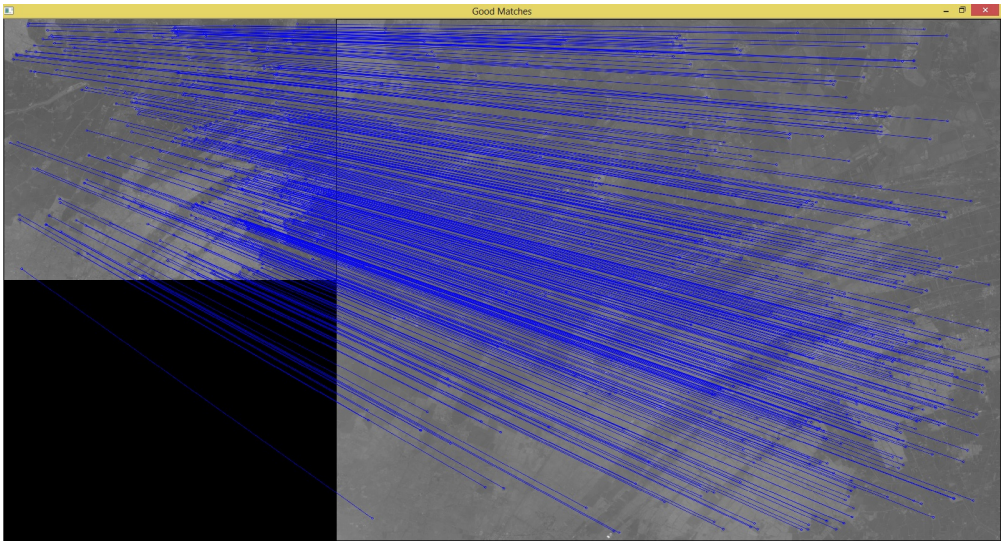


Figure 6.8: Landsat-8 dataset: the inlier feature matches in target and reference images.

Resourcesat-1 AWiFS and GeoEye-1: In this dataset, multispectral images with different spatial resolutions acquired by two different satellites of the part of Trivandrum area were registered. A multispectral image acquired by the Indian Resourcesat-1 AWiFS sensor at 56 m spatial resolution acquired on 14 October 2013 and a multispectral image acquired by GeoEye-1 satellite at 2 m spatial resolution acquired on 31 January 2013 were registered and the resulting false colour composite images are shown overlayed in Fig.6.9. The match points identified in the two images are shown in Fig.6.10. The Geo-Eye multispectral image was downsampled before feature matching to deal with the large variation in resolution of the target and reference images.

AISA airborne hyperspectral imagery: In this dataset, two hyperspectral images of the part of Isabena river catchment area acquired during two different times in a year were considered for registration. The images were acquired by the AISA airborne sensor on 2 April 2011 and 9 August 2011, and are of 4 m spatial resolution. The automatic registered false colour composite images are shown overlayed in Fig.6.11. The match points identified are shown in Fig.6.12.

Figure 6.9: The automatic registered false color composite of the Resourcesat-1 AWiFS multispectral image with 56 m spatial resolution overlayed with the GeoEye-1 multispectral image with 2 m spatial resolution.

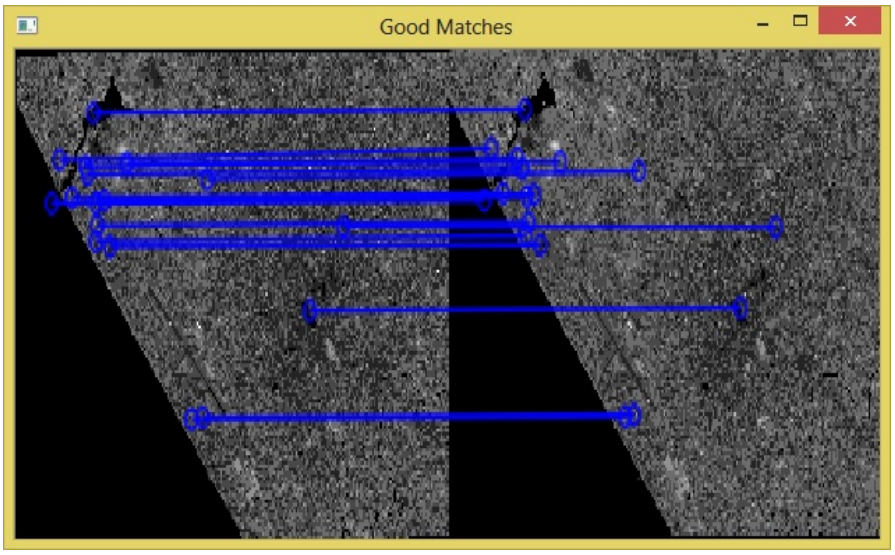


Figure 6.10: Landsat-8 dataset: the inlier feature matches between target and down-sampled reference images.

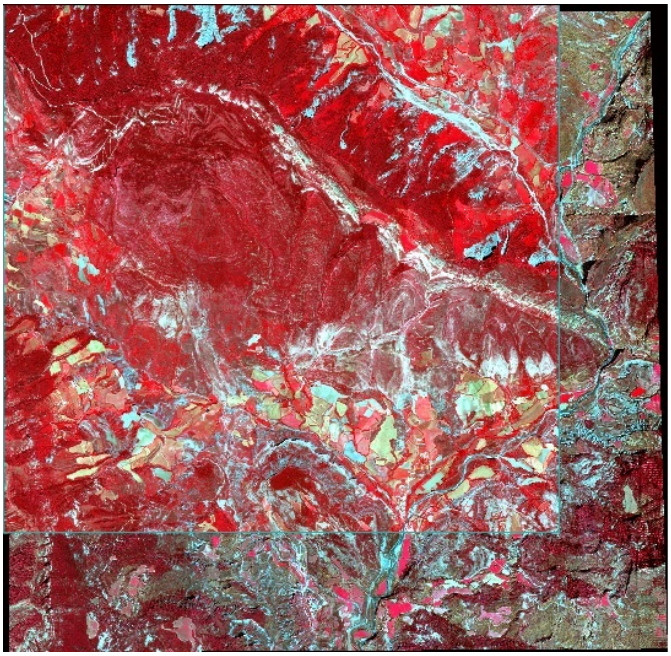


Figure 6.11: The automatic registered false color composite of the two AISA airborne hyperspectral images with 4 m spatial resolution acquired on 2 April 2011 and 9 August 2011.

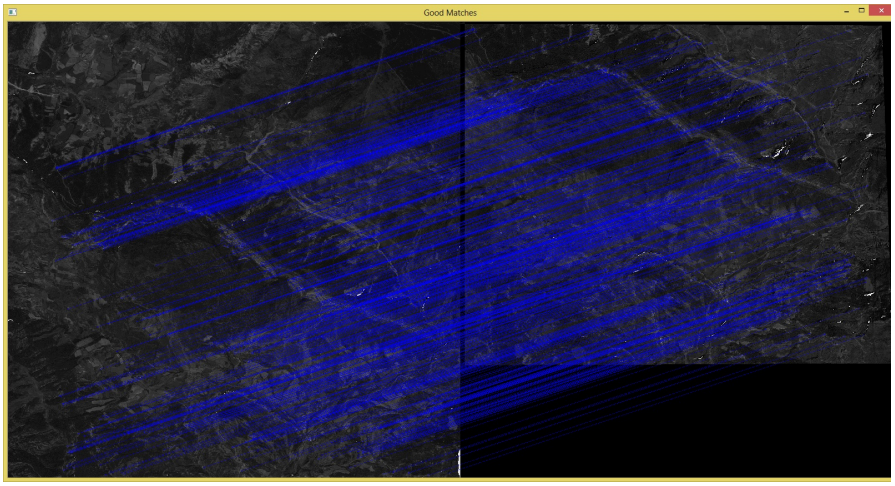


Figure 6.12: AISA airborne hyperspectral dataset: the inlier feature matches between target and reference images.

Comparison with RANSAC: The performance of the proposed robust estimation method on estimating the transformation parameters for automatic satellite image registration was compared with that of the de-facto standard algorithm RANSAC. The accuracy and computational cost of the two methods on estimating the transformation from a set of feature corresponding points were compared. The comparative results obtained for the considered datasets are summarised in Table 6.2. It can be observed

that the proposed robust estimation technique performs better than RANSAC both in terms of accuracy and computational time. This emphasises that the proposed technique is generic and can be applied to both 3D point cloud registration and image registration tasks.

Table 6.2: Assessment of the accuracy and computational cost of image registration.

Image Pair	RMS error		Time (seconds)	
	RANSAC	ProLoSAC	RANSAC	ProLoSAC
Pleiades -1A multispectral image (target), and panchromatic image (reference)	1.47867	1.46878	41.957	11.474
GeoEye-1 Multispectral image(target), and panchromatic image(reference)	1.74499	1.27038	81.507	15.742
Landsat-8 OLI multispectral image (target), and panchromatic image (reference)	1.1113	1.01434	0.042	0.178
Resourcesat-1 -AWiFS multispectral image (target), and GeoEye-1 multispectral image (reference)	1.50698	1.49765	3.813	0.616
AISA hyperspectral image April 2011 (target), and August 2011 (reference)	1.50397	1.36693	6.432	0.665

For high resolution images like Pleiades, Geo-Eye and AISA, it can be observed that the proposed method has a consistent computational advantage over RANSAC. For low resolution images like Landsat, RANSAC converged faster than ProLoSAC, even though the accuracy obtained by ProLoSAC is superior. The percentage of inliers in the set of matches is high (near to 0.5%) for Landsat dataset, compared to the other datasets (near to 1%). This indicates that when the inlier percentage is high, both the algorithms converge faster to a global minimum. In this case, the local op-

timisation stage by ProLoSAC slightly increases the computational load. However, it can be observed that the local optimisation still leads to a better (accurate) solution when compared to RANSAC, with a slight overhead of computational load. When the percentage of inliers is low, the guided sampling by the proposed approach offers superior results compared to RANSAC. In this case, the computational overhead by local optimisation is insignificant compared to the computational advantage obtained by probing the good matches initially by the guided sampling.

A lot of self repeating structures in the urban land cover settings in the Pleiades dataset, which lead to incorrect matches may be attributed to the low accuracy obtained by both the algorithms. For the Resourcesat and Geo-Eye dataset registration, images which differ more than 25 times in the spatial resolution were considered. It was observed that the application of automated image registration methods in such cases leads to large registration errors. A downsampled version of the high resolution image was used for registration in this case in order to match the features reliably. As can be seen for the hyperspectral images, there is relatively higher RMS error by both methods. This can be explained from the fact that the airborne hyperspectral images were acquired at different temporal period and over agricultural dominant areas. The temporal changes in the land cover classes pose serious limitations on the matching pixels generation. This would have caused the relatively lower accuracy for these images.

The number of inliers among the feature matches identified by the proposed algorithm and RANSAC was estimated and compared. A feature match is considered an inlier if the reprojection error of the point pair is less than a given threshold. The reprojection error of a point pair is the distance between the reference point and the transformed target point. The total number of iterations in the main transformation estimation loop of RANSAC and the proposed approach were compared. The maximum number of iterations was limited to 50001. Table 6.3 presents the comparison between the proposed approach and RANSAC in terms of the number of inliers and the number of iterations for convergence. From the Table 6.3, it is observed that the proposed algorithm requires lesser number of iterations for convergence compared to RANSAC. The inner loop inside the proposed algorithm is not considered. Higher number of inliers to the detected transformation by the proposed algorithm indicates better registration accuracy.

Table 6.3: Assessment of the number of inliers and iterations for convergence.

Image Pair	Iterations		Number of Inliers	
	RANSAC	ProLoSAC	RANSAC	ProLoSAC
Pleiades -1A multispectral image (target), and panchromatic image (reference)	31440	7819	349	350
GeoEye-1 Multispectral image(target), and panchromatic image(reference)	50001	8888	298	321
Landsat-8 OLI multispectral image (target), and panchromatic image (reference)	108	3	479	482
Resourcesat-1 -AWiFS multispectral image (target), and GeoEye-1 multispectral image (reference)	33075	5170	25	25
AISA hyperspectral image April 2011 (target), and August 2011 (reference)	3818	4	928	936

Validation using Inverse Consistency: The inverse consistency is confirmed by interchanging the target and reference images for performing registration. Both the reverse consistency and forward consistency RMS errors are compared in Table 6.4. From the Table 6.4, it is observed that the RMSE difference for both the methods are comparable in both the forward and reverse cases for the different datasets considered. In both cases, the general observation is that proposed approach outperforms RANSAC in terms of accuracy. While there is relative variation in the overall accuracy in the reverse and forward registrations, the difference between RANSAC and the proposed methods is consistent in terms of the accuracy metrics. The trend of superior performance of the proposed method across the high spatial and high spectral resolution images confirms the suitability of the proposed method for registration of high resolution remote sensing images.

Table 6.4: Assessment of inverse consistency of the image registration.

Image Pair	Inverse consistency (RMS error)		Forward consistency (RMSE) from Table 6.2	
	RANSAC	ProLoSAC	RANSAC	ProLoSAC
Pleiades -1A multispectral image (target), and panchromatic image (reference)	0.716082	0.688713	1.47867	1.46878
GeoEye-1 Multispectral image(target), and panchromatic image(reference)	1.9786	0.763202	1.74499	1.27038
Landsat-8 OLI multispectral image (target), and panchromatic image (reference)	0.723795	0.712386	1.1113	1.01434
Resourcesat-1 -AWiFS multispectral image (target), and GeoEye-1 multispectral image (reference)	1.62465	1.61726	1.50698	1.49765
AISA hyperspectral image April 2011 (target), and August 2011 (reference)	1.40624	1.38328	1.50397	1.36693

From the results detailed, it is observed that the proposed method exhibits superior performance in terms of accuracy and computational time when compared to the standard algorithm RANSAC. The computational advantage of the proposed approach is obtained by improving the random sampling strategy of RANSAC by making use of the quality of feature matches. The superior performance in accuracy is obtained by local optimisation of the estimated transformation. The model obtained by RANSAC may be sub-optimal and most often lies near to the actual optimum. The local refinement strategy helps in enhancing the accuracy. The fast but sub-optimal solution obtained by the guided sampling is compensated by the better accuracy obtained by local refinement. Thus the proposed algorithm ProLoSAC finds an accurate estimate in lesser computational time. The proposed algorithm has converged to a global solu-

tion even when the percentage of correct matches are as low as 1%, where RANSAC failed to converge to a global solution. The proposed algorithm was able to generate more number of matches which conformed to the estimated transformation (inliers), when compared to RANSAC. Thus the results suggest that the proposed algorithm can be adapted for automatic registration of remote sensing satellite images of various resolutions.

6.5 Summary

Automatic registration of images is a widely used and demanded task in computer vision related applications. Extension and application of the approaches to solve complex image registration scenarios encountered in large scale remote sensing images is a challenge. Several methods are being developed aimed at some or all the steps involved in remote sensing images registration.

This part of the thesis employed and adapted the robust estimation algorithm developed in Chapter 3 for automatic registration of satellite images which differ widely in spatial, spectral and environmental settings. The algorithm is implemented and evaluated for the registration of various remote sensing images. The method proposed has also been tested to register images of high resolution airborne hyperspectral remote sensing images. Results indicate that the proposed algorithm estimates an accurate transformation faster than the compared algorithms. Extensive validation of the results show that the proposed method is accurate and offer substantial computational advantage compared to the standard RANSAC based implementation of automated image registration. Especially, the method has been found apt to the automated registration of high spectral and spatial resolution remote sensing images with good accuracy while maintaining computational efficiency.

CHAPTER 7

Summary and Conclusion

The sensors to acquire 3D data has become ubiquitous and this necessitates the availability of advanced techniques to process the scanned 3D data. Due to the advent of high precision and low cost 3D sensors in the market, many industries which utilise 3D data for innovative products and processes are springing up at an unprecedented rate. The registration of 3D data captured from multiple view angles is often essential in order to employ them for further processing. This may be required in multiple scenarios such as generation of a complete 3D model of an object or a complete 3D scene view etc. Even when a complete model or scene is not generated, registration of multiple views may be necessary to deal with the holes caused by occlusions. The registration techniques are also used in object recognition problems which are often encountered in robotic 3D vision tasks. The automatic alignment of the captured 3D scans is essential for fast processing of 3D data in many application areas like robotic vision.

As answers to the research questions and objectives set in Chapter 1, this chapter summarises the major techniques, methods, experimental set up and results presented in this thesis. A complete framework for automatic registration of 3D point clouds is developed. The efficiency of the proposed approaches is demonstrated in the case of pairwise and multiview registration. The following section summarises the results and observations for each of the research problems addressed in this thesis. Some of the problems that can be addressed as an extension to the current work are also presented. For ease of readability, we summarise the major conclusions chapter-wise.

7.1 Pairwise registration of 3D point clouds

The role of pairwise registration in the 3D object reconstruction from scans is critical. The subsequent stages depend upon the success of pairwise alignment. The 3D point clouds captured from several view points are registered in multiple steps - pairwise registration and multiview registration. Sufficient overlap is required between the

pair of scans considered for automatic pairwise registration. The main advantage of 3D data over 2D images is the availability of geometric information. This geometric information of the object, present in the overlapping region of a pair of scans, is effectively utilised to find the transformation between them. The geometric features are extracted and matched across the pair of scans to form the set of corresponding points automatically. Chapter 3 presented a robust estimation method named ProLoSAC which estimates the best rigid body transformation supported by the maximum number of corresponding points.

The proposed ProLoSAC algorithm finds a global solution to the pairwise alignment problem. The major factors which affect the pairwise alignment problem include the percentage of overlap, availability of discriminable geometric features, presence of noise, presence of holes in data, variation in point density, presence of self similar (symmetric regions) or repetitive structures etc. The minimum number of scans required to reliably construct a complete 3D point cloud is still an open research problem and depends on many factors like the complexity of the object surface and self-occlusions.

The correctness of the feature matches depends on the richness of the object surface. A number of the 3D sensors in use today also features a 2D camera which captures the intensity information. If the objects do not have rich geometric details, the intensity or colour information can be utilised to aid the geometric features. If intensity information is used, illumination changes are to be considered. The features are selected based on the reliability of the basic information used to find the features. For example, if the normal estimation module is not accurate enough, 3D features like SHOT, which are constructed based on the distribution of normals in the neighbourhood, may not perform well. In such cases, a feature which is built on the point density information like 3D Shape Context may be more discriminable. The keypoint detection and feature extraction algorithms have to be selected considering many factors based on the nature of the point clouds to be aligned like variation in point density, presence of symmetrical structures, presence of holes etc. The scale at which the algorithms execute also depends upon these factors.

A benchmarking study of the combination of various 3D keypoints and feature extractors is suggested as an extension to the current study to select the best keypoint-

feature combination for feature matching. The use of multiple features which extract complementary information from the point clouds is expected to enhance the feature matching process and may be considered as a future enhancement for point cloud registration.

The online registration pipelines may not include a coarse registration and employ only a fine registration stage in some cases. This requires capturing of a large number of closely spaced scans of the scene to obtain a solution. However, this leads to huge computational and memory requirements. In addition, it stands a chance of entering a local minimum. A coarse registration stage in the pipeline eliminates the requirement of huge number of closely spaced scans as it is able to find a global minimum. Also as the coarse registration works on a small subset of points called keypoints, instead of the original point clouds which may contain a large number of points, the memory and computational requirements are minimised. In addition to this, as the fine registration tries to refine an already registered point cloud, the convergence is achieved at a faster rate. Hence a coarse - to - fine registration strategy is ideal in terms of memory and computational requirements.

The significance of the proposed robust algorithm lies in the fact that the corresponding point pairs estimated by feature matching are largely corrupted by outliers or wrong matches. In order to filter out the wrong matches, the existing standard approaches like RANSAC test many randomly generated transformation matrices to find the one with maximum support. The proposed approach presented a computationally efficient algorithm that finds a more accurate solution. The proposed algorithm is effective in finding a coarse alignment, which is then refined using a standard fine alignment algorithm. It outperforms the standard RANSAC in terms of computational time as well as accuracy. This in turn aids in making the fine alignment stage converge faster.

In the case of large percentage of outliers, the proposed algorithm performs better than RANSAC in terms of accuracy due to the local refinement of the solution. Since more promising match pairs are examined initially by guided sampling, a good sample is expected to be found early in the loop compared to a random search among the wrong matches. This accounts for the computational advantage of the proposed approach. In the case of high percentage of correct matches, RANSAC is also ex-

pected to find a good sample early even with a random search. In such cases, the computational advantage due to guided sampling may not be evident and may be even overshadowed by the time spent in local refinement. If a good accuracy is attained, then the local refinement stage may be avoided in such cases.

In the case of fine alignment, the standard ICP algorithm, along with a correspondence refinement stage is used in this study. Application of more advanced methods of fine alignment including the various methods detailed in section 2.1.2 may be included in the pairwise registration pipeline in future studies.

Thus the results obtained by the thesis indicate that the development of a robust estimation algorithm can improve the pairwise alignment and thus the automatic registration framework.

7.2 Multiview registration of 3D point clouds

Once the overlapping pairs of scans are registered using pairwise registration, the transformation from all of the point clouds to a common coordinate system is to be estimated in order to make a single 3D point cloud. Chapter 4 addressed this problem of finding the global transformations. Since sequentially multiplying the transformation matrices from the pairwise registration results leads to accumulation of errors, a multiview registration stage is often necessary to mitigate the alignment errors.

This study has considered the case where there is redundant overlap among the sequence of scans, i.e. a scan overlaps with scans other than its immediate neighbours. In such case, all the overlapping pairs of scans are registered using pairwise registration and these redundant transformations are used as additional constraints in estimating the global transformations. This is especially beneficial in the case of registering the last scan with the first in a cycle of scans, which helps to account for the loop closing constraint.

Some of the scan pairs considered for pairwise registration may not have sufficient overlap and may lead to incorrect estimation of relative motions (transformations). This set of relative motions can be averaged by a least squares algorithm to find the global motions. If the relative motions are incorrect, this corrupts the results of the

global motions as L2 averaging cannot handle outliers. Chapter 4 presented an approach to identify the outliers or incorrect transformations from the set of relative motions. The global motions are initialised using a sequential approach. Then the outliers from the set of relative motions are identified based on their agreement with the global motions and the set of other relative motions with common scan pairs. The agreement between the transformations is defined utilising the Lie group structure of the rigid body transformations. The distance between transformation matrices is defined by the first order approximation of the Riemannian distance on the Lie algebra space corresponding to their Lie groups. This distance metric is employed to quantify the measure of agreement of a relative transformation matrix with the global and other relative motion estimates. The measure of agreement endows the relative motion estimates with a weight, which is then utilised in a graph based approach to find the inliers in the set of relative motions. Once the correct relative motions are identified, they are averaged to calculate the global transformations.

The multiview approach considered in this study mainly included registration of a set of scans of an object captured in a sequence. This can be extended to the case where the views are captured in a random order. In the present system, the set of pairs of scans which are expected to overlap are provided as input and the system automatically performs the pairwise registration. The set of neighbours (extended neighbours) of a scan in the scan sequence are considered while forming the pairs. The initialisation of global motions is also performed by making use of the sequence of scans. In the case of a random acquisition, the overlapping pairs of scans should be identified first before pairwise registration. Including all possible pairs of scans for pairwise registration may lead to combinatorial expansion of the problem space and affects computational time. Hence the research queries considered in this thesis may be further extended by carrying out the following investigations. The problem of identifying the set of overlapping pairs from the input set of scans can be considered as a future study. Initialisation of global motions reliably in such a case can also be considered as an extension to the present study.

The robust motion averaging algorithm developed in this thesis thus improves the multiview registration stage. A successful multiview alignment is shown to be essential for object reconstruction from 3D scans.

7.3 Automatic framework for 3D point cloud registration

Chapter 5 presented the complete registration pipeline for obtaining a single aligned point cloud from a set of partially overlapping 3D point clouds. The cases of object reconstruction as well as scene reconstruction were considered.

The 3D registration pipeline performance was evaluated using the 3D scans captured using different modalities of 3D data acquisition. Object reconstruction using laser scanned data of a set of objects, obtained from two different datasets - Stanford and University of Western Australia, and from an RGB-D dataset was investigated. The case of object reconstruction using 3D data of objects captured using the low-cost Microsoft Kinect sensor was also considered. The data captured by Kinect are often corrupted by noise and holes. The performance of the registration pipeline using low-quality 3D scans was investigated.

Scene reconstruction using data captured by Kinect sensor was performed. Another sensor considered in this study was the terrestrial laser scanner (TLS). Unlike the previous cases, the considered TLS provides a multi-angle view in a single scan. Multiple scans were acquired to account for the presence of holes due to occlusion. The proposed registration pipeline could successfully register the considered cases of object and scene reconstruction.

The laser scanned object databases possess only the coordinate information. The case where the point clouds possess colour information in addition to the coordinate information was also considered. The features which make use of colour data in addition to the geometry information were employed for feature matching. The change in illumination variation can affect the feature match performance and future studies may investigate the influence of illumination variation in such instances. In case of a long sequence of scans, the accumulated error may affect the initialisation of global motions, which may in turn affect the inlier calculation and motion averaging. An investigation into the performance of multiview registration in the case of a long sequence of views may be attempted as a future study.

The investigations carried out in this thesis is based on the assumption of rigid body

transformation between the local coordinate system (camera coordinate system) of the scan views. The current study does not deal with the case of non-rigid body motions or warps in the scene. The study on non-rigid body registration is gaining popularity in many fields like human motion tracking. The proposed robust estimation algorithm may be considered in the case of estimation of a higher order motion matrix.

The development of a complete automatic framework for registration in the case of object and scene reconstruction is achieved by the thesis.

7.4 Automatic registration of remote sensing satellite images

Chapter 6 presented an application where the robust estimation algorithm for pairwise registration was adapted and implemented. The case of affine transformation estimation between two remote sensing satellite images varying in spatial, spectral and temporal variations was considered for study. The main stages involved in satellite image registration are the feature extraction and matching, transformation estimation, and resampling. The proposed method was adapted in the transformation estimation phase where the parameters of the affine transformation were estimated using the corresponding point pairs obtained by matching 2D features.

The adaptation of the proposed robust estimation method to the problem of image registration highlights the generalisation ability of the algorithm. The algorithm can be adapted to other robust estimation tasks where a measure of the quality of the data points (here the corresponding point pair distance) can be specified. The adaptation of the proposed method to other application areas may be carried out in a future investigation.

7.5 Contributions of the thesis

The major contributions of this thesis work are :

- *An automatic 3D registration framework for 3D object reconstruction from mul-*

tiple partial 3D scans is developed.

- *A novel robust algorithm for coarse registration is developed. It improves the pairwise registration stage by achieving fast convergence and improved accuracy. It is shown to be superior to existing approaches.*
- *A novel robust multiview alignment is developed. It helps to estimate the global transformations reliably even in presence of outlier relative transformations. It exhibits better performance with less memory and computational requirements compared to existing approaches.*
- *The developed pairwise registration algorithm is highly adaptive and general compared to other approaches. The algorithm is adapted to the automatic registration of remote sensing satellite images. This highlights the generalization capability of the algorithm.*
- *The proposed registration framework is easily generalizable. It is successfully applied in the case of scene reconstruction from 3D scans captured with Kinect sensor and a terrestrial laser scanner.*
- *The entire registration framework is developed in an open source C++ platform thus making it feasible for efficient sharing.*

7.6 Summary and Future work

This thesis has addressed the important research requirement of developing an automatic complete 3D registration framework. As part of this, novel robust and efficient algorithms are developed to improve the fundamental building blocks of registration. A detailed survey on the basic problems to be addressed in the registration pipeline and the current solutions is presented. Novel robust algorithms are developed and implemented to address pairwise as well as multiview registration problems. A complete framework for automatically generating a single point cloud from a sequence of partially overlapping 3D scans is implemented and evaluated using scan data from different datasets. Scans acquired using Microsoft Kinect sensor and terrestrial laser scanner (TLS) are also considered for assessing the generalizability of the methodology in scene modelling.

Future Directions:

The present work can be extended to improve the registration pipeline by studying the effects of the amount of overlap, sensor noise, variation in point density and peculiarities of object surface. Benchmarking of the registration algorithms can be undertaken by considering scans acquired using strictly controlled environments, known calibration targets or synthetic scans. Future studies could evaluate how well an approach scales with the size of the target. The challenges in the case of real world scenes and huge cultural artefacts may be addressed. The effect of the various keypoints and features on the performance of pairwise registration may be evaluated. In the case of multiview registration, the automatic detection and quantification of the amount of overlap between scans in a set is one area for consideration. The correctness of pairwise registration does not necessarily reflect a correct multiview alignment and benchmarking of multiview registration algorithms also needs to be done. The distance between the reconstructed model and the actual model is one way of examining the correctness. The case of non-rigid body motions like moving and deforming shapes is a challenging research area which can be undertaken as a future study.

Another area of future improvements is the adaptation of parallel algorithms in this work. The adaptation of CPU based or GPU based parallel processing for pairwise alignment may be considered for adaptation in the pipeline.

In conclusion, I hope that the ideas discussed in this thesis add value to the current knowledge and inspire future work.

7.7 Acknowledgements for data

I would like to acknowledge Stanford University and the University of Western Australia for providing the 3D scan dataset freely online. I would also like to acknowledge the satellite image data provided by various space agencies - CNES, France for the Pleaides-1A satellite imagery, DLR, Germany for the airborne AISA hyperspectral imagery and NRSC-Hyderabad for AWiFS data from Bhuvan platform and USGS for Landsat-8 data.

REFERENCES

1. Aiger, D., Mitra, N. J., and Cohen-Or, D. (2008). 4points congruent sets for robust pairwise surface registration. In *ACM SIGGRAPH 2008 Papers* SIGGRAPH'08 (pp. 85:1–85:10). New York, NY, USA: ACM.

2. Akagunduz, E., and Ulusoy, I. (2007). 3d object representation using transform and scale invariant 3d features. In *2007 IEEE 11th International Conference on Computer Vision* (pp. 1–8). Rio de Janeiro, Brazil.

3. Albarelli, A., Rodol, E., Bergamasco, F., and Torsello, A. (2011). A non-cooperative game for 3d object recognition in cluttered scenes. In *Proceedings - 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2011* (pp. 252–259). Hangzhou, China.

4. Aldoma, A., Tombari, F., Rusu, R. B., and Vincze, M. (2012). Our-cvfh - oriented, unique and repeatable clustered viewpoint feature histogram for object recognition and 6dof pose estimation. In A. Pinz, T. Pock, H. Bischof, and F. Leberl (Eds.), *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 113–122). Berlin, Heidelberg: Springer Berlin Heidelberg volume 7476 LNCS.

5. Aldoma, A., Vincze, M., Blodow, N., Gossow, D., Gedikli, S., Rusu, R. B., and Bradski, G. (2011). Cad-model recognition and 6dof pose estimation using 3d cues. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (pp. 585–592). Barcelona, Spain.

6. Alexandre, L. (2012). 3d descriptors for object and category recognition: a comparative evaluation. In *Workshop on Color-Depth Camera Fusion in Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* 3 (p. 7). Vilamoura, Portugal.

7. Andersen, M. R., Jensen, T., Lisouski, P., Mortensen, A. K., Hansen, M. K., Gregersen,

T., and Ahrendt, P. (2015). Kinect depth sensor evaluation for computer vision applications. *Technical Report Electronics and Computer Engineering*, *1*.

8. Arun, K. S., Huang, T. S., and Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. *IEEE Trans. Pattern Anal. Mach. Intell. ()*, *9*, 698–700.

9. Assfalg, J., Bertini, M., Bimbo, a. D., and Pala, P. (2007). Content-based retrieval of 3-d objects using spin image signatures. *IEEE Transactions on Multimedia*, *9*, 589–599.

10. Bagchi, P., Bhattacharjee, D., and Nasipuri, M. (2016). Registration of range images using a novel technique of centroid alignment. In R. Chaki, A. Cortesi, K. Saeed, and N. Chaki (Eds.), *Advanced Computing and Systems for Security. Springer India* (pp. 81–89). Springer India.

11. Bariya, P., Novatnack, J., Schwartz, G., and Nishino, K. (2012). 3d geometric scale variability in range images: features and descriptors. *International Journal of Computer Vision*, *99*, 232–255.

12. Batlle, E., Matabosch, C., and Salvi, J. (2007). Summarizing image/surface registration for 6dof robot/camera pose estimation. *Pattern Recognition and Image Analysis, Pt 2, Proceedings*, *4478*, 105–112.

13. Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In A. Leonardis, H. Bischof, and A. Pinz (Eds.), *Computer Vision – ECCV 2006: 9th European Conference on Computer Vision, Proceedings, Part I* (pp. 404–417). Graz, Austria.: Springer Berlin Heidelberg.

14. Bayramoglu, N., and Alatan, A. A. (2010). Shape index sift: Range image recognition using local features. In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 352–355). Istanbul, Turkey.

15. Benjemaa, R., and Schmitt, F. (1998). A solution for the registration of multiple 3d point sets using unit quaternions. In H. Burkhardt, and B. Neumann (Eds.), *Computer Vision — ECCV'98: 5th European Conference on Computer Vision, June 2–6, 1998 Proceedings, Volume II* (pp. 34–50). Freiburg, Germany.: Springer Berlin Heidelberg.

16. Benjemaa, R., and Schmitt, F. (1999). Fast global registration of 3d sampled surfaces using a multi-z-buffer technique. *Image and Vision Computing*, *17*, 113–123.

17. Bergevin, R., Laurendeau, D., and Poussart, D. (1992). Estimating the 3d rigid transformation between two range views of a complex object. In *Pattern Recognition, 1992. Vol.I. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on* (pp. 478–482). The Hague, Netherlands.

18. Bergevin, R., Soucy, M., Gagnon, H., and Laurendeau, D. (1996). Towards a general multi-view registration technique. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*, 540–547.

19. Berkmann, J., and Caelli, T. (1994). Computation of surface geometry and segmentation using covariance techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*, 1114–1116.

20. Besl, P. J. (1988). Geometric modeling and computer vision. *Proceedings of the IEEE*, *76*, 936–958.

21. Besl, P. J., McKay, N. D. et al. (1992). A method for registration of 3-d shapes. *IEEE Transactions on pattern analysis and machine intelligence*, *14*, 239–256.

22. Biber, P., and Strasser, W. (2003). The normal distributions transform: a new approach to laser scan matching. In *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on* (pp. 2743–2748). Las Vegas, USA. volume 3.

23. Blais, G., and Levine, M. D. (1995). Registering multiview range data to create 3d computer objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*, 820–824.

24. Blanco, J. L. (2010). *A tutorial on se (3) transformation parameterizations and on-manifold optimization*. University of Malaga, Tech. Rep.

25. Borrmann, D., Elseberg, J., Lingemann, K., Nüchter, A., and Hertzberg, J. (2008a). Globally consistent 3d mapping with scan matching. *Robotics and Autonomous Systems*, *56*, 130–142.

26. Borrmann, D., Elseberg, J., Lingermann, K., Nuchter, A., and Hertzberg, J. (2008b). The efficient extension of globally consistent scan matching to 6 dof. *Knowledge Based Systems*, *1*, 20.

27. Bouaziz, S., Tagliasacchi, A., and Pauly, M. (2013). Sparse iterative closest point. *Computer Graphics Forum*, *32*, 113–123.

28. Boyer, E., Bronstein, A. M., Bronstein, M. M., Bustos, B., Darom, T., Horaud, R., Hotz, I., Keller, Y., Keustermans, J., Kovnatsky, A., Litman, R., Reininghaus, J., Sipiran, I., Smeets, D., Suetens, P., Vandermeulen, D., Zaharescu, A., and Zobel, V. (2011). {SHREC} 2011: robust feature detection and description benchmark. In *Proc. Eurographics Workshop 3D Object Retrieval (3DOR),CoRR*. Llandudno, UK. volume abs/1102.4.

29. Breuckmann, B. (2014). 25 years of high definition 3d scanning: History, state of the art, outlook. In *Proceedings of the EVA London 2014 on Electronic Visualisation and the Arts* (pp. 262–266). London, UK.: BCS.

30. Bronstein, A. M., Bronstein, M. M., Bustos, B., Castellani, U., Crisani, M., Falcidieno, B., Guibas, L. J., Kokkinos, I., Murino, V., Ovsjanikov, M., and Others (2010). Shrec 2010: robust feature detection and description benchmark. In *Proc. Eurographics Workshop 3D Object Retrieval,3DOR* (p. 6). Norrkoping, Sweden. volume 2.

31. Buch, A. G., Kraft, D., Kamarainen, J. K., Petersen, H. G., and Krüger, N. (2013). Pose estimation using local structure-specific shape and appearance context. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on* (pp. 2080–2087). Karlsruhe, Germany.

32. Caetano, T. S., and Caelli, T. (2006). Approximating the problem, not the solution: An alternative view of point set matching. *Pattern Recognition*, *39*, 552–561.

33. Caetano, T. S., Caelli, T., and Barone, D. a. C. (2004). An optimal probabilistic graphical model for point set matching. In A. Fred, T. M. Caelli, R. P. W. Duin, A. C. Campilho, and D. de Ridder (Eds.), *Structural, Syntactic, and Statistical Pattern Recognition: Joint IAPR International Workshops, SSPR 2004 and SPR 2004, Proceedings* (pp. 162–170). Lisbon, Portugal: Springer Berlin Heidelberg.

34. Campbell, D., and Petersson, L. (2015). An adaptive data representation for robust point-set registration and merging. In *The IEEE International Conference on Computer Vision (ICCV)*. Santiago, Chile.

35. Campbell, R. J., and Flynn, P. J. (2001). A survey of free-form object representation and recognition techniques. *Computer Vision and Image Understanding*, *81*, 166–210.

36. Carmichael, O., Huber, D., and Hebert, M. (1999). Large data sets and confusing scenes in 3-d surface matching and recognition. In *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on* (pp. 358–367). Ottawa, Ontario, Canada.

37. Castellani, U., and Bartoli, A. (2012). 3d shape registration. In N. Pears, Y. Liu, and P. Bunting (Eds.), *3D Imaging, Analysis and Applications* (pp. 221–264). London: Springer London.

38. Castellani, U., Cristani, M., Fantoni, S., and Murino, V. (2008). Sparse points matching by combining 3d mesh saliency with statistical descriptors. *Computer Graphics Forum*, *27*, 643–652.

39. Chatterjee, A., Govindu, V. M., and Madhav Govindu, V. (2013). Efficient and robust large-scale rotation averaging. In *The IEEE International Conference on Computer Vision (ICCV)* (pp. 521–528). Sydney, Autstralia.

40. Chen, H., and Bhanu, B. (2007). 3d free-form object recognition in range images using local surface patches. *Pattern Recognition Letters*, *28*, 1252–1262.

41. Chen, Y., and Medioni, G. (1992). Object modelling by registration of multiple range images. *Image and Vision Computing*, *10*, 145–155.

42. Chetverikov, D., Svirko, D., Stepanov, D., and Krsek, P. (2002). The trimmed iterative closest point algorithm. In *In Pattern Recognition, 2002. Proceedings. 16th International Conference on* (pp. 545–548). Quebec City, Canada. volume 3.

43. Christensen, G. E., Geng, X., Kuhl, J. G., Bruss, J., Grabowski, T. J., Pirwani, I. A., Vannier, M. W., Allen, J. S., and Damasio, H. (2006). Introduction to the non-rigid image registration evaluation project (nirep). In J. P. W. Pluim, B. Likar, and F. A. Gerritsen (Eds.), *Biomedical Image Registration: Third International Workshop, WBIR 2006, Utrecht, The Netherlands, July 9-11, 2006. Proceedings* (pp. 128–135). Berlin, Heidelberg: Springer Berlin Heidelberg.

44. Chua, C. S., and Jarvis, R. (1996). 3d free-form surface registration and object recognition. *International Journal of Computer Vision*, *17*, 77–99.

45. Chua, C. S., and Jarvis, R. (1997). Point signatures: A new representation for 3d object recognition. *International Journal of Computer Vision*, *25*, 63–85.

46. Chum, O., and Matas, J. (2005). Matching with prosac - progressive sample consensus. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (pp. 220–226 vol.1). San Diego, CA, USA.

47. Chum, O., Matas, J., and Kittler, J. (2003). Locally optimized ransac. In B. Michaelis, and G. Krell (Eds.), *Pattern Recognition: 25th DAGM Symposium, Proceedings* (pp. 236–243). Magdeburg, Germany.: Springer Berlin Heidelberg.

48. Cognex (2016). Cognex sensors. `http://www.cognex.com/products/machine-vision/ds-1000-displacement-sensor-laser-profiler/`.

49. Creusot, C., Pears, N., and Austin, J. (2013). A machine-learning approach to keypoint detection and landmarking on 3d meshes. *International Journal of Computer Vision*, *102*, 146–179.

50. Cunnington, S. J., and Stoddart, A. J. (1999). N-view point set registration: A comparison. In *BMVC* (pp. 1–11). Notingham, U.K.

51. Curless, B., and Levoy, M. (1996). A volumetric method for building complex models from range images. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques* SIGGRAPH '96 (pp. 303–312). New York, NY, USA.: ACM.

52. Dalley, G., and Flynn, P. (2002). Pair-wise range image registration: A study in outlier classification. *Computer Vision and Image Understanding*, *87*, 104–115.

53. Darom, T., and Keller, Y. (2012). Scale-invariant features for 3-d mesh models. *IEEE Transactions on Image Processing*, *21*, 2758–2769.

54. Das, A., and Waslander, S. L. (2012). Scan registration with multi-scale k-means normal distributions transform. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2705–2710). Vilamoura-Algarve, Portugal.

55. Das, A., and Waslander, S. L. (2014). Scan registration using segmented region growing ndt. *The International Journal of Robotics Research*, *33*, 1645–1663.

56. Davis, T. A. (2006). *Direct methods for sparse linear systems* volume 2. Society for Industrial and Applied Mathematics.

57. Dey, T. K., Li, G., and Sun, J. (2005). Normal estimation for point clouds: a comparison study for a voronoi based method. In *Proceedings Eurographics/ IEEE VGTC Symposium Point-Based Graphics* (pp. 39–46). Stony Brook, NY, USA.

58. Diebel, J. (2006). Representing attitude: Euler angles, unit quaternions, and rotation vectors. *Matrix*, *58*, 1–35.

59. Dinh, H. Q., and Kropac, S. (2006). Multi-resolution spin-images. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (pp. 863–870). New York, NY, USA. volume 1.

60. Do Nascimento, E. R., Oliveira, G. L., Vieira, A. W., and Campos, M. F. M. (2013). On the development of a robust, fast and lightweight keypoint descriptor. *Neurocomputing*, *120*, 141–155.

61. Domey, J., and Picard, M. (2002). Active optical 3d imaging for heritage applications. *IEEE Computer Graphics Applications*, *22*, 24–36.

62. Dorai, C., Wang, G., Jain, A. K., and Mercer, C. (1996). From images to models: automatic 3d object model construction from multiple views. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on* (pp. 770–774). Vienna, Austria. volume 1.

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

64. Drost, B., Ulrich, M., Navab, N., and Ilic, S. (2010). Model globally, match locally: Efficient and robust 3d object recognition. In *IEEE Conference on Computer Vision and Pattern Recognition* (p. 5). San Francisco, CA. volume 1.

65. Druon, S., Aldon, M. J., and Crosnier, A. (2006). Color constrained icp for registration of large unstructured 3d color data sets. In *2006 IEEE International Conference on Information Acquisition* (pp. 249–255). Weihai, China.

66. Du, S., Zheng, N., Ying, S., and Liu, J. (2010). Affine iterative closest point algorithm for point set registration. *Pattern Recognition Letters*, *31*, 791–799.

67. Duan, X., Sun, H., and Peng, L. (2013). Riemannian means on special euclidean group and unipotent matrices group. *The Scientific World Journal*, *2013*.

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

69. Ebrahim, M., and Mostafa, D. (2011). 3d laser scanners: History, applications, and future. Civil Engineering Department, Faculty of Engineering, Assiut University.

70. Eggert, D. W., Fitzgibbon, A. W., and Fisher, R. B. (1996). Simultaneous registration of multiple range views for use in reverse engineering. In *Proceedings - International Conference on Pattern Recognition* (pp. 243–247). Vienna, Austria. volume 1.

71. El-Hakim, S. F., Beraldin, J. A., and Blais, F. (1995). Comparative evaluation of the performance of passive and active 3d vision systems. In *SPIE Proc. 2646 St. Petersburg Conf on Digital Photogrammetry* (pp. 14–25). Orlando, FL, USA. volume 2646.

72. Evangelidis, G. D., Kounades-Bastian, D., Horaud, R., and Psarakis, E. Z. (2014). A generative model for the joint registration of multiple point sets. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014: 13th European Conference Proceedings, Part VII* (pp. 109–122). Zurich, Switzerland.: Springer International Publishing.

73. Fan, B., Huo, C., Pan, C., and Kong, Q. (2013). Registration of optical and sar satellite images by exploring the spatial relationship of the improved sift. *IEEE Geoscience and Remote Sensing Letters*, *10*, 657–661.

74. FARO (2016). Faro 3d sensors. `http://www.faro.com/products/3d-documentation`.

75. Faugeras, O., and Hebert, M. (1986). The representation, recognition, and locating of 3-d objects. *The International Journal of Robotics Research*, *5*, 27–52.

76. Favaro, P., and Soatto, S. (2002). Learning shape from defocus. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen (Eds.), *Computer Vision — ECCV 2002: 7th European*

Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part II (pp. 735–745). Copenhagen, Denmark.: Springer Berlin Heidelberg.

77. Filipe, S., and Alexandre, L. A. (2014). A comparative evaluation of 3d keypoint detectors in a rgb-d object dataset. In *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on* (pp. 476–483). Lisbon, Portugal. volume 1.

78. Fischler, M. A., and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, *24*, 381–395.

79. Fletcher, P. T., Lu, C., and Joshi, S. (2003). Statistics of shape via principal geodesic analysis on lie groups. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (pp. 95–101). Madison, Wisconsin. volume 1.

80. Flint, A., Dick, A. R., and Van Den Hengel, A. (2007). Thrift: Local 3d structure recognition. In *Proceedings - Digital Image Computing Techniques and Applications: 9th Biennial Conference of the Australian Pattern Recognition Society, DICTA 2007* (pp. 182–188). Glenelg, Australia. volume 7.

81. Friedman, J. H., Bentley, J. L., and Finkel, R. A. (1977). An algorithm for finding best matches in logarithmic expected time. *ACM Trans. Math. Softw.*, *3*, 209–226.

82. Frome, A., Huber, D., Kolluri, R., Bülow, T., and Malik, J. (2004). Recognizing objects in range data using regional point descriptors. In T. Pajdla, and J. Matas (Eds.), *Computer Vision - ECCV 2004: 8th European Conference on Computer Vision, May 11-14, 2004. Proceedings, Part III* (pp. 224–237). Prague, Czech Republic.: Springer Berlin Heidelberg volume 3023.

83. Gagnon, H., Bergevin, R., Laurendeau, D., Soucy, M., Bergevin, R., and Laurendeau, D. (1994). Registration of multiple range views for automatic 3-d model building. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94* (pp. 581–586). Seattle, USA.

84. Gal, R., and Cohen-Or, D. (2006). Salient geometric features for partial shape matching and similarity. *ACM Transactions on Graphics*, *25*, 130–150.

85. Gelfand, N., Ikemoto, L., Rusinkiewicz, S., and Levoy, M. (2003). Geometrically stable sampling for the icp algorithm. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on* (pp. 260–267). Banff, Alta.

86. Gelfand, N., Mitra, N. J., Guibas, L. J., and Pottmann, H. (2005). Robust global registration. In *Symposium on geometry processing* (p. 5). Vienna, Austria. volume 2.

87. Geng, J. (2011). Structured-light 3d surface imaging: a tutorial. *Adv. Opt. Photon.*, *3*, 128–160.

88. Girardeau-Montaut, D. (2015). Cloud compare - 3d point cloud and mesh processing software. Open Source Project.

89. Głomb, P. (2009). Detection of interest points on 3d data: Extending the harris operator. In M. Kurzynski, and M. Wozniak (Eds.), *Computer Recognition Systems 3* (pp. 103–111). Springer Berlin Heidelberg.

90. Godin, G., Rioux, M., and Baribeau, R. (1994). Three-dimensional registration using range and intensity information. In *Proceedings of SPIE videometric III* (pp. 279–290). Boston, Massachusetts. volume 2350.

91. Gold, S., and Rangarajan, A. (1996). A graduated assignment algorithm for graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*, 377–388.

92. Gomes, L., Regina Pereira Bellon, O., and Silva, L. (2014). 3d reconstruction methods for digital preservation of cultural heritage: A survey. *Pattern Recognition Letters*, *50*, 3–14.

93. Goncalves, H., Corte-Real, L., and Goncalves, J. A. (2011). Automatic image registration through image segmentation and sift. *IEEE Transactions on Geoscience and Remote Sensing*, *49*, 2589–2600.

94. Gonzalez, R., and Woods, R. (1992). *Digital Image Processing*. Addison-Wesley world student series. Addison-Wesley.

95. Govindu, V. M. (2004). Lie-algebraic averaging for globally consistent motion estimation. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of*

the 2004 IEEE Computer Society Conference on (pp. I–684–I–691). Washington DC, USA. volume 1.

96. Govindu, V. M. (2006). Robustness in motion averaging. In P. J. Narayanan, S. K. Nayar, and H.-Y. Shum (Eds.), *Computer Vision – ACCV 2006: 7th Asian Conference on Computer Vision, Proceedings, Part II* (pp. 457–466). Hyderabad, India.: Springer Berlin Heidelberg volume 3852 LNCS.

97. Govindu, V. M., and Pooja, A. (2014). On averaging multiview relations for 3d scan registration. *IEEE Transactions on Image Processing*, *23*, 1289–1302.

98. Guehring, J. (2000). Dense 3d surface acquisition by structured light using off-the-shelf components. In *Proc. of Videometrics and Optical Methods for 3D Shape Measurings (SPIE)* (pp. 220–231). San Jose, [California] USA. volume 4309.

99. Guo, Y., Bennamoun, M., Sohel, F., Lu, M., Wan, J., and Kwok, N. M. (2016). A comprehensive performance evaluation of 3d local feature descriptors. *International Journal of Computer Vision*, *116*, 66–89.

100. Guo, Y., Bennamoun, M., Sohel, F., Min Lu, and Jianwei Wan (2014). 3d object recognition in cluttered scenes with local surface features: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *36*, 2270–2287.

101. Guo, Y., Sohel, F., Bennamoun, M., Lu, M., and Wan, J. (2013a). Rotational projection statistics for 3d local surface description and object recognition. *International Journal of Computer Vision*, *105*, 63–86.

102. Guo, Y., Sohel, F. A., Bennamoun, M., Lu, M., and Wan, J. (2013b). Trisi: A distinctive local surface descriptor for 3d modeling and object recognition. In *GRAPP/IVAPP* (pp. 86–93). Barcelona, Spain.

103. Han, J., Wang, F., Guo, Y., Zhang, C., and He, Y. (2015). An improved ransac registration algorithm based on region covariance descriptor. In *Chinese Automation Congress (CAC), 2015* (pp. 746–751). Hubei Province, China.

104. Han, J., Yin, P., He, Y., and Gu, F. (2016). Enhanced icp for the registration of large-scale 3d environment models: An experimental study. *Sensors*, *16*, 228.

105. Hansard, M., Lee, S., Choi, O., and Horaud, R. P. (2012). *Time-of-flight cameras: principles, methods and applications*. Springer Science & Business Media.

106. Harris, C., and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference* (p. 50). Manchester, England. volume 15.

107. Hébert, P., Tubic, D., Gignac, O., Saint-Pierre, É., Rochette, F., and Carette, A. (2016). Sensor positioning for 3d scanning. US Patent 9,325,974.

108. Hersch, M., Billard, A., and Bergmann, S. (2012). Iterative estimation of rigid-body transformations. *Journal of Mathematical Imaging and Vision*, *43*, 1–9.

109. Ho, H. T., and Gibbins, D. (2008). Multi-scale feature extraction for 3d surface registration using local shape variation. In *2008 23rd International Conference Image and Vision Computing* (pp. 1–6). Christchurch, New Zealand.

110. Ho, H. T., and Gibbins, D. (2009). Curvature-based approach for multi-scale feature extraction from 3d meshes and unstructured point clouds. *IET Computer Vision*, *3*, 201–212.

111. Holzer, S., Shotton, J., and Kohli, P. (2012). Learning to efficiently detect repeatable interest points in depth data. In A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid (Eds.), *Computer Vision – ECCV 2012: 12th European Conference on Computer Vision, Proceedings, Part I* (pp. 200–213). Florence, Italy.: Springer Berlin Heidelberg.

112. Hong, H., and Lee, B. (2016). Colored point cloud registration with improved hue-assisted normal distributions transform. *International Journal of Computer Theory and Engineering*, *8*, 63.

113. Horn, B. K. P. (1987). Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, *4*, 629–642.

114. Hu, J., and Hua, J. (2009). Salient spectral geometric features for shape matching and retrieval. *The Visual Computer*, *25*, 667–675.

115. Hu, T., Zhang, H., Shen, H., and Zhang, L. (2014). Robust registration by rank minimization for multiangle hyper/multispectral remotely sensed imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*, 2443–2457.

116. Hua, J., Lai, Z., Dong, M., Gu, X., and Qin, H. (2008). Geodesic distance-weighted shape vector image diffusion. *IEEE Transactions on Visualization and Computer Graphics*, *14*, 1643–1650.

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

118. Hurley, J. R., and Cattell, R. B. (1962). The procrustes program: Producing direct rotation to test a hypothesized factor structure. *Behavioral Science*, *7*, 258–262.

119. Hussmann, S., Gonschior, M., Buttgen, B., Peter, C., Schwope, S., Perwass, C., and Hallstig, J. (2013). A review on commercial solid state 3d cameras for machine vision applications. *Recent Advances in Topography Research*, *1*, 303–351.

120. iDS (2016). ids ensenso. `https://en.ids-imaging.com/ensenso-stereo-3d-camera.html`.

121. Jian, B., and Vemuri, B. C. (2011). Robust point set registration using gaussian mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*, 1633–1645.

122. Johnson, A. E., and Hebert, M. (1997). Recognizing objects by matching oriented points. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on* (pp. 684–689). San Juan, Puerto Rico, USA.

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

124. Jolliffe, I. (2014). Principal component analysis. In *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd.

125. Joung, J. H., An, K. H., Kang, J. W., Chung, M. J., and Yu, W. (2009). 3d environment reconstruction using modified color icp algorithm by fusion of a camera and a 3d laser range finder. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3082–3088). St. Louis, MO, USA.

126. Joyner, D., Van Nguyen, M., and Cohen, N. (2010). *Algorithmic graph theory*. Free Software Foundation.

127. Jun, L., Wei, L., Donglai, D., and Qiang, S. (2015). Point cloud registration algorithm based on ndt with variable size voxel. In *Control Conference (CCC), 2015 34th Chinese* (pp. 3707–3712). Hangzhou, China.

128. van Kaick, O., Zhang, H., Hamarneh, G., and Cohen-Or, D. (2011). A survey on shape correspondence. *Computer Graphics Forum*, *30*, 1681–1707.

129. Kanatani, K. (1994). Analysis of 3-d rotation fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*, 543–549.

130. Kanatani, K. (2012). *Group-theoretical methods in image understanding* volume 20. Springer Science & Business Media.

131. Khoshelham, K. (2016). Closed-form solutions for estimating a rigid motion from plane correspondences extracted from point clouds. *{ISPRS} Journal of Photogrammetry and Remote Sensing*, *114*, 78–91.

132. Kim, H., and Hilton, A. (2013). Evaluation of 3d feature descriptors for multi-modal data registration. In *2013 International Conference on 3D Vision - 3DV 2013* (pp. 119–126). Tokyo, Japan.

133. Kim, J. W., and Lee, B. H. (2016). Robust and fast 3-d scan registration using normal distributions transform with supervoxel segmentation. *Robotica*, *34*, 1630–1658.

134. Kim, T., and Im, Y.-J. (2003). Automatic satellite image registration by combination of matching and random sample consensus. *IEEE Transactions on Geoscience and Remote Sensing*, *41*, 1111–1117.

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

136. Knopp, J., Prasad, M., and Van Gool, L. (2010a). Orientation invariant 3d object classification using hough transform based methods. In *Proceedings of the ACM Workshop on 3D Object Retrieval* 3DOR '10 (pp. 15–20). New York, NY, USA: ACM.

137. Knopp, J., Prasad, M., Willems, G., Timofte, R., and Van Gool, L. (2010b). Hough transform and 3d surf for robust three dimensional classification. In K. Daniilidis, P. Maragos, and N. Paragios (Eds.), *Computer Vision – ECCV 2010: 11th European*

Conference on Computer Vision, Proceedings, Part VI (pp. 589–602). Heraklion, Crete, Greece.: Springer Berlin Heidelberg volume 6316 LNCS.

138. Krishnan, S., Lee, P. Y., Moore, J. B., Venkatasubramanian, S., and Others (2005). Global registration of multiple 3d point sets via optimization-on-a-manifold. In *Symposium on Geometry Processing* (pp. 187–196). Vienna, Austria.

139. Kümmerle, R., Grisetti, G., Strasdat, H., Konolige, K., and Burgard, W. (2011). G2o: A general framework for graph optimization. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on* (pp. 3607–3613). Shanghai, China.

140. Kytö, M., Nuutinen, M., and Oittinen, P. (2011). Method for measuring stereo camera depth accuracy based on stereoscopic vision. In *Three-Dimensional Imaging, Interaction, and Measurement* (pp. 78640I–78640I–9). California, USA. volume 7864.

141. Lamdan, Y., and Wolfson, H. J. (1988). Geometric hashing: A general and efficient model-based recognition scheme. New York University, Department of Computer Science, Courant Institute of Mathematical Sciences.

142. Langming, Z., Xiaohu, Z., and Banglei, G. (2014). A flexible method for multi-view point clouds alignment of small-size object. *Measurement*, *58*, 115–129.

143. Lara López, G., Peña Pérez Negrón, A., De Antonio Jiménez, A., Ramírez Rodríguez, J., and Imbert Paredes, R. (2016). Comparative analysis of shape descriptors for 3d objects. *Multimedia Tools and Applications*, (pp. 1–48).

144. Lawson, C. L., and Hanson, R. J. (1995). *Solving least squares problems* volume 15. SIAM.

145. Leordeanu, M., and Hebert, M. (2005). A spectral technique for correspondence problems using pairwise constraints. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (pp. 1482–1489 Vol. 2). Beijing, China: IEEE volume 2.

146. Levoy, M., Gerth, J., Curless, B., and Pull, K. (2005). The stanford 3d scanning repository. `http://www-graphics.stanford.edu/data/3dscanrep`.

147. Li, H., and Hartley, R. (2007). The 3d-3d registration problem revisited. In *2007 IEEE 11th International Conference on Computer Vision* (pp. 1–8). Rio de Janeiro, Brazil.: IEEE.

148. Li, X., and Guskov, I. (2005). Multiscale features for approximate alignment of point-based surfaces. In *Symposium on geometry processing* (pp. 217–226). Vienna, Austria. volume 255.

149. Li, Z., Zhu, J., Lan, K., Li, C., and Fang, C. (2014). Improved techniques for multi-view registration with motion averaging. In *2014 2nd International Conference on 3D Vision* (pp. 713–719). Tokyo, Japan. volume 1.

150. LiDAR (2016). Lidar. `http://www.lidar-uk.com/how-lidar-works/`.

151. Lindeberg, T. (1998). Feature detection with automatic scale selection. *International Journal of Computer Vision*, *30*, 79–116.

152. Liu, Y., Dominicis, L. D., Wei, B., Chen, L., and Martin, R. R. (2015). Regularization based iterative point match weighting for accurate rigid transformation estimation. *IEEE Transactions on Visualization and Computer Graphics*, *21*, 1058–1071.

153. LMI (2016). Gocator sensors. `http://lmi3d.com/quality-control-solutions/gocator`.

154. Lo, T.-W. R., and Siebert, J. P. (2009). Local feature extraction and matching on range images: 2.5d {sift}. *Computer Vision and Image Understanding*, *113*, 1235–1250.

155. Lobay, A., and Forsyth, D. A. (2006). Shape from texture without boundaries. *International Journal of Computer Vision*, *67*, 71–91.

156. Lomonosov, E., Chetverikov, D., and Ekárt, A. (2006). Pre-registration of arbitrarily oriented 3d surfaces using a genetic algorithm. *Pattern Recognition Letters*, *27*, 1201–1208.

157. Lorusso, A., Eggert, D. W., and Fisher, R. B. (1995). *A comparison of four algorithms for estimating 3-D rigid transformations*. University of Edinburgh, Department of Artificial Intelligence.

158. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*, 91–110.

159. Lu, F., and Milios, E. (1997). Globally consistent range scan alignment for environment mapping. *Autonomous Robots*, *4*, 333–349.

160. Lu, G., Han, K., DeSouza, G. N., Armer, J., and Shyu, C.-R. (2014). A new algorithm for 3d registration and its application in self-monitoring and early detection of lymphedema. *{IRBM}*, *35*, 370–384.

161. Ma, Y., Soatto, S., Kosecka, J., and Sastry, S. S. (2012). *An invitation to 3-d vision: from images to geometric models* volume 26. Springer Science & Business Media.

162. Magnusson, M., Lilienthal, A., and Duckett, T. (2007). Scan registration for autonomous mining vehicles using 3d-ndt. *Journal of Field Robotics*, *24*, 803–827.

163. Magnusson, M., Nuchter, A., Lorken, C., Lilienthal, A. J., and Hertzberg, J. (2009). Evaluation of 3d registration reliability and speed - a comparison of icp and ndt. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on* (pp. 3907–3912). Kobe, Japan.

164. Malassiotis, S., G.Strintzis, M., and Strintzis, M. G. (2007). Snapshots: A novel local surface descriptor and matching algorithm for robust 3d surface alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*, 1285–1290.

165. Marani, R., Renò, V., Nitti, M., D'Orazio, T., and Stella, E. (2016). A modified iterative closest point algorithm for 3d point cloud registration. *Computer-Aided Civil and Infrastructure Engineering*, *31*, 515–534.

166. Markelj, P., Tomaževič, D., Likar, B., and Pernuš, F. (2012). A review of 3d/2d registration methods for image-guided interventions. *Medical Image Analysis*, *16*, 642–661.

167. Markets, and Markets (2016). *3D Scanner Market by Offering (Hardware, Aftermarket Service), Type (Laser, Structured Light), Range, Product (Tripod Mounted, Fixed CMM Based, Portable CMM Based, Desktop), Application, Industry, and Geography - Global Forecast to 2022*. Technical Report SE 3566 MarketsandMarkets.

168. Marton, Z.-C., Pangercic, D., Blodow, N., and Beetz, M. (2011). Combined 2d-3d categorization and classification for multimodal perception systems. *The International Journal of Robotics Research*, *30*, 1378–1402.

169. Masuda, T. (2002). Registration and integration of multiple range images by matching signed distance fields for object shape modeling. *Computer Vision and Image Understanding*, *87*, 51–65.

170. Masuda, T. (2009). Log-polar height maps for multiple range image registration. *Computer Vision and Image Understanding*, *113*, 1158–1169.

171. Masuda, T., Sakaue, K., and Yokoya, N. (1996). Registration and integration of multiple range images for 3-d model construction. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on* (pp. 879–883 vol.1). Vienna, Austria. volume 1.

172. Matabosch, C., Fofi, D., Salvi, J., and Batlle, E. (2008). Registration of surfaces minimizing error propagation for a one-shot multi-slit hand-held scanner. *Pattern Recognition*, *41*, 2055–2067.

173. Matei, B., Shan, Y., Sawhney, H. S., Tan, Y., Kumar, R., Huber, D., and Hebert, M. (2006). Rapid object indexing using locality sensitive hashing and joint 3d-signature space estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 1111–1126.

174. Mateo, X., Orriols, X., and Binefa, X. (2014). Bayesian perspective for the registration of multiple 3d views. *Computer Vision and Image Understanding*, *118*, 84–96.

175. Mavridis, P., Andreadis, A., and Papaioannou, G. (2015). Efficient sparse icp. *Computer Aided Geometric Design*, *35Ű36*, 16–26.

176. Meng, D., Du, S., Zhu, J., Li, Z., and Yuan, Z. (2014). Robust registration of partially overlapping point sets via genetic algorithm with growth operator. *IET Image Processing*, *8*, 582–590.

177. Mian, A., Bennamoun, M., and Owens, R. (2010). On the repeatability and quality of keypoints for local feature-based 3d object retrieval from cluttered scenes. *International Journal of Computer Vision*, *89*, 348–361.

178. Mian, A. S., Bennamoun, M., and Owens, R. (2006). Three-dimensional model-based object recognition and segmentation in cluttered scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 1584–1601.

179. Mian, A. S., Bennamoun, M., and Owens, R. A. (2004). From unordered range images to 3d models: a fully automatic multiview correspondence algorithm. In *Theory and Practice of Computer Graphics, 2004. Proceedings* (pp. 162–166). Bournemouth, United Kingdom.

180. Mian, A. S., Bennamoun, M., and Owens, R. A. (2005). Automatic correspondence for 3d modeling: An extensive review. *International Journal of Shape Modeling*, *11*, 253–291.

181. Mian, A. S., and Bennamoun, R. A., M.and Owens (2006). A novel representation and feature matching algorithm for automatic pairwise registration of range images. *International Journal of Computer Vision*, *66*, 19–40.

182. Mitra, N. J., Gelfand, N., Pottmann, H., and Guibas, L. (2004). Registration of point cloud data from a geometric optimization perspective. In *Proceedings of the 2004 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing* SGP '04 (pp. 22–31). New York, NY, USA: ACM.

183. Mokhtarian, F., Khalili, N., and Yuen, P. (2001). Multi-scale free-form 3d object recognition using 3d models. *Image and Vision Computing*, *19*, 271–281.

184. Moons, T., Van Gool, L., and Vergauwen, M. (2009). *3D Reconstruction from multiple images, part 1: Principles*. Now Publishers Inc.

185. Morell-Gimenez, V., Saval-Calvo, M., Azorin-Lopez, J., Garcia-Rodriguez, J., Cazorla, M., Orts-Escolano, S., and Fuster-Guillo, A. (2014). A comparative study of registration methods for rgb-d video of static scenes. *Sensors*, *14*, 8547–8576.

186. Muja, M., and Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In *International Conference on Computer Vision Theory and Applications (VISAPP(1) '09)* (pp. 2.331–2.340). Lisbon, Portugal.

187. Myatt, D. R., Torr, P. H. S., Nasuto, S. J., Bishop, J. M., and Craddock, R. (2002). Napsac: High noise, high dimensional robust estimation - it's in the bag. In *Proceedings of the British Machine Vision Conference* (pp. 44.1–44.10). U.K.: BMVA Press.

188. Myronenko, A., Xubo Song, and Song, X. (2010). Point set registration: Coherent point drift. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*, 2262–2275.

189. Neugebauer, P. J. (1997a). Geometrical cloning of 3d objects via simultaneous registration of multiple range images. In *Shape Modeling and Applications, 1997. Proceedings., 1997 International Conference on* (pp. 130–139). AizuWakamatsu, Japan.

190. Neugebauer, P. J. (1997b). Reconstruction of real-world objects via simultaneous registration and robust combination of multiple range images. *International Journal of Shape Modeling*, *03*, 71–90.

191. Nishino Ko Ikeuchi, K. (2002). Robust simultaneous registration of multiple range images. In *ACCV 2002: Fifth Asian conference on computer vision, Proceedings* (pp. 454–461). Melbourne, Victoria.

192. Novatnack, J., and Nishino, K. (2008). Scale-dependent/invariant local 3d shape descriptors for fully automatic registration of multiple sets of range images. In D. Forsyth, P. Torr, and A. Zisserman (Eds.), *Computer Vision – ECCV 2008: 10th European Conference on Computer Vision, Proceedings, Part III* (pp. 440–453). Marseille, France.: Springer Berlin Heidelberg.

193. Novatnack, J., Nishino, K., and Shokoufandeh, A. (2006). Extracting 3d shape features in discrete scale-space. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on* (pp. 946–953). Chapel Hill, USA.

194. Pankaj, D., and Nidamanuri, R. (2016). A robust estimation technique for 3d point cloud registration. *Image Analysis & Stereology*, *35*, 15–28.

195. Pankaj, D. S., Nidamanuri, R. R., and Prasad, P. B. (2013). 3-d imaging techniques and review of products. In *Proceedings of International Conference on Innovations in Computer Science and Engineering*. Hyderabad, India.

196. Park, I. K., Germann, M., Breitenstein, M. D., and Pfister, H. (2010). Fast and automatic object pose estimation for range images on the gpu. *Machine Vision and Applications*, *21*, 749–766.

197. Pasqualotto, G., Zanuttigh, P., and Cortelazzo, G. M. (2013). Combining color and shape descriptors for 3d model retrieval. *Signal Processing: Image Communication*, *28*, 608–623.

198. Pauly, M., Gross, M., and Kobbelt, L. P. (2002). Efficient simplification of point-sampled surfaces. In *Visualization, 2002. VIS 2002. IEEE* Section 4 (pp. 163–170). Boston, Massachusetts.

199. Pauly, M., Keiser, R., and Gross, M. (2003). Multi-scale feature extraction on point-sampled surfaces. *Computer Graphics Forum*, *22*, 281–289.

200. Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., and Chamzas, C. (2007). Methods for 3d digitization of cultural heritage. *Journal of Cultural Heritage*, *8*, 93–98.

201. Pennec, X. (1996). Multiple registration and mean rigid shapes: Application to the 3d case. In *Image Fusion and Shape Variability Techniques (16th Leeds Annual Statistical Workshop)* (pp. 178–185). Leeds, UK.

202. Person, L. (2014). *3D Technology Market by Product (3D printer, 3D material, 3D glasses, 3D display technology, 3D imaging software, 3D camera) - Global Opportunity Analysis and Industry Forecast, 2012 - 2020*. Technical Report SE 14152 Emerging and Next Generation Technologies.

203. Petrelli, A., and Di Stefano, L. (2016). Pairwise registration by local orientation cues. *Computer Graphics Forum*, *35*, 59–72.

204. Phillips, J. M., Liu, R., and Tomasi, C. (2007). Outlier robust icp for minimizing fractional rmsd. In *3-D Digital Imaging and Modeling, 2007. 3DIM '07. Sixth International Conference on* (pp. 427–434). Montreal, Quebec, Canada.

205. Pieraccini, M., Guidi, G., and Atzeni, C. (2001). 3d digitizing of cultural heritage. *Journal of Cultural Heritage*, *2*, 63–70.

206. Pomerleau, F., Colas, F., and Siegwart, R. (2015). A review of point cloud registration algorithms for mobile robotics. *Foundations and Trends in Robotics*, *4*, 1–104.

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

208. Pratt, W. K. (2001). *Digital Image Processing: PIKS Inside*. (3rd ed.). New York, NY, USA: John Wiley & Sons, Inc.

209. Pulli, K. (1999). Multiview registration for large data sets. In *3-D Digital Imaging and Modeling, 1999. Proceedings. Second International Conference on* (pp. 160–168). Kyoto, Japan.

210. Restrepo, M. I., and Mundy, J. L. (2012). An evaluation of local shape descriptors in probabilistic volumetric scenes. In *Procedings of the British Machine Vision Conference 2012* November (pp. 46.1–46.11). Guildford, Surrey, United Kingdom.

211. Rohr, K., Stiehl, H. S., Sprengel, R., Buzug, T. M., Weese, J., and Kuhn, M. H. (2001). Landmark-based elastic registration using approximating thin-plate splines. *IEEE Transactions on Medical Imaging*, *20*, 526–534.

212. Rosten, E., Porter, R., and Drummond, T. (2010). Faster and better: A machine learning approach to corner detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*, 105–119.

213. Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, *79*, 871–880.

214. Rousseeuw, P. J., and Leroy, A. M. (2005). *Robust regression and outlier detection* volume 589. John Wiley & Sons.

215. Ruiz-Correa, S., Shapiro, L. G., and Melia, M. (2001). A new signature-based method for efficient 3-d object recognition. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on* (pp. I–769–I–776 vol.1). Kauai, HI, USA. volume 1.

216. Rusinkiewicz, S., Hall-Holt, O., and Levoy, M. (2002). Real-time 3d model acquisition. *ACM Trans. Graph.*, *21*, 438–446.

217. Rusinkiewicz, S., and Levoy, M. (2001). Efficient variants of the icp algorithm. In *Proceedings of International Conference on 3-D Digital Imaging and Modeling, 3DIM* (pp. 145–152). Quebec City, Canada.

218. Rusu, R. B. (2010). *Semantic 3d object maps for everyday manipulation in human living environments*. Ph.D. thesis.

219. Rusu, R. B., Blodow, N., and Beetz, M. (2009a). Fast point feature histograms (fpfh) for 3d registration. In *Robotics and Automation, 2009. ICRA '09. IEEE International Conference on* (pp. 3212–3217). Kobe, Japan.

220. Rusu, R. B., Blodow, N., Marton, Z. C., and Beetz, M. (2008). Aligning point cloud views using persistent feature histograms. In *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3384–3391). Nice, France.

221. Rusu, R. B., Bradski, G., Thibaux, R., and Hsu, J. (2010). Fast 3d recognition and pose using the viewpoint feature histogram. In *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on* (pp. 2155–2162). Taipei, Taiwan.

222. Rusu, R. B., and Cousins, S. (2011). 3d is here: Point cloud library (pcl). In *Robotics and Automation (ICRA), 2011 IEEE International Conference on* (pp. 1–4). Shanghai, China.

223. Rusu, R. B., Holzbach, A., Beetz, M., and Bradski, G. (2009b). Detecting and segmenting objects for mobile manipulation. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009* (pp. 47–54). Kyoto, Japan.

224. Salti, S., Petrelli, A., Tombari, F., Di Stefano, L., and Stefano, L. D. (2012). On the affinity between 3d detectors and descriptors. In *Proceedings - 2nd Joint 3DIM/3DPVT Conference: 3D Imaging, Modeling, Processing, Visualization and Transmission, 3DIMPVT 2012* (pp. 424–431). Zurich, Switzerland.

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

226. Salvi, J., Pagès, J., and Batlle, J. (2004). Pattern codification strategies in structured light systems. *Pattern Recognition*, *37*, 827–849.

227. Sansoni, G., Trebeschi, M., and Docchio, F. (2009). State-of-the-art and applications of 3d imaging sensors in industry, cultural heritage, medicine, and criminal investigation. *Sensors*, *9*, 568–601.

228. Schaller, C. (2011). *Time-of-flight-a new modality for radiotherapy*. Ph.D. thesis University of Erlangen-Nuremberg.

229. Segal, A., Haehnel, D., and Thrun, S. (2009). Generalized-icp. In *Robotics: Science and Systems Conference*. Seattle, Washington. volume 2.

230. Serafin, J., and Grisetti, G. (2014). Using augmented measurements to improve the convergence of icp. In D. Brugali, J. F. Broenink, T. Kroeger, and B. A. MacDonald (Eds.), *Simulation, Modeling, and Programming for Autonomous Robots: 4th International Conference, SIMPAR 2014, Proceedings* (pp. 566–577). Bergamo, Italy,: Springer International Publishing.

231. Sharp, G. C., Lee, S. W., and Wehe, D. K. (2004). Multiview registration of 3d scenes by minimizing error between coordinate frames. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *26*, 1037–1050.

232. Shih, S. W., Chuang, Y. T., and Yu, T. Y. (2008). An efficient and accurate method for the relaxation of multiview registration error. *IEEE Transactions on Image Processing*, *17*, 968–981.

233. Shiratori, T., Berclaz, J., Harville, M., Shah, C., Li, T., Matsushita, Y., and Shiller, S. (2015). Efficient large-scale point cloud registration using loop closures. In *Proceedings - 2015 International Conference on 3D Vision, 3DV 2015* (pp. 232–240). Lyon, France.

234. SICK (2016). Sick 3d cameras. `https://www.sick.com/us/en/product-portfolio/vision/3d-vision/c/g138560`.

235. SICKRanger (2016). Sick 3d ranger camera. `http://www.sickinsight-online.com/ranger-3d-camera-grading-70000-oysters-per-hour/`.

236. Sipiran, I., and Bustos, B. (2011). Harris 3d: A robust extension of the harris operator for interest point detection on 3d meshes. *Visual Computer*, *27*, 963–976.

237. Smith, S. M., and Brady, J. M. (1997). Susan—a new approach to low level image processing. *International Journal of Computer Vision*, *23*, 45–78.

238. Soucy, M., and Laurendeau, D. (1992). Multi-resolution surface modeling from multiple range views. In *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR '92., 1992 IEEE Computer Society Conference on* (pp. 348–353). Illinois, USA.

239. Soucy, M., and Laurendeau, D. (1995). A general surface approach to the integration of a set of range views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *17*, 344–358.

240. Spreeuwers, L. (2011). Fast and accurate 3d face recognition. *International Journal of Computer Vision*, *93*, 389–414.

241. Sprickerhof, J., Nüchter, A., Lingemann, K., and Hertzberg, J. (2009). An explicit loop closing technique for 6d slam. In *Proc. European Conference on Mobile Robots, ECMR* (pp. 229–234). Mlini, Croatia.

242. Stamos, I., and Leordeanu, M. (2003). Automated feature-based range registration of urban scenes of large scale. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on* (pp. II–555–Ii–561). Madison, Wisconsin. volume 2.

243. Steder, B., Rusu, R. B., Konolige, K., and Burgard, W. (2010). Narf: 3d range image features for object recognition. In *Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*. Taipei, Taiwan. volume 44.

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

245. Stein, F., and Medioni, G. (1992). Structural indexing: Efficient 3-d object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *14*, 125–145.

246. Stoddart, A. J., and Hilton, A. (1996). Registration of multiple point sets. In *Pattern Recognition, 1996., Proceedings of the 13th International Conference on* (pp. 40–44). Vienna, Austria. volume 2.

247. Strecha, C., Lindner, A., Ali, K., and Fua, P. (2009). Training for task specific keypoint detection. In J. Denzler, G. Notni, and H. Süße (Eds.), *Pattern Recognition: 31st DAGM Symposium, Proceedings* (pp. 151–160). Jena, Germany.: Springer Berlin Heidelberg.

248. Sun, J., Ovsjanikov, M., and Guibas, L. (2009). A concise and provably informative multi-scale signature based on heat diffusion. *Computer graphics forum*, *28*, 1383–1392.

249. Sun, Y., and Abidi, M. A. (2001). Surface matching by 3d point's fingerprint. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on* (pp. 263–269). Vancouver, BC, Canada. volume 2.

250. Sun, Z., van de Ven, J., Ramos, F., Mao, X., and Durrant-Whyte, H. (2012). Inferring laser-scan matching uncertainty with conditional random fields. *Robotics and Autonomous Systems*, *60*, 83–94.

251. Taati, B., Bondy, M., Jasiobedzki, P., and Greenspan, M. (2007). Variable dimensional local shape descriptors for object recognition in range data. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 1–8). Rio de Janeiro, Brazil.

252. Taati, B., and Greenspan, M. (2011). Local shape descriptor selection for object recognition in range data. *Computer Vision and Image Understanding*, *115*, 681–694.

253. Takeuchi, E., and Tsubouchi, T. (2006). A 3-d scan matching using improved 3-d normal distributions transform for mobile robotic mapping. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 3068–3073). Beijing, China.

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

255. Tao, J., and Jiyong, K. (2007). A 3-d point sets registration method in reverse engineering. *Computers & Industrial Engineering*, *53*, 270–276.

256. Teran, L., and Mordohai, P. (2014). 3d interest point detection via discriminative learning. In D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars (Eds.), *13th European Conference on Computer Vision (ECCV), Proceedings, Part IV* (pp. 1–8). Zurich, Switzerland.: Springer International Publishing.

257. Tomasi, C., and Kanade, T. (1991). *Detection and tracking of point features*. School of Computer Science, Carnegie Mellon Univ. Pittsburgh.

258. Tombari, F., Salti, S., and Di Stefano, L. (2010a). Unique shape context for 3d data description. In *Proceedings of the ACM Workshop on 3D Object Retrieval* 3DOR '10 (pp. 57–62). New York, NY, USA: ACM.

259. Tombari, F., Salti, S., and Di Stefano, L. (2010b). Unique signatures of histograms for local surface description. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6313 LNCS*, 356–369.

260. Tombari, F., Salti, S., and Di Stefano, L. (2013). Performance evaluation of 3d keypoint detectors. *International Journal of Computer Vision*, *102*, 198–220.

261. Tombari, F., Salti, S., and Stefano, L. D. (2011). A combined texture-shape descriptor for enhanced 3d feature matching. In *2011 18th IEEE International Conference on Image Processing* (pp. 809–812). Brussels, Belgium.

262. Tombari, F., and Stefano, L. D. (2012). Hough voting for 3d object recognition under occlusion and clutter. *IPSJ Transactions on Computer Vision and Applications*, *4*, 20–29.

263. Ton, J., and Jain, A. K. (1989). Registering landsat images by point matching. *IEEE Transactions on Geoscience and Remote Sensing*, *27*, 642–651.

264. Torr, P. H. S., Szeliski, R., and Anandan, P. (2001). An integrated bayesian approach to layer extraction from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 297–303.

265. Torr, P. H. S., and Zisserman, A. (2000). Mlesac: A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, *78*, 138–156.

266. Turk, G., and Levoy, M. (1994). Zippered polygon meshes from range images. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques* SIGGRAPH '94 (pp. 311–318). New York, NY, USA: ACM.

267. Umeyama, S. (1991). Least-squares estimation of transformation parameters between two point patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *13*, 376–380.

268. Unnikrishnan, R., and Hebert, M. (2008). Multi-scale interest regions from unorganized point clouds. In *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW '08. IEEE Computer Society Conference on* (pp. 1–8). Anchorage, Alaska, USA.

269. VC, H. P. (1962). Method and means for recognizing complex patterns. US Patent 3,069,654.

270. Verdie, Y., Yi, K. M., Fua, P., and Lepetit, V. (2015). Tilde: A temporally invariant learned detector. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5279–5288). Boston, MA, USA.

271. Viola, P., and Jones, M. J. (2004). Robust real-time face detection. *International Journal of Computer Vision*, *57*, 137–154.

272. VisionComponents (2016). Vc nano 3d. `http://www.vision-components.com/de/produkte/oem-embedded-vision-systeme/3d-systeme/`.

273. Walker, M. W., Shao, L., and Volz, R. A. (1991). Estimating 3-d location parameters using dual number quaternions. *CVGIP: Image Understanding*, *54*, 358–367.

274. Wang, H. (2004). *Robust statistics for computer vision: model fitting, image segmentation and visual motion analysis*. Ph.D. thesis.

275. Wang, X., Li, Y., Wei, H., and Liu, F. (2015). An asift-based local registration method for satellite imagery. *Remote Sensing*, *7*, 7044–7061.

276. Weik, S. (1997). Registration of 3-d partial surface models using luminance and depth information. In *3-D Digital Imaging and Modeling, 1997. Proceedings., International Conference on Recent Advances in* (pp. 93–100). Ottawa, Ontario, Canada.

277. Weise, T., Wismer, T., Leibe, B., and Gool, L. V. (2011). Online loop closure for real-time interactive 3d scanning. *Computer Vision and Image Understanding*, *115*, 635–648.

278. Williams, J. A., and Bennamoun, M. (2001). Simultaneous registration of multiple corresponding point sets. *Comp Vis Imag Under*, *81*, 117–142.

279. Wohlkinger, W., and Vincze, M. (2011). Ensemble of shape functions for 3d object classification. In *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on* (pp. 2987–2992). Phucket, Thailand.

280. Woodham, R. J. (1980). Photometric method for determining surface orientation from multiple images. *Optical Engineering*, *19*, 191139–191139.

281. Wu, Y., Wang, W., Lu, K., Wei, Y., and Chen, Z. (2015). A new method for registration of 3d point sets with low overlapping ratios. *Procedia CIRP*, *27*, 202–206.

282. Wyngaerd, J. V. (2003). Combining texture and shape for automatic crude patch registration. In *3-D Digital Imaging and Modeling, 2003. 3DIM 2003. Proceedings. Fourth International Conference on* (pp. 179–186). Banff, Alta., Canada.

283. Yamany, S. M., and Farag, A. A. (2002). Surface signatures: an orientation independent free-form surface representation scheme for the purpose of objects registration and matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 1105–1120.

284. Yang, B., and Zang, Y. (2014). Automated registration of dense terrestrial laser-scanning point clouds using curves. *{ISPRS} Journal of Photogrammetry and Remote Sensing*, *95*, 109–121.

285. Yang, J., Li, H., and Jia, Y. (2013). Go-icp: Solving 3d registration efficiently and globally optimally. In *The IEEE International Conference on Computer Vision (ICCV)* (pp. 1457–1464). Sydney, Australia.

286. Yang, Y., Zhang, P. P., Qiao, Y., Yang, J., and Wang, S. Z. (2014). A robust cpd approach based on shape context. In *Proceedings of the 33rd Chinese Control Conference* (pp. 4930–4935). Nanjing, China.: IEEE.

287. Zaharescu, A., Boyer, E., Varanasi, K., and Horaud, R. (2009). Surface feature detection and description with applications to mesh matching. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on* (pp. 373–380). Miami, FLorida.

288. Zhang, L., Sturm, J., Cremers, D., and Lee, D. (2012). Real-time human motion tracking using multiple depth cameras. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 2389–2395). Vilamoura, Portugal.

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

290. Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. *International Journal of Computer Vision*, *13*, 119–152.

291. Zhong, Y. (2009). Intrinsic shape signatures: A shape descriptor for 3d object recognition. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on* (pp. 689–696). Kyoto, Japan.

292. Zou, G., Hua, J., Dong, M., and Qin, H. (2008). Surface matching with salient keypoints in geodesic scale space. *Computer Animation and Virtual Worlds*, *19*, 399–410.

293. Zou, G., Hua, J., Lai, Z., Gu, X., and Dong, M. (2009). Intrinsic geometric scale space by shape diffusion. *IEEE Transactions on Visualization and Computer Graphics*, *15*, 1193–1200.

LIST OF PAPERS BASED ON THESIS

Papers in Refereed International Journals

(i) Pankaj, D. S., & Nidamanuri, R. R. (2016). A robust estimation technique for 3d point cloud registration. *Image Analysis & Stereology*, 35 (1), 15-28.

(ii) Pankaj, D.S, & Nidamanuri, R. R. (2017). Robust algorithm for multiview registration. *IET Computer Vision*, 11 (1), 96-103 (7).

Presentations in Conferences

(i) Pankaj, D. S., Nidamanuri, R. R., & Pinnamaneni, B. P. (2014, December). Prolosac: An improved ransac algorithm for pairwise registration of 3D point clouds. *In Proceedings of the 2014 Indian Conference on Computer Vision Graphics and Image Processing (ICVGIP)* (pp. 30:1-30:8). ,Bangalore, India.: ACM .

(ii) Pankaj, D. S., Nidamanuri, R. R. (2016, July). Robust multiview registration of point clouds. *In Communication Systems and Networks (ComNet), International Conference on*, (pp.50-55). , Trivandrum, India.: IEEE

(iii) Pankaj, D. S., Nidamanuri, R. R., & Prasad, P. B. (2013). 3-d imaging techniques and review of products. *In Proceedings of International Conference on Innovations in Computer Science and Engineering.*(pp.21-27), Hyderabad, India.

Manuscripts under processing

(i) Pankaj, D.S, & Nidamanuri, R. R. A robust estimation based method for automatic registration of remote sensing images. *ISPRS Journal of Photogrammetry and Remote Sensing*, accepted for publication.

(ii) Pankaj, D.S, & Nidamanuri, R. R. An end-to-end framework for automatic registration of TLS 3D point clouds. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS)*.

(iii) Pankaj, D.S, & Nidamanuri, R. R. A review of 3D point cloud registration techniques. *IET Computer Vision*.