



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Research Commons

<http://waikato.researchgateway.ac.nz/>

Research Commons at the University of Waikato

Copyright Statement:

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

The thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of the thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from the thesis.

3D Shape Reconstruction from Multiple Range Image Views

A thesis
submitted in partial fulfillment
of the requirements for the degree
of

Master of Science
in Physics and Electronic Engineering

at

The University of Waikato

by

Kartick Ganapathi Annadurai



THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

Hamilton
2006

Dedicated To

My Mum and Dad

My Brother Shathiyakkumar

My Sister Indumathi

Acknowledgement

I am extremely grateful and privileged to be educated in Image Processing by a distinguished physicist and chief supervisor Dr Michael Cree. Personal gratitude for his patience, guidance and immense support over the last two years as I have worked towards this thesis.

In addition, without the help and support of many people it may not have been possible to complete this thesis. To Dr Adrian Dorrington I am more grateful, for his efforts during my work on simulated study. Also thank you Mr Richard Convoy for your help and support during acquisition with the Waikato Range Imager. Thanks to staff Heidi and Bruce in the Department of Physics & Electronic Engineering for their timely assistance. Thanks to flat mate Mr Anandajothi and Mr Dipu for taking extreme care of my health and helping with thesis writing.

Many thanks must go to friends Kel, Jono, Inky, AJ, Dipu and Bobby for their valuable time. Special thanks to Shene Kurien for his timely advice and encouragement during my thesis. Above all, very special thanks to Mum and Dad for their patience and support.

Abstract

Shape reconstruction of different three dimensional objects using multiple range images has evolved recently within the recent past. In this research shape reconstruction of a three dimensional object using multiple range image views is investigated.

Range images were captured using the Waikato Range Imager. This range images camera is novel in that it uses heterodyne imaging and is capable of acquiring range images with precision less than a millimeter simultaneously over a full field. Multiple views of small objects were taken and the FastRBF was explored as a mean of registration and surface rendering. For comparison to the real range data, simulated range data under noise free condition were also generated and reconstructed with the FastRBF tool box.

The registration and reconstruction of simple object was performed using different views with the FastRBF toolbox. Analysis of the registration process showed that the translation error produced due to distortion during registration of different views hinders the process of reconstructing a complete surface. While analyzing the shape reconstruction using the FastRBF tool it is also determined that a small change in accuracy values can affect the interpolation drastically. Results of reconstruction of a real 3D object from multiple views are shown.

Table of contents

Acknowledgement-----	iii
Abstract -----	v
Table of contents -----	vii
List of figures -----	xi
List of Tables-----	xv
1 Introduction -----	1
2 Background Theory-----	5
2.1 Digital image processing -----	5
2.1.1 Image Acquisition -----	6
2.1.2 Image Enhancement -----	6
2.1.3 Image Restoration -----	7
2.1.4 Image Analysis-----	7
2.1.5 Image Synthesis-----	8
2.2 Image data handling-----	9
2.3 Range imaging-----	10
2.4 Geometrical Representation-----	11

2.4.1	Euclidean coordinate -----	11
2.5	Projections-----	12
2.5.1	Parallel projections -----	12
2.5.2	Central projections-----	13
2.6	Coordinate Transformation -----	14
2.7	Distortions-----	16
2.7.1	Radial Distortion-----	16
2.7.2	Geometrical Distortion-----	17
2.8	Background Segmentation -----	18
2.9	3D Shape Reconstruction Methods Reviewed-----	19
2.10	Overview of Registration Process -----	21
3	Equipment and hardware -----	25
3.1	Image Capturing-----	25
3.1.1	Imaging Lidar -----	25
3.1.2	Image Intensifier-----	27
3.1.3	Signal Generator -----	28
3.1.4	Ranger Camera-----	29
3.1.5	CCD image Sensor -----	29
3.1.6	Light Source-----	30
3.1.7	Grabber-----	31
3.2	Ranger Precision-----	31
4	Experimental Methodologies -----	33
4.1	FastRBF tool -----	33
4.1.1	Fitting RBF to surface data-----	35
4.1.2	Mesh Generation-----	36

4.1.3	Using FastRBF -----	38
4.2	Simulation Work -----	39
4.2.1	Principle-----	39
4.2.2	Setup Description -----	40
4.2.3	Procedure -----	41
4.2.4	Setup Values used for the simulation study -----	47
4.3	Experiment using Real Range Data -----	48
4.4	Registration and Reconstruction -----	57
5	Results -----	59
5.1	Simulated Reconstruction results -----	59
5.2	Experiment Reconstruction results-----	65
6	Conclusion -----	79
	References -----	83

List of figures

Figure 2.1 : Block diagram showing the stages in Digital Image Processing.....	5
Figure 2.2 : A surface point S in Euclidean coordinate system	12
Figure 2.3 : Image captured using parallel projection	13
Figure 2.4 : Scene point captured on image plane using central projection	14
Figure 2.5 : The three three-dimensional rotations.	15
Figure 2.6 : Distortion produced on a checked plane while image acquiring. (Source: www.umich.edu/~lowbrows/guide/opticaljargon.html)	16
Figure 2.7 : Geometrical distortion resulting from increased path length of off axis objects.	18
Figure 3.1 : Heterodyne imaging lidar systems (Source: Payne <i>et al.</i> , 2006).....	26
Figure 3.2 : Image intensifier depicting how a single photon is multiplied to many due to electron acceleration (Source: Payne <i>et al.</i> , 2006).....	28
Figure 3.3 : CCD used in Ranger camera	30
Figure 4.1 : Plot showing Fitter accuracy and evaluation accuracy. (Source : Farfield technology).....	34
Figure 4.2 : Creating density data from surface normals (Source: Farfield technology)	36
Figure 4.3 : Comparison of RBF evaluations (Source : Farfield technology)	37
Figure 4.4 : The results for different mesh optimisation methods (Source: Farfield technology).....	38
Figure 4.5 : Scene used for simulation work	39
Figure 4.6: Simulation setup showing image plane, centre of projection and the scene.	41
Figure 4.7 : Ray tracing - Intersection point of the ray originating at R_0 with directional vector R_d on the finite plane at $R(t)$	43
Figure 4.8 : A ray intersecting at two points R and R1. The shortest range value for the is recorded.	45
Figure 4.9 : Simulated range image	46

Figure 4.10 : Experimental setup used for capturing the test object.	48
Figure 4.11 : Test object bear used in this study as scene is placed on a turntable.....	50
Figure 4.12 : Range image front view of bear captured from ranger camera.....	51
Figure 4.13 : Range image cropped to 351×291 from 512×512	52
Figure 4.14 : Establishing world coordinates for every pixel in the range image.....	53
Figure 4.15 : Range image showing 0° overlapped with 180° (dashed line) and the x coordinate centre	55
Figure 4.16 : Range image showing calculation of z axis centre (see explanation for detail).....	56
Figure 5.1 : Simulated 0° Range image for 101×101 resolution image	59
Figure 5.2 : Simulated 270° Range image for 101×101 resolution image	60
Figure 5.3 : Simulated 0° Range image for 512×512 resolution image	60
Figure 5.4 : Simulated 270° Range image for 512×512 resolution image	61
Figure 5.5 : Reconstructed surface using 0° , 30° , 180° and 330° range data (front on)	63
Figure 5.6 : Reconstructed surface using 0° , 30° , 180° and 330° range data (side view)	63
Figure 5.7 : Reconstructed surface using 0° , 30° , 180° and 330° range data (side view)	64
Figure 5.8: Reconstructed surface using 0° , 30° , 180° and 330° range data (rear view)	64
Figure 5.9 : a) Intensity image and b) range image for the 0° view.....	65
Figure 5.10 : a) Intensity image and b) range image for 30° view.....	66
Figure 5.11 : a) Intensity image and b) range image for 270° view.....	66
Figure 5.12 : Segmentation of bear based on range values only	67
Figure 5.13 : Segmentation based on using both range and intensity values	68
Figure 5.14 : Shape reconstructed 0° degree range image.....	69
Figure 5.15 : Shape reconstruction using Fast RBF for two set of data (0° and 30°)..	70
Figure 5.16 : Scatter plot showing two set of range data 0° and rotated 30°	72
Figure 5.17 : Error between two data sets determined using point evaluation.....	72

Figure 5.18 : Deviation error between two data sets across the image at five locations	73
Figure 5.19 : Reconstructed bear with corrected range data.....	74
Figure 5.20 : Side view of the reconstructed bear.....	75
Figure 5.21 : Side view of the reconstructed bear showing the rotated 30° range data	75
Figure 5.22 : Scatter plot showing two set of range data (0° and new rotated 30°) ...	76
Figure 5.23 : Error between two new data sets determined using point evaluation....	76
Figure 5.24 : Deviation error between two data sets across the image at five locations for the coordinate registration values.....	77
Figure 5.25 : Linear fit for a single row across the deviation error image.....	77

List of Tables

Table 1 : Scene coordinate values used in the simulated robot scene.....	40
Table 2 : Showing the values used for simulation	47
Table 3 : Experimental Setup values.....	49
Table 4: Showing the number pixels and corresponding measurement across the range image and the bear	54

1 Introduction

Methods to reconstruct the shapes of different three dimensional objects have evolved rapidly in recent years. The speed and accuracy of digitizing technologies have given much to advance in the areas of physics and electrical engineering, including the development of lasers and high speed sampling and timing circuitry. Such technologies allow us to take detailed shape measurements with precision better than 1 part per 1000 at rates exceeding 10000 samples per second. To capture the complete shape of an object, many thousands of samples must be acquired. The resulting mass of data requires algorithms that can efficiently and reliably generate models from these samples. The images used in this study are range images in which the distance from the camera to a point on the scene is encoded in the image as a grey value. Each range image provides a detailed description of an object as seen from one point of view. Before attempting to reconstruct the entire shape of an object, we require multiple range images. A single range image is taken from one point of view and much of the object being viewed may be obscured, therefore multiple range images are taken. Some objects are too complex to be captured in detail by a small number of range images. For example, faces lying in mutually orthogonal directions and their opposites sides of a six faced cube. Due to self occlusions, an object may require a large number of range scans to see every visible point. Thus multiple range images are required for the reconstruction of the surface.

The principle goal of this thesis is to explore, merging multiple range images taken from different views of an object, to reconstruct the full 3D surface. In this study a simulated reconstruction and real world reconstruction are undertaken. The simulated reconstruction study is carried out to know how well the reconstruction algorithm works in a noise free situation and under known conditions. The simulated object which resembles a roboshape is generated by computer (see Figure 4.5). The simulated object is rotated and a number of range images are recorded from different views.

The real test object includes a bear (see Figure 4.11). The test object is of uniform colouration. However the presence of little discolouration were permitted. The test object is manually rotated on a turntable, and a number of range images are recorded from different views, as the object is rotated full circle. The three dimensional data for the scene is obtained from the range images. Shape reconstruction used in this study is the generation of surface data and does not include the generation of interior data as in the case of computer tomography, nor does it include further information on colour. The objects or scene to be imaged are also assumed to be of a static nature, as it does take some minutes to acquire the range data. The acquiring of range image is performed using the Waikato ranger system. The main advantage of using Waikato ranger system is its high precision in range values. After capturing, in order to merge a set of range images into a single description of an object, it is necessary to place them all in the same coordinate system; that is they must be registered or aligned with respect to each other. The alignment arises from prior knowledge of the pose of the rangefinder when acquiring the range image. After alignment, the range images provide the starting point for performing a surface reconstruction.

The proposed work uses FastRBF toolbox for interpolation and reconstruction. This tool takes the aligned range data as input; it interpolates them and constructs it in to a complete surface with given resolution, which is the final goal of this research. Exploring the FastRBF tool for reconstruction using the Waikato ranger data forms one of the goals of this research. There are only few models available that uses FastRBF tool for reconstruction using range images. Also there have been no shape reconstruction model yet published that uses FastRBF software for the Waikato ranger data. So the detailed study and results of my proposed work forms a platform for those uses the Waikato ranger data for reconstruction.

The thesis is structured in the following manner. First in chapter two we introduce background theory which forms the fundamental basis of the research. In this chapter image processing techniques that play a large role in the processing of the visual data are detailed as; fundamental theorems, definitions about the representation and acquisition of the scene. Finally an overview of the registration process and

review of the available reconstruction models is presented to finish up the background theory. The chapter three the equipment and hardware used in the study is described. Following that in chapter four the testing methodology and the proposed work on this study is discussed in the experimental methodology chapter. The chapter begins with discussion on the FastRBF software, the tool used for interpolating the data points to represent a surface. Then discussion about the experimental setup and process of performing shape reconstruction for both simulated study and the test objects is explained. This chapter contains some original content introduced by myself such as exploring the FastRBF software for reconstruction of surface using the range data obtained from the Waikato ranger system. The results of which were discussed in the following chapter. In chapter five results for the reconstructed outputs of the simulated work and test object are presented. A discussion detailing the parameters at which particular results were obtained as well as features or characteristics worthy of attention are also identified. Some errors identified are rectified and new results with rectified errors were discussed. In the final chapter the overall conclusion of the research is given, and the proposal for future work on the research topic covered.

2 Background Theory

This chapter discusses the basics of the digital image processing and the stages involved in that, following an overview of range imaging, geometrical representation of the surface and major distortions affecting the acquisition process. Finally, an overview of models available for shape reconstruction using multiple range images is discussed.

2.1 Digital image processing

Digital image processing is the processing of digital images by means of digital computation (Gonzalez and Woods, 2002). Digital images are normally expressed as an array of numbers where the elements of the array are called pixels. Normally the value of each pixel represents the intensity of light received however in this thesis we shall often use images in which the value of each pixel is range between the scene and camera. The five major process involved in digital image processing are image acquisition, image enhancement, image restoration, image analysis and image synthesis (see Figure 2.1). These processes are explained in the following sections.

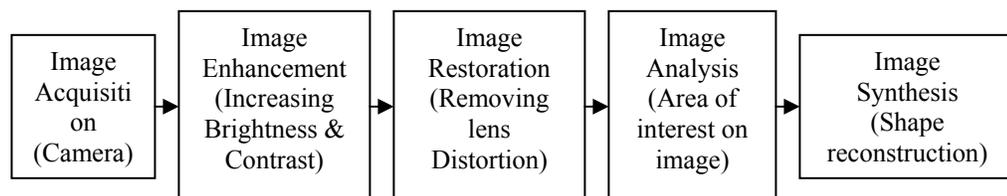


Figure 2.1 : Block diagram showing the stages in Digital Image Processing

2.1.1 Image Acquisition

A scene captured by some imaging device that converts the received light to an electrical signal. This electrical signal cannot directly be handled by digital circuits so the signals are converted to a discrete form by a digitizer. The resulting image can then directly be used in digital image processing applications. The digitizer performs sampling and quantization tasks. The values of continuous signals are sampled at specific locations in the image. After sampling the signals real values are discretized into digital numbers in the quantization process which gives a digital image (Baxes, 1994). The two major steps involved in image acquisition are choosing a camera and choosing the acquisition mode. The choice of choosing a camera may be difficult since there are so many different options. The two main types of camera are analog and digital camera. Analog cameras are cameras that generate a video signal in analog format. Digital cameras have several advantages over analog cameras. By digitizing at the camera level rather than at the image acquisition device, the signal-to-noise ratio is typically higher, resulting in better accuracy. Digital cameras use charge coupled device (CCD) for acquiring images. A Charge coupled device (CCD) is an array of interconnected photosensitive elements that stores and transfers its charge to a shift register, which converts the spatial array of charges in the CCD imager into a time-varying video signal. Timing information for the vertical and horizontal positions of the CCD and the sensor value combine to form the video signal. The various acquisition modes available are grab mode, snap mode, ring mode and sequence mode. The processing and display can then take place after the acquisition is complete.

2.1.2 Image Enhancement

The process of improving the quality of a digitally stored image by manipulating the image is image enhancement. To make an image lighter or darker, or to increase or decrease contrast, or reduce its noise contents, are examples of image enhancement. Image enhancement techniques are often used to improve an

image for a subsequent image analysis operation. Image enhancement techniques may be grouped in to subjective enhancement or objective enhancement (Baxes, 1994).

Subjective enhancements are used to make an image visually appealing for subsequent processing. A sharpening operation may be applied to improve the appearance of the image. This may be repeated until the image yields the details necessary for the particular application. Objective image enhancement is the method to enhance the details of an image for particular area in digital image. These enhancements need not be visually appealing. Edge enhancement is one of the processes wherein only edge details are highlighted. In this research the edge enhancement is used for finding the coordinate centre of the scene image as explained in chapter 4.

2.1.3 Image Restoration

Image restoration removes or minimizes some known distortions in an image. Image restoration techniques may be used to restore images with problems like geometric distortion, improper focus and camera motion (Gonzalez and woods, 1993). Sometimes the distortion induced into an image may not be known. In these cases it may be possible to estimate how the image was distorted and recreate a close approximation of the original. Photometric restoration corrects an image for poor intensity responses. Similarly geometric distortions can be caused by the imaging system and also by the camera lens. One such distortion by camera lens is pincushion or barrel distortion. This distortion makes the image appear bloated in the centre. A method to restore this kind of image is discussed in section 2.7. Images with misfocus appear fuzzy, with a lack of detail. By making assumptions about the blurring function that caused the distortion, it is possible to remove the distortion by applying inverse filtering process.

2.1.4 Image Analysis

Image analysis combines techniques that compute statistics and measurements based on the grey-level intensities of the image pixels. The aim of image analysis is to understand an image by separating the elements of interest, generally objects in an

image. Their quantification includes such things as measure of size, range and description of outlines. Image analysis processes can be subdivided into three steps. The first is image preprocessing, in which useless and distracting information from the image is removed. Care must be taken that this preprocess operation must not degrade the image in ways that interfere with the overall image analysis. The second stage is initial object discrimination, where objects with like characteristics are grouped separately. Each process in this stage works to isolate objects in an image, either by highlighting similar objects with a common brightness or by their edges. The final stage is object boundary cleanup, where object boundaries are reduced to single pixel widths.

Image analysis algorithms for monochrome images are generally based on one of two basic properties of grey level values; namely discontinuity and similarity. The first approach is to partition an image based on abrupt changes in grey level. The areas of interest in this approach are detection of lines and edges in an image. The second approach is based on thresholding and region splitting or merging (Gonzalez and Woods, 1993). Image analysis also yields various image statistics. One important statistic is brightness histogram, which is the distribution of grey level in an image. This distribution is displayed in a graphical form. The histogram describes the overall or regional contrast attributes of an image, and can therefore be used to determine contrast enhancement parameters. Other image statistics such as the frequency content can also be useful information in carrying out the subsequent processes. In this research extracting the range information, avoiding the areas of less intensity near edges where the error is high are carried out through this image analysis process.

2.1.5 Image Synthesis

Image synthesis is the process where images are created from other image or non-image data. There are two primary forms of image synthesis operations (Baxes, 1994). The first is the reconstruction of any image using multiple projection images. The second form is visualization, where images are created for presentation purposes that may or may not be based on physical objects.

Images of three-dimensional objects can be created from multiple two dimensional images of the same scene. As long as the geometrical relationship between the images is controlled, the depth dimension can be recreated. This type of operation is used to synthesize images showing attributes that are not readily apparent in the original individual images. In particular three dimensional attributes such as depth features can be made more visible.

In this thesis range images of the scene are captured at different views and are merged together by applying rotation and translation techniques to the data, to give complete three dimensional data. Using certain software this three dimensional data can be reconstructed into a complete model.

2.2 Image data handling

Image data handling has two major processes; image display and image storage. Image display is the process to viewing digital image on any display monitor, for this the digital image has to be converted back to an analogue video signal. The image display process is the reverse of the image digitization process. Image display is simple in comparison with the image digitization because of the necessary timing signals are already available from the digitization process.

Image storage is the other critical process during image data handling. In some applications the digital image may be processed in real time (Baxes, 1994). In those cases the image data can flow directly from the digitizer, through the digital image processor and back out to the display function. But most of the applications require storing the image. It is because of reasons like processing is too complex to be carried out in real time or the processing operation requires access to all pixels in the image frame before processing can begin. There are two kinds of storage; the working image storage and permanent image storage. Working image storage is the process wherein images were stored temporarily. These are commonly referred to as image store. These have memory devices which can store the digitized video data produced by the image digitizer. The image data also leaves the image store at the same rate as they were stored to feed the image display circuitry. In permanent image storage, the processed digital image may be required for further processing or need to

be transferred to another digital image processing system so they were stored on permanent digital storage medium. It is necessary to store the digital image in a standardized format before archiving.

2.3 Range imaging

Range images are a special class of digital image. Range images have pixel values that represent the distance from the point of observation to the point on the object. Range images contain information that can be interpreted to give the three dimensional layout of the part of the object visible to the range camera. Range images are also referred to as depth images, depth maps, surface profiles and 2.5D images. In older range image cameras, the pixels had a depth of 8 bits. This was found to be so coarse that quantisation errors had a significant effect on the processing. More recent camera designs have given data with depths of 10, 12 or 16 bits per pixel (Russ, 2002). There are two main methods of collecting range images that are known as time of flight and structured lighting. The time of flight method uses a technique similar to radar. A pulse of laser light is directed to a point on the object. The time taken for the pulse to travel outwards to the object plus that for the reflected pulse to return to the camera is measured. The distance that the pulse travels can then be calculated. Range images can be represented in two basic forms. One is a list of 3D coordinates in a given reference frame (cloud of points), for which no specific order is required. The other is a matrix of depth values of points along the directions of the x, y image axes, which makes spatial organisation explicit.

Range images are acquired with range sensors. Range sensors can be usefully characterised as active and passive range sensors. Active range sensors project energy (e.g. light) on the scene and detect the reflected light to exploit the effect of controlled changes of some sensor parameters. Passive range sensors rely only on naturally present light on the scene.

Active range sensors exploit a variety of physical principles. The most common sensor techniques are triangulation, radar/sonar, moiré interferometry and active focusing/defocusing. Triangulation uses a light projector and an intensity camera,

which is placed at a certain distance from the projector. The projector emits a light pattern. The most common patterns are planes and single beams. We shall use a projected plane for illustration. The intersection of the plane with the scene surface is a planar curve called the strip, which is observed by the camera. By using triangulation we get the depth map of the surface points under the strip. Radar/sonar uses a short electromagnetic or acoustic wave and detect the return (echo) reflected from surrounding surfaces. Distance is obtained as a function of the time taken by the wave to hit a surface and come back. Moiré sensors project two gratings with regularly spaced patterns onto the surface and measure the phase differences of the observed interference pattern. Other phase difference sensors measure the phase shift of the observed return beam. Distance is a function of the phase difference. Active focusing/defocusing sensors use two or more images of the same scene, which are acquired under varying focus settings. Once the best focused image is determined, a model linking focus values and distance yields the distance.

2.4 Geometrical Representation

There are many types of coordinate systems used to represent surface points geometrically in the three dimensional space, but in general the Spherical and Euclidean coordinate systems are satisfactory for the purpose of surface reconstruction. In order to obtain a full geometric description about the scene object, the Euclidian coordinate system is ideally suited in this instance.

2.4.1 Euclidean coordinate

The Euclidean coordinate system locates a surface point $S(x, y, z)$ in three dimensional space spanned by the horizontal x -axis, vertical y -axis and with the z -axis usually pointing towards the camera or image sensor along the optical axis (see Figure 2.2). The coordinates x , y and z may be positive or negative lengths from the origin O . The surface point $S(x, y, z)$ is located at a distance $\|(x, y, z)\|$ from the origin O and pointing in the direction $\hat{S} = (x, y, z) / \|(x, y, z)\|$.

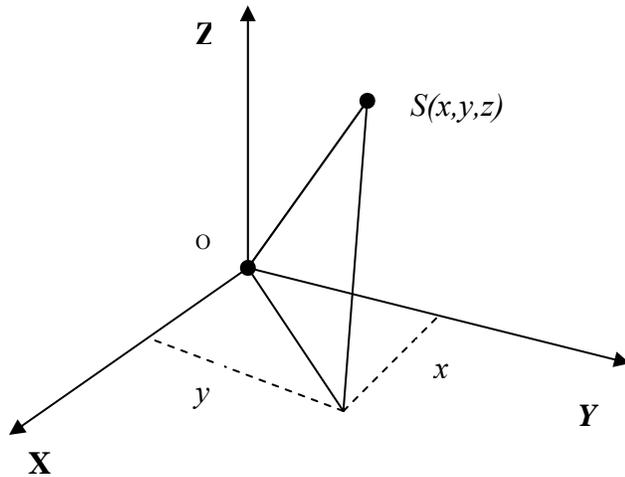


Figure 2.2 : A surface point S in Euclidean coordinate system

2.5 Projections

In computer graphics a projection can be stated as mapping of 3D scene coordinates on to a 2D screen coordinates (Klette *et al.*, 1998). There are two different types of projection used in common practice: the orthographic or parallel projection and the perspective or central projection.

2.5.1 Parallel projections

In the parallel projections the centre of projection is located at infinity and the projected rays are parallel to the optical axis (see Figure 2.3).

The parallel projection is efficient for computational purposes. The image formed is not affected by the distance between the scene and image plane. The projection equation for the image coordinates are given by

$$x = X \text{ and } y = Y, \quad (2.1)$$

where x, y are the pixel coordinates on the image plane and X, Y, Z are the scene coordinates.

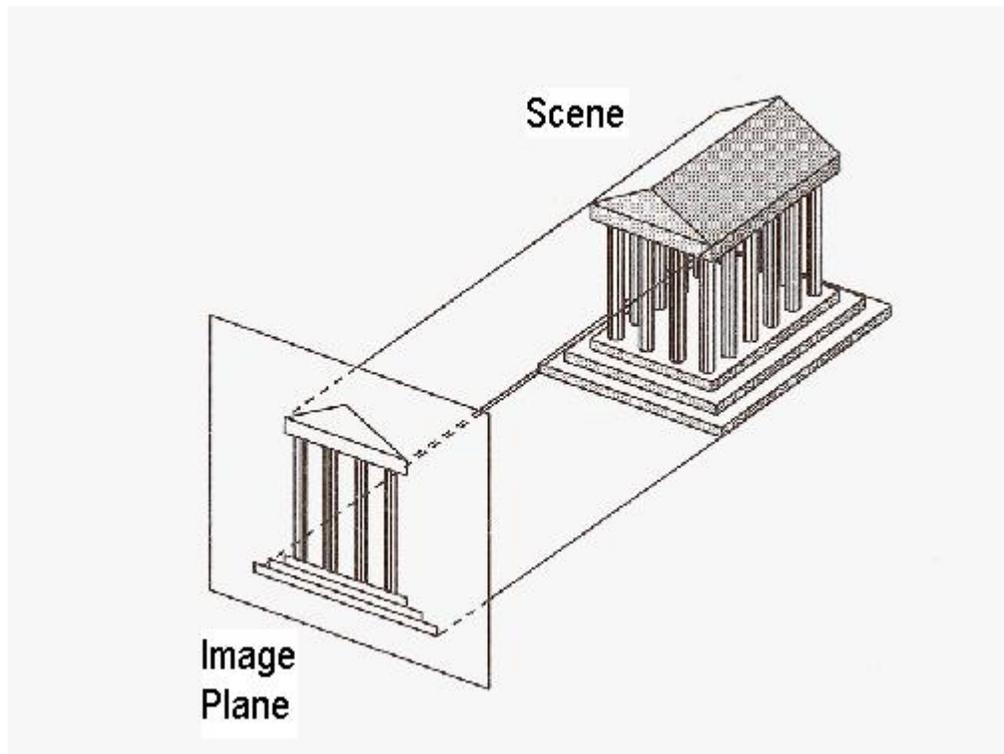


Figure 2.3 : Image captured using parallel projection

[source:<http://www.cs.princeton.edu/courses/archive/fall99/cs426/lectures/view/img016.gif>]

2.5.2 Central projections

In the central projection the centre of projection or focal point is at a finite distance from the image plane. The point on the image that corresponds to a particular point in the scene is found by following the line that passes through the scene point and the centre of projection (see Figure 2.4). In the central projection a camera lens is modelled as a pinhole at the focal point.

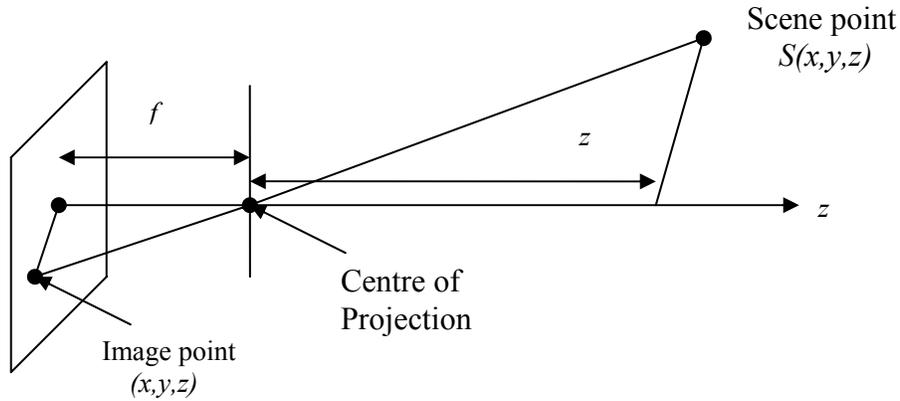


Figure 2.4 : Scene point captured on image plane using central projection

The captured image is inverted with respect to the scene and the size of the image decreases with increase in the distance of the scene from image plane (Klette *et al.*, 1998). The projection equation for the image coordinates are given by

$$x = \frac{fX}{f + Z} \quad \text{and} \quad y = \frac{fY}{f + Z}, \quad (2.2)$$

where x, y are the pixel coordinates and X, Y, Z are the scene coordinates and f is the distance from the image plane to the focal point.

2.6 Coordinate Transformation

A Coordinate transformation is one of the important processes during the reconstruction of surface. The registration phase is done by coordinate transformation, where multiple views are registered to give a complete set of data points for the surface. In generating a view of a scene, they are used to achieve the effect of different viewing positions and directions. The two important terms used in transformation are translation and rotation.

The coordinate transform will be written as

$$\vec{S}' = \text{Rm}\vec{S} + \vec{T} \quad (2.3)$$

where \vec{T} is the translation, Rm is the rotation matrix, \vec{S} is the scene coordinate and \vec{S}' is the transformed coordinate.

Rotation matrix about x axis through an angle θ , is achieved by the following transformation. The transformation matrix affects only the y and z coordinates. Figure 2.5c

$$\text{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & -\sin \theta \\ 0 & \sin \theta & \cos \theta \end{bmatrix} \quad (2.4)$$

Similarly rotation about y axis affects x and z coordinates. Figure 2.5b

$$\text{R}_y = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \quad (2.5)$$

The rotation about z axis affects x and y coordinates and rotation matrix is given by, (see Figure 2.5a)

$$\text{R}_z = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (2.6)$$

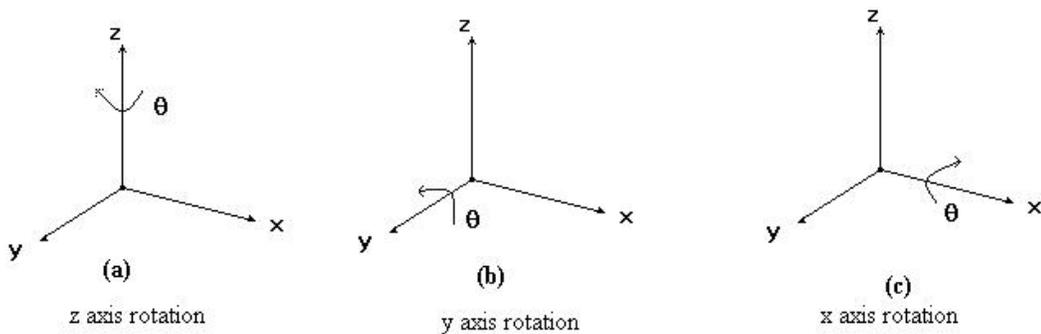


Figure 2.5 : The three three-dimensional rotations.

2.7 Distortions

Distortions due to the use of an imperfect lens cause a change in the shape of the image. Also the range measured with the Waikato Image Ranger for any objects off the optical axis is affected by geometrical distortion. These two major distortions need to be considered that need to be considered.

2.7.1 Radial Distortion

Lens distortions are caused by the use of imperfect lenses. Any surface may appear as a bent surface on the image. The two major radial distortions are barrel distortion and pincushion distortion. Barrel distortion appears with minimal zoom lenses. It causes the image to appear slightly curved outward like a barrel at the edges of the image, whereas for pincushion distortion, edges of the image appear to be bent inwards. Pincushion is caused due to maximum zoom lenses. The pincushion distortion is often less noticeable than barrel but are equally visible near the edges (see Figure 2.6). These distortions can be eliminated easily by systematic means.

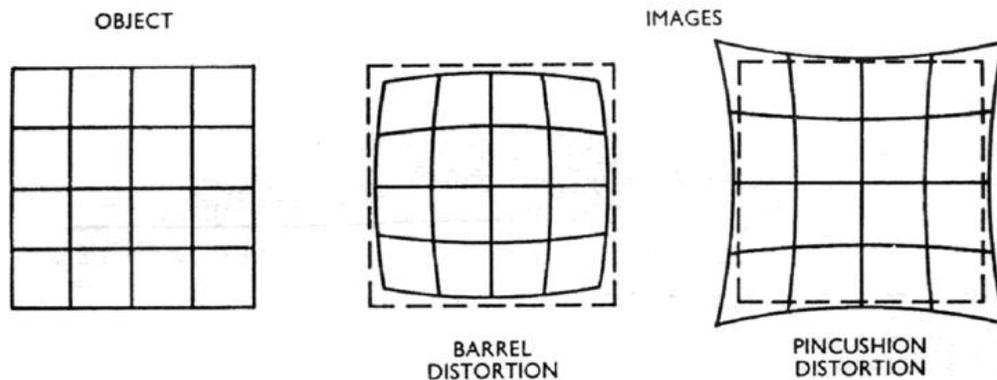


Figure 2.6 : Distortion produced on a checked plane while image acquiring. (Source: www.umich.edu/~lowbrows/guide/opticaljargon.html)

Radial distortions can be calibrated using polynomial approximation methods. Let (X_d, Y_d) be the image coordinates which are distorted, and (X_0, Y_0) be the image centre pixel. The polynomial model for radial distortion can be expressed as

$$D(R_L, K) = \left(1 + K_1 R_L^1 + K_2 R_L^2 + K_3 R_L^3 + \dots + K_n R_L^n\right) \quad (2.7)$$

Where K_n is the distortion coefficient, R_L is the distance between (X_0, Y_0) and (X_d, Y_d) which is given by

$$R_L = \sqrt{(X_d - X_0)^2 + (Y_d - Y_0)^2} \quad (2.8)$$

A third order of distortion polynomial is usually considered sufficient, with higher order terms set to zero. After calculating the distortion polynomial, radial distortion can be eliminated from the distorted image by

$$(X, Y) = (X_d, Y_d) \cdot D. \quad (2.9)$$

2.7.2 Geometrical Distortion

Geometrical distortion occurs in the Waikato range imager when the scene is located away from the optical axis. Figure 2.7 shows two objects located at the same z coordinate (i.e. $z = d$) from the camera, but one object is located on the optical axis and the other off the optical axis. Though both objects have the same z coordinate, the range image shows object 2 as closer and object 1 farther away due to the difference in path length that the light has to travel. This error in the range can be considerably reduced by placing the light source closer to the camera. Ideally the light source should be co-located with the camera. Nevertheless off axis objects still have a greater range due to the increased light path length and a correction must be applied to calculate the correct z coordinate. This correction is dependent on the x , y coordinates as well as the z coordinate of the object.

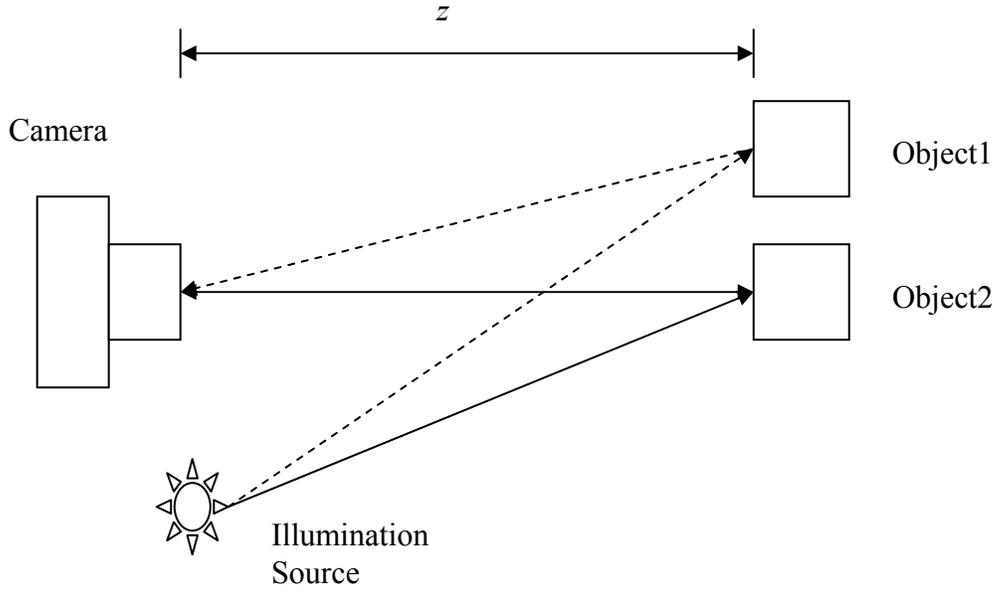


Figure 2.7 : Geometrical distortion resulting from increased path length of off axis objects.

The correction for this distortion can be explained using Pythagoras theorem. Let z be the actual range or distance from the camera to the object1 and recorded range value for the object 1 is d . The error in the range value is given by

$$\varepsilon = \sqrt{z^2 - d^2} \quad (2.10)$$

Thus by adding the error in distance to the range value, the exact range value for the object can be determined. In this research the light source is located close to the camera so the one type of distortion caused due to distance between light source and camera is avoided.

2.8 Background Segmentation

If objects of interest have a significantly different pixel value than that of a background pixel, then we may identify the objects $f(x, y)$ by

$$t_1 < f(x, y) \leq t_2 \quad (2.11)$$

with the threshold values t_1 and t_2 chosen appropriately. If the above is satisfied for a given pixel, then it is part of an object. The remaining pixels make up the background. In many cases, only a single threshold t is needed for background segmentation if a controlled environment is established such as a plain black background. Creating such a controlled environment will improve the segmentation process.

To further improve the efficiency of the segmentation, the background image may be subtracted from the input image before thresholding; provided the environment remains the same while each image is acquired. In many cases, it is useful to assign the background pixels to zero or black and the representing object points are set to 1, thus creating a binary image f_b that is

$$f_b(x, y) = \begin{cases} 1 & : E(x, y) - E_{bg}(x, y) \leq t, \\ 0 & : \text{otherwise} \end{cases} \quad (2.12)$$

where E is the image to threshold and E_{bg} is the background image. This method of background segmentation is useful for locating the object contours.

2.9 3D Shape Reconstruction Methods Reviewed

A surface reconstruction procedure cannot guarantee the recovery of the surface exactly, as there is only certain information about the surface through finite set of sample points. Therefore, the correct reconstruction method depends on the application. Some of the available models for shape reconstructions are discussed below.

Shape from shading, the three-dimensional shape of an object is determined from its irradiances in a single image using its reflection properties (Horn *et al.*, 1989). The irradiance is given

$$I(b) = \rho I_s \cos \phi(b), \quad (2.13)$$

where b denotes a point in the image, $I(b)$ the reflected light intensity observed at b , I_s the illuminant intensity, ρ the constant albedo on the surface and $\phi(b)$ the angle between the light source direction and the surface normal at the three dimensional point on the object surface corresponding to b . The reconstruction problem is to determine the surface shape from the irradiances $I(x, y)$.

Shape from texture is a model in which texture information is calculated first from the image, and then the three dimensional orientation and shape of the scene surface are inferred from the texture distortion (Kender, 1978). *Shape from focus/defocus* uses blurring phenomena to determine the depth of the scene (Aizawa, 1994). There are two approaches to the utilization of the blurring phenomena for shape reconstruction. The first approach is the depth from focus which is a search based method. A sequence of images is taken with a changing focus in small steps and the setting that optimizes image focus is determined. The second approach is depth from defocus, where a couple of images are taken with different optical settings. The depth map is reconstructed by solving the inverse problem of the blurring process based on camera and models edge.

All the above models used intensity images for the shape reconstruction. In intensity images pixel values are related to surface geometry only indirectly. Therefore, the shape is not straight forward to calculate using intensity images. Range images encode the position of the surface directly. Shape reconstruction process that uses range images are straighter forward and tend to produce more accurate results. Reconstruction algorithms using range images can be usefully classified according to the quality of input data: unstructured point cloud and structured point cloud (Fabio, 2003).

Unstructured point cloud algorithms works on unorganized data having no other information on the input data except their spatial position. They do not use any assumption on the object geometry and therefore, before generating mesh surface, these usually structure the points according to their coherence. These algorithms need a good distribution of the input data and if the points are not uniformly distributed they easily fail. The structure point cloud algorithms are based on structured data that can take additional information of the points in to account. Some of the additional information is edge, fold, etc. Apart from the above two major classifications there are other ways to classify shape reconstruction algorithms; algorithms assuming fixed topological types where topological surfaces are assume to be known a prior; approximated surfaces algorithm do not always contains all the points but points as near as possible to them, and interpolated surface algorithms are used when precise

models are required. The conversion of point cloud data in to a consistent meshed surface is generally based on four steps. First the preprocessing stage where erroneous data are eliminated or points are sampled to reduce the sampling time. The second stage is determination of global topology of the object surface. This stage needs global sorting mainly to preserve special features like edges. The third stage is the mesh generation, triangular or tetrahedral meshes are created. Finally post processing stage where editing operations are carried on to perfect the polygonal surface.

2.10 Overview of Registration Process

To create a correct and consistent model, the scans have to be merged in one coordinate system. This process is called registration. Registration plays an important role in three dimensional model acquisitions, object recognition and geometry processing. The most general approach for registration is the Iterative closest point (ICP) algorithm (Huber, 2003). ICP method attempts to register a pair of range images by defining a metric describing the amount of error in a candidate registration. The minimum of this error function indicates the optimal registration, and can be found with standard function optimization techniques. A significant problem in ICP-based methods, however, is their reliance on shared points in the two range images. That is, these algorithms expect a significant number of points in one range image to also be seen in the other image. Let $P = \{p_1, p_2, \dots, p_n\}$ and $Q = \{q_1, q_2, \dots, q_n\}$ be two point cloud data. The main aim of the registration algorithm is to find a rigid body transform α composed of rotation matrix R and translation vector t that best aligns the two set of point cloud data. Registration algorithms based on Iterative closest point (ICP) works as, known initial position of the data with respect to the model, the algorithm chooses a set of k point pairs (p_n, q_n) from the two set of data. The distance between the two sets of data is approximated by the sum of the distances between the point pairs. The algorithm searches for the rigid transform that minimizes the residual distance, ξ , between the given data and the transformed data:

$$\xi(\alpha) = \sum_{i=1}^k d^2(\alpha(q_i), p_i), \quad (2.14)$$

where d can be point to point distance. However the basic assumption is that the sum of squared distances between pairs of points is a good approximation for the distances between two point cloud data.

Mitra *et al.*, (2004) proposed a framework for pairwise registration of shapes represented by point cloud data is one of the recent approaches for registration. To show theoretical bounds on convergence behaviour of registration algorithms based on their distance function, they pose the problem of registration of two point clouds as an optimization problem over the space of rigid transforms. This leads to the following optimization problem for the search of best rigid transform $\alpha = (R, t)$ that minimizes the error given by

$$\xi(\alpha) = \sum_{i=1}^m d^2(Rq_i + t, \alpha_p), \quad (2.15)$$

where R is the rotation matrix and t is the translation vector. The function $d^2(Rq_i + t, \alpha_p)$ is the distance from the transformed data point q_i to the surface represented by point cloud data P . By providing the optimization problem to the above equation they found convergence behaviour to be dependent on the distance function d^2 . The advantage of this algorithm is that it has more stable convergence behaviour when the initial distance between two set of point cloud data is large.

Fully automatic registration was proposed by Huber and Hebert (2003). Their process is also called multi view surface matching. The problem they solved is, given an unordered set of overlapping three dimensional views of a static scene and no additional information, automatically recover the viewpoints from the views themselves, thereby registering the views in a common coordinate system. The algorithm begins by converting the input point data in to a surface mesh. There are two main phases in this algorithm registration and integration. In registration phase different views were aligned on to a common coordinate. In the integration phase obtained views were combined in to single surface. The three problems occurring during the registration phase are overlapping scenes, transform between each overlaps and finally position of all views in global coordinate. Initially surface matching

involves the conversion of input point cloud data in to surface mesh. Then model construction is carried on from global coordinate measure. Local surface consistency can be achieved by having a smaller value for the overlap distance, a general measure that applies to any pair of surfaces and two measures based on visibility consistency that are tailored to surface derived from range images. Later global surface consistency was achieved if every pair of view is locally surface consistent. The main advantage of this algorithm is that no additional input like the position of the scene, needs to be given for the registration process to proceed.

3 Equipment and hardware

In this chapter the working principle of the Waikato Range Imager is discussed initially which is followed by the components and hardware used in the ranger system and finally about the precision of the ranger data captured.

3.1 Image Capturing

The process of capturing range information encoded in the image is range imaging. Systems for range imaging can be classified into four categories: laser scanning, stereo vision, structured light and imaging lidar. Each method has its advantages and disadvantages. The imaging lidar system will be discussed in detail as it is the principle used in The Waikato Range Imager.

3.1.1 Imaging Lidar

A modulated light source illuminates the scene. The reflected light is captured with a camera, after passing through a high speed shutter with the same frequency as that of the light source. The received light at the shutter has a phase delay that is due to the time of flight of the light. The major difference between imaging lidar systems arise in the modulation control signals (Dorrington *et al.*, 2006). The three major modulation control signals used in general are pulsed systems, homodyne systems and heterodyne systems. In pulsed systems, both the illumination and the high speed shutter are controlled with a single pulse in the nano-seconds region. Consider the situation where both the illumination and shutter are pulsed simultaneously. Then the scattered illumination light pulse arriving at the camera is delayed due to propagation delay. Light scattered from a distant object arrives later and will be cut off by the shutter resulting in a darker image than for a closer object. Homodyne systems are in principle similar to pulsed systems, but the illumination and shutter modulation signals are a continuous square, sinusoidal or triangle wave generally in the 10 to 100 MHz region. The modulation signal of the light entering the shutter can be considered

as being mixed with the shutter modulation signal, resulting in the objects range being encoded as intensity of light. Some decoding is required to derive actual range values from brightness, and often multiple measurements are performed at different relative phase of the modulation signals to perform quadrature type decoding.

In heterodyne systems, the modulated frequency applied to the light source and shutter differs slightly in frequency (see Figure 3.1). This produces an offset frequency (beat frequency) at the camera which is equal to the difference in the modulation frequencies (Dorrington *et al.*, 2006). Range information for each pixel can be determined by acquiring a video sequence of the scene and calculating the beat signal phase for each pixel. These systems have high precision in range because the range is determined from time varying intensity. Moreover the range determination is not directly affected by background lighting. The Waikato ranger system used in this study is a heterodyne imaging system.

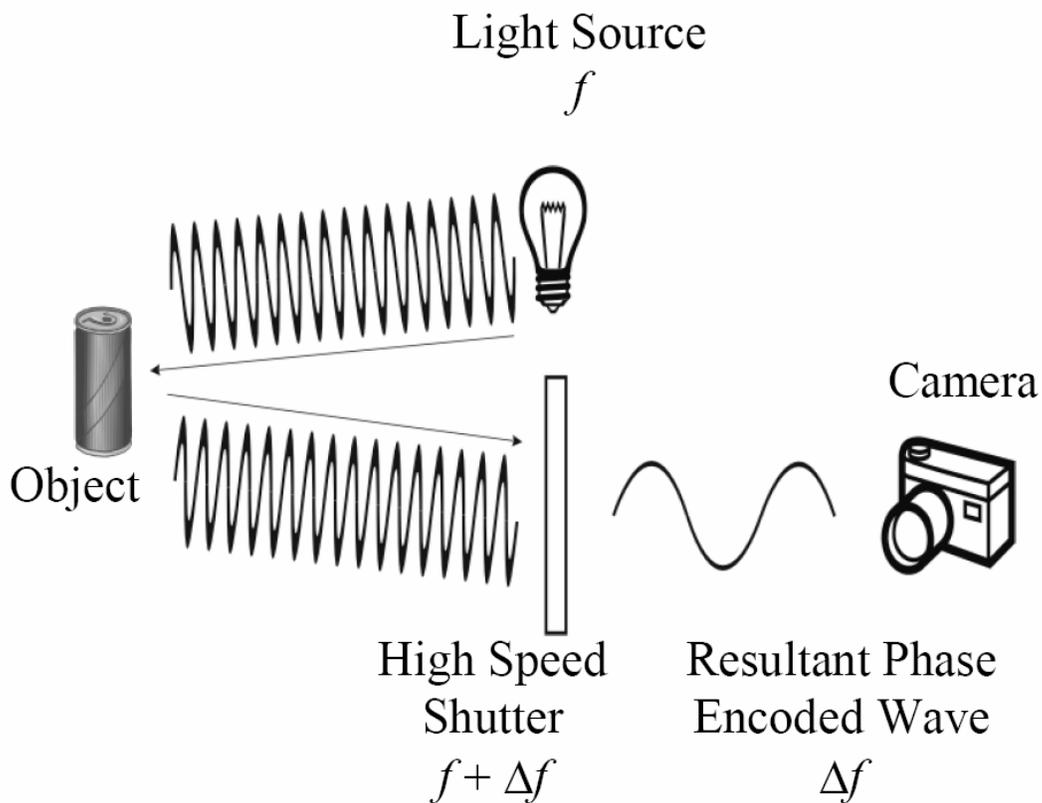


Figure 3.1 : Heterodyne imaging lidar systems (Source: Payne *et al.*, 2006)

Digital cameras are used to capture the flashing beat signal in heterodyne systems. The beat signal must be limited to the tens or singles of hertz to stay within the Nyquist sampling criteria because of typical camera technology. Stability is generally achieved by using frequency locked signal generators. By generating a camera synchronization signal from the same set of frequency locked signal generators, the sampling video frame clock is synchronised. The beat signal lies exactly on a Fourier transform bin allowing simplification of the phase determination process and high speed processing.

3.1.2 Image Intensifier

In the Waikato Range Imager, image intensifier is used as a high speed shutter. An image intensifier is the device which amplifies any light that is incident on it from the scene. Light focussed on the photocathode is converted into photoelectrons (see figure 3.2). The number of electrons emitted at this point is proportional to the number of input photons. These electrons are accelerated by the voltage applied between the photocathode and micro channel plate (MCP), which then enter individual channels of the MCP. Each channel of the MCP acts as an independent electron multiplier, the input electrons hitting the channel will produce secondary electrons. This is repeated by the potential gradient across both ends of the MCP and the number of electrons is amplified by up to four orders of magnitude. The electrons are accelerated with an electric field, impinging on a phosphor screen where they are converted back into photons and can be measured with a digital video camera.

Continuous modulation is preferred for the application to maximise the light collected. Pulses are used with a voltage of +50 V (off) to -200 V (on) at low repetition rates, however continuous modulation at these voltages is impractical due to heat dissipated within the intensifier (Payne *et al.*, 2006). To reduce this power dissipation, the photocathode voltage can be lowered, although this reduces the gain and defocuses the image as the electric field is responsible for focussing the emitted electrons on to the MCP input. So -80 V is used with minimal loss of focus. The output intensity using -80 V is reduced by approximately half to that of -180 V, but as the CCD is integrating over time for this application, two successive pulses

achieve the same final intensity. The power dissipated within the intensifier during two -80 V pulses is less than 20% of a single -180 V pulse.

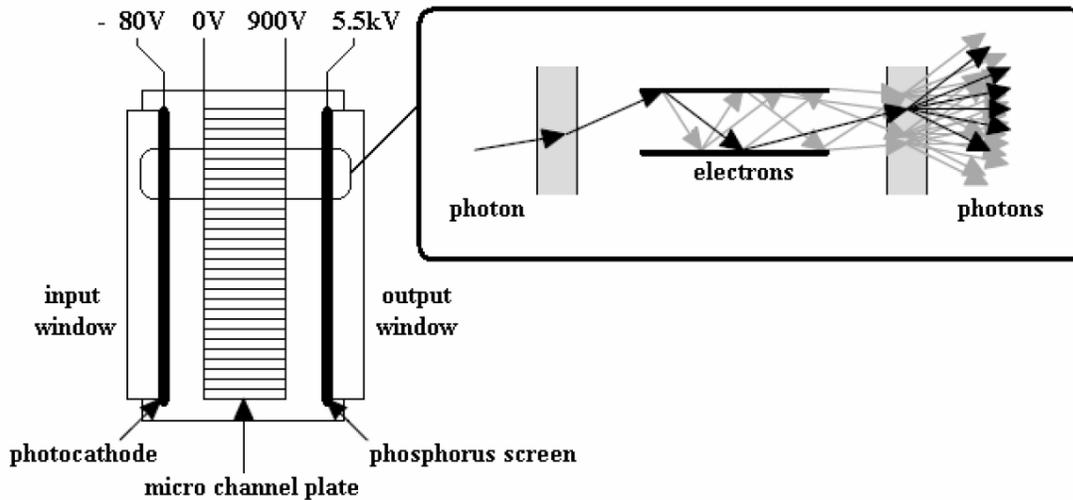


Figure 3.2 : Image intensifier depicting how a single photon is multiplied to many due to electron acceleration (Source: Payne *et al.*, 2006)

3.1.3 Signal Generator

Direct Digital Synthesis (DDS) is used in the signal generator because of its high stability and frequency tuning. A DDS uses logic and memory circuits to digitally give the expected output signal and a data conversion device to convert it from a digital to an analogue signal. Major advantages of DDS are fast switching, excellent phase noise and phase continuous frequency changes (Payne *et al.*, 2006). DDS have three major components, the phase accumulator, a mapping device and a DAC (digital-to-analogue converter). The output of the phase accumulator is a correlation between the desired frequency and the clock, in the form of a phase ramp. This signal is passed through a sine wave lookup table to the DAC. By changing the step size through the lookup table the output can be precisely adjusted in very small increments from DC to the Nyquist frequency. Three outputs from the DDS are given as input to the light source, the image intensifier and third to synchronise the camera frame trigger with the rest of the system.

3.1.4 Ranger Camera

The camera used is the Dalsa Pantera TF 1M60 which offers 60 frames per second, 1024×1024 pixel resolution and high sensitivity 12-bit images. The low noise, digitised video signal makes the camera useful for capturing low contrast images. The camera frame trigger is synchronised to the system through the direct digital synthesis board (DDS), allowing the scene to be recorded at an exact multiple of the low frequency input light signal. Simple Fourier analysis of the known frequency bin provides the phase measurement of each pixel (without any spectral leakage occurring), from which the range is calculated. For the studies described herein the camera is running in the 2×2 pixel binning mode giving better sensitivity with 512×512 pixel resolution. In this mode the camera can run at 100 frames per second.

3.1.5 CCD image Sensor

Modern cameras suitable for image processing are often based on CCD semiconductor sensors and CMOS sensors. The interline transfer architecture and the frame transfer architecture are the two major concepts for CCD chips design. They differ in the way how the charge is transferred out of the light sensitive sensor elements. An interline transfer sensor contains light sensitive sensor elements that are arranged in columns. Every column is connected to an adjacent vertical shift register through a transfer gate. The stored charge from light sensitive model in vertical shift register is shifted to the horizontal readout register. In a frame transfer sensor the light sensitive elements and the shift registers are not arranged in column but in two separate grids. The advantage of the frame transfer is that the light sensitive elements are also horizontally adjacent to each other and hence less image distortion arise. The CCD image sensor design used in this research is frame transfer architecture. The sensor used in this is a monochrome progressive scan frame transfer image sensor which offers 1024 × 1024 pixels up to 60 frames per second (see Figure 3.3). The

combination of high speed and a high linear dynamic range makes this device the perfect solution for high-end real time application.

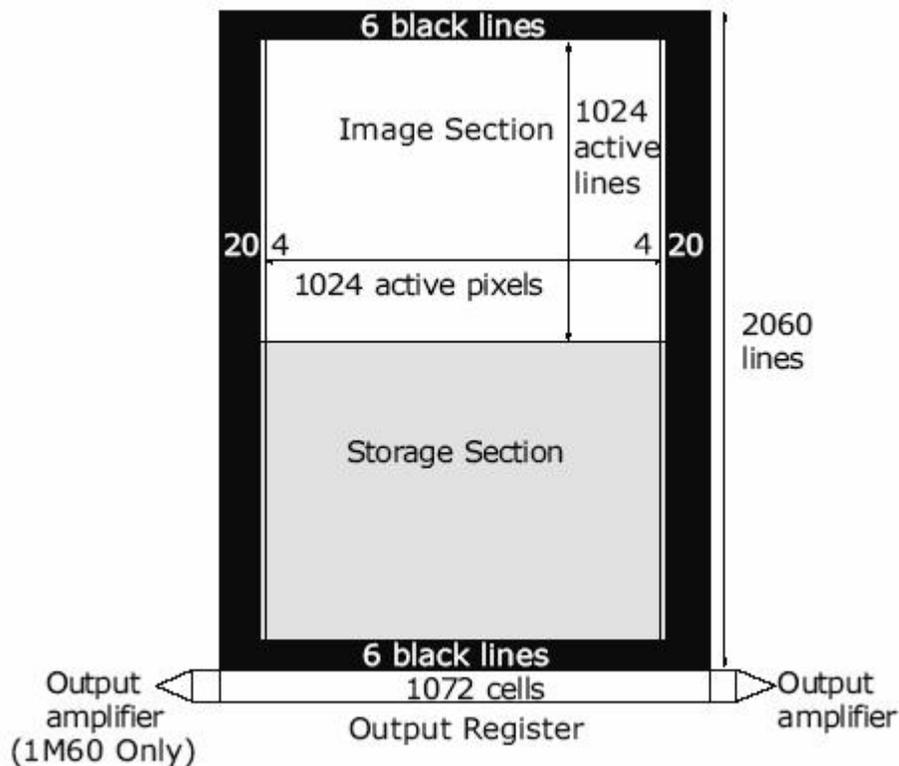


Figure 3.3 : CCD used in Ranger camera

3.1.6 Light Source

A light source with a modulation bandwidth of 100 MHz is required to illuminate the entire field of view. Since LED's can be modulated to only tens of MHz, laser diodes are preferred for the bandwidth required here. Moreover range precision is proportional to phase determination and modulation wavelength, so by increasing the modulation frequency, range precision can be increased. The laser diode used in the Waikato ranger is the Mitsubishi ML120G21. Its operating wavelength is 658 nm which is not optimal for image intensifier but the low cost and high power makes it more advantageous. A laser which operates at 155 MHz is utilised to provide a constant bias current up to 150 mA in addition to a pulsed current up to 700 mA.

3.1.7 Grabber

Frame Grabbers are image processing computer boards that capture and store image data for certain applications such as computer vision. For frame grabbers that can handle camera outputs in digital format the input pixel acquisition depth is important to consider. Increasing the pixel depth increases the amount of detail that will be reproduced in the scanned image.

Common acquisition features for frame grabbers include digitization accuracy, number of video inputs, number of input lookup tables, on-board video memory, overlay buffer memory, trigger input, high resolution cameras, gain scaling, and offset scaling. Signal-to-Noise ratio is defined as the peak-to-peak camera signal output current to the RMS noise in the output current. This ratio represents how prevalent the noise component of a signal, and thus the image uncertainty, is in the total signal. Noise sources include sensor "dark current", electromagnetic interference, and any other spurious non-image signal elements. Higher SNR numbers represent less image degradation from noise. The number of video inputs is the number of video signals the frame grabber can handle simultaneously. High-resolution cameras support inputs greater than 1000 x 1000 pixels. An important environmental parameter to consider is the operating temperature.

3.2 Ranger Precision

The Waikato range imager used in this study has a precision of 1mm in the acquired range image (Dorrington *et al.*, 2006). The main factors that affect the precision in the ranger system are signal-to-noise ratio and modulation frequency used in the light source. The Waikato range imager has a better signal to noise ratio because of the following; image intensifier which can provide higher intensities, camera sensitivity because of larger pixels, noise performance in the camera itself and higher illumination drive. The ranger precision is directly proportional to the modulated frequency so higher the frequency better the precision will be (Dorrington *et al.*, 2006).

The values used in the experimental setup are explained in the following chapter.

4 Experimental Methodologies

In this chapter it is discussed in detail about the FastRBF tool which is used for the interpolation, then the simulation study and finally the experimental work using the range data obtained from the Waikato Range Imager.

4.1 FastRBF tool

Fast RBF toolbox is one of the applications wherein the interpolation of surface points can be done. The Fast RBF toolbox allows scattered three dimensional data sets to be described by a single mathematical function, a Radial Basis Function (RBF). The RBF is a function of the form

$$F_r(x) = p_y(x) + \sum_{i=1}^N \alpha_i \Psi(x - x_i) \quad , \quad (4.1)$$

where, F_r is the radial basis function, x is the surface point, p_y is a low degree polynomial, α_i are the RBF coefficient, Ψ is the real valued basis function and the x_i are the RBF centres.

Some of the most widely used basic functions are as follows: The thin-plate spline function given by $\Psi(x - x_i) = (x - x_i)^2 \log(x - x_i)$ is used for fitting smooth functions of two variables. The multi quadratic function given by $\Psi(x - x_i) = \sqrt{(x - x_i)^2 + c^2}$ is used in fitting topographical data.

Given a set of N points x_i and values g_i , the process of finding an interpolating RBF, F_r , such that,

$$F_r(x) = g_i \quad i = 1, 2, \dots, N \quad (4.2)$$

is called fitting.

FastRBF uses fast approximation methods to fit RBF and to evaluate RBF. This means that true equality at the interpolation nodes is never achieved. The maximum difference between the fitted RBF value and the given values,

$$\max_{i=1, \dots, N} |F_r(x_i) - g_i|, \quad (4.3)$$

is called the fitting accuracy. Given a surface point x , calculating the value $F_r(x)$ is called evaluating the RBF. If w_i is the approximate values of the RBF at the point x_i , then evaluation accuracy is the value which is given by

$$\max_{i=1,\dots,N} |F_r(x_i) - w_i|. \quad (4.4)$$

Generally the evaluation accuracy should be higher than the fitting accuracy (Fairfield Technology, 2001). Figure 4.1 illustrates how best the FastRBF can interpolate data. It also shows the fitting accuracy and evaluated accuracy.

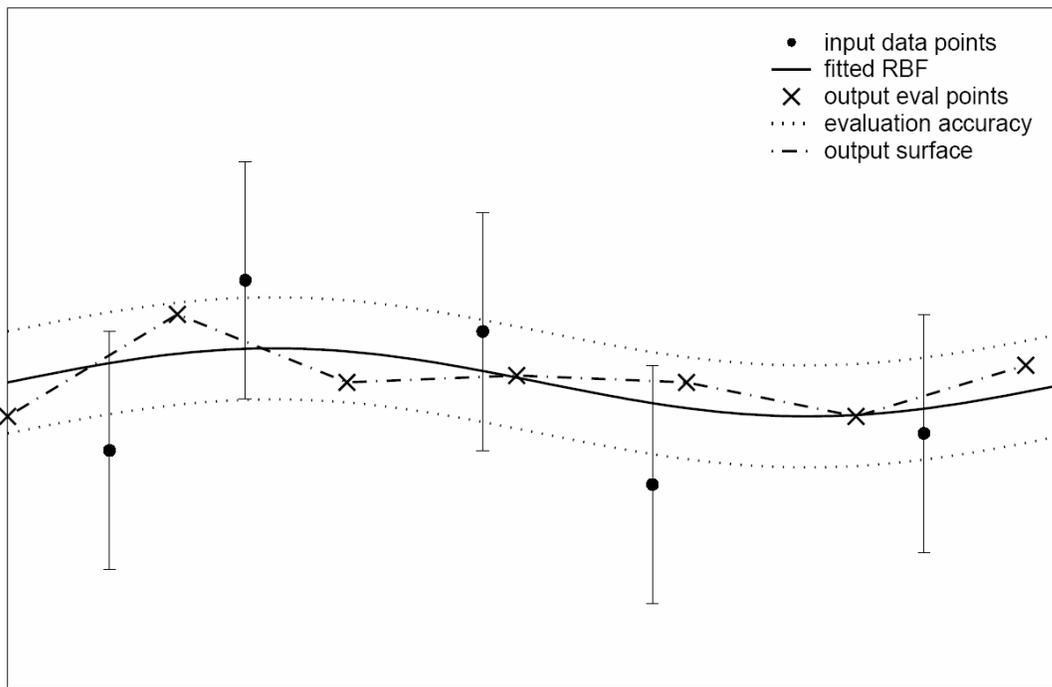


Figure 4.1 : Plot showing Fitter accuracy and evaluation accuracy. (Source : Fairfield technology)

In general fitting an RBF requires $E(N^2/2)$ storage and $E(N^3)$ computation, where N is the number of data points to be interpolated. With the FastRBF these requirements are substantially reduced, so the storage requires $E(N)$ and $E(N \log N)$ for the computation.

A set of values g_i at points x_i where all the x_i are used as RBF centres is interpolated and the resulting equations are solved for the RBF coefficients. In FastRBF this type of standard fit is called a direct fit. FastRBF provides a greedy

algorithm for fitting RBFs which reduces the number of centres used to interpolate the data which in turn results in even faster fitting and evaluation times. Centre reduction is most suited to data that have significant regions where the data is smooth compared to the sampling density. This is not least square fitting but it is fitting so that the infinity norm error at the input points is less than a specified accuracy.

Aliasing is the undersampling of signal that causes artefacts. The minimization of the appearance of jagged edges of polygons or lines during the visualization is called anti-aliasing. FastRBF provides a low-pass filter option in all routines that evaluate RBFs. Data smoothing is a process where unwanted signal or noise is removed. This is achieved during evaluation or by approximating the data when fitting the RBF. FastRBF offers two methods for fitting an approximating RBF: spline smoothing and error-bar fitting. The error-bar fitter fits a smooth RBF to the data. The details lost in the raw data can not be recovered from the RBF during evaluation. This method requires an estimate of the magnitude of noise present at each data point. Spline smoothing is traditional RBF smoothing technique. It also approximates the raw data and is applied during the fitting process. Though it can produce good results choosing the smoothing threshold value can be difficult.

4.1.1 Fitting RBF to surface data

This section explains the process of fitting a three dimensional RBF to a set of N points that lie on a surface. When a surface normal n exists for a surface point S , off-surface points S_o outward and S_i inward are generated by $S_o = S + \gamma n$ and $S_i = S - \gamma n$, where γ and ν are the projection distances of the off-surface points. When normals on both sides are used, every normal generates two off-surface points. This can increase the size of the density data set to three times the original number of points which is not desirable and the default is 1.0, which means that no more than N off surface points will be added to surface points. Figure 4.2 demonstrates this parameter

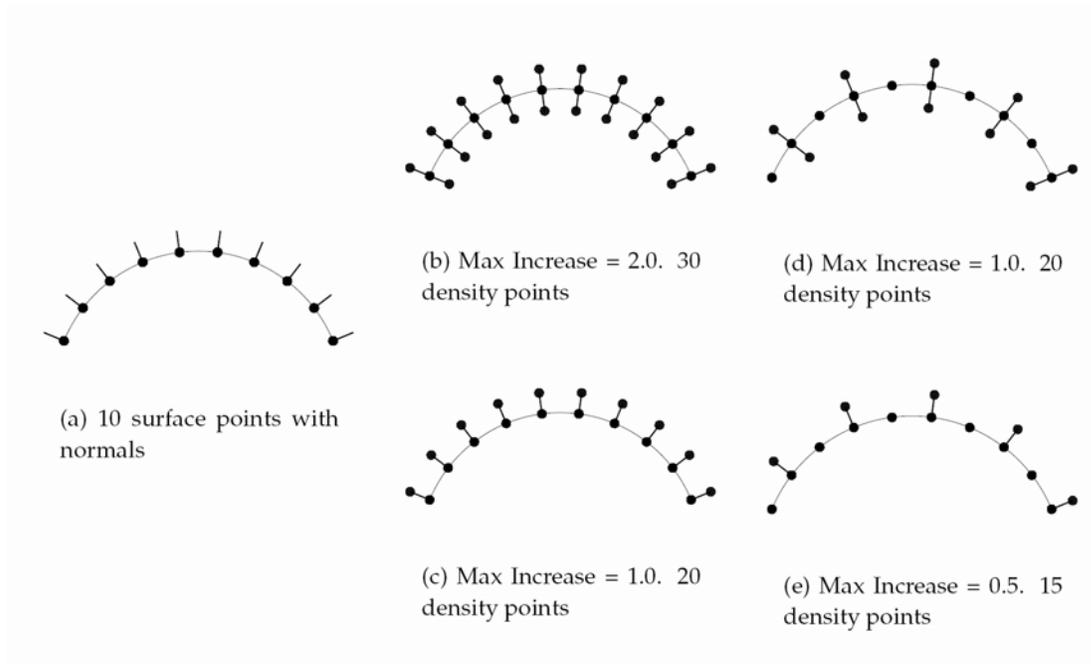


Figure 4.2 : Creating density data from surface normals (Source: Farfield technology)

The consistency of the data is done by validating projection distance along normals. When validating the projection distance of noisy data, undesirably short normals may occur. This can cause curves on the surface and reduce the efficiency of data smoothing when fitting accuracy is less than the noise level. In some case density values were validated because of the bad choice of projection distance. By knowing the noise, errorbar option can be used with a fitting accuracy at the level of noise. If fitting accuracy is too close then having a minimum projection distance around twice the expected level of noise can improve surface smoothness.

4.1.2 Mesh Generation

The final stages of reconstruction is mesh generation for the given data points. The process of reconstructing an isosurface using mesh of polygons is called isosurfacing. FastRBF has two isosurfacing functions: Marching cube and surface following. Both use the fitted RBF at regular intervals to construct a manifold mesh of polygons representing the desired isosurface at a specified resolution. Surface following

algorithm which is particularly suited for isosurfacing signed distance data as in our study will be discussed here. This algorithm minimizes the number of evaluation of density function s by only evaluating s near the surface. Figure 4.3 shows a comparison of RBF evaluations using both algorithms

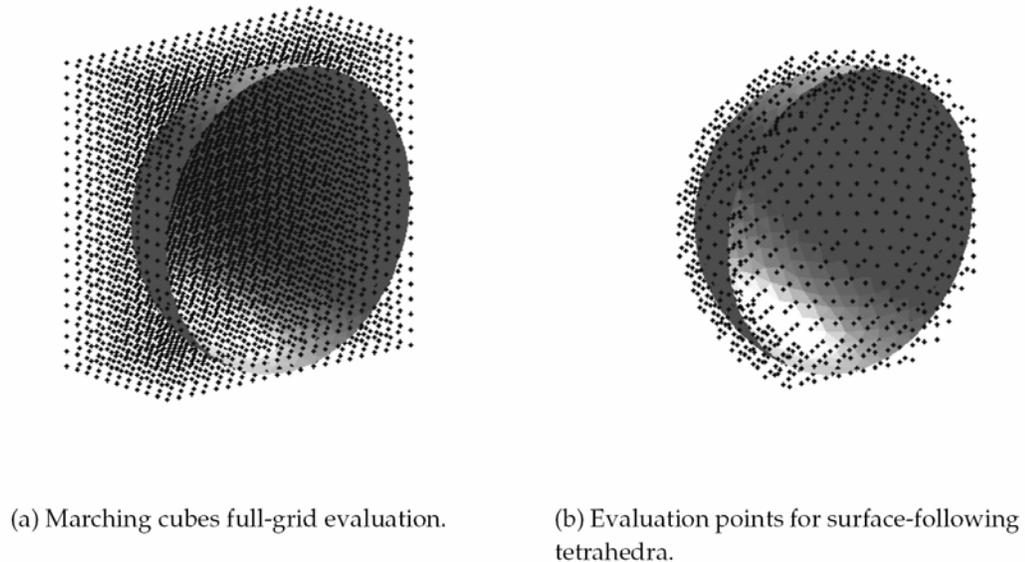


Figure 4.3 : Comparison of RBF evaluations (Source : Fairfield technology)

In general for the density data, the isosurface of interest may lie far from the RBF centres and even the original data points. This may result in not detecting all isosurface components (Fairfield Technologies, 2002). For visualizing density data, initially grid evaluation followed by marching cubes algorithm will produce good results. Mesh optimization improves the aspect ratio of triangular facets and minimizes the number of faces produced. Surface topology is always preserved when mesh optimization is used. Figure 4.4 shows optimization in three dimensional data.

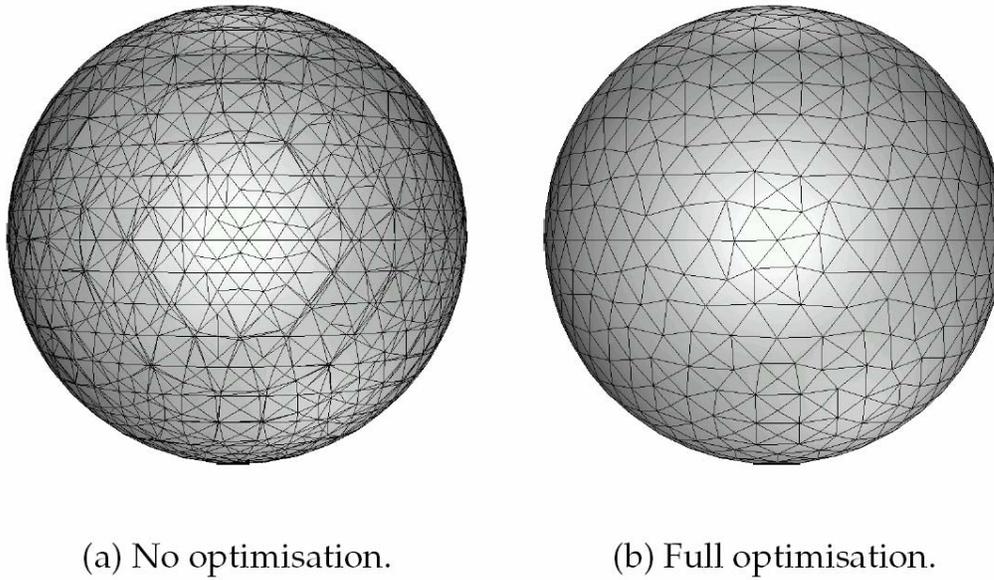


Figure 4.4 : The results for different mesh optimisation methods (Source: Farfield technology)

4.1.3 Using FastRBF

For a given point cloud data five major steps are carried in FastRBF to give reconstructed surface. In the first step inputting the data to the FastRBF is done, for which a structure is defined with x , y and z coordinates of the scene in desired units. The length of the point cloud and the origin of the scene coordinate are also defined in the structure. The second step, calculates the surface normal for the input by fitting a plane to a subset of points in the neighborhood of each point. The minimum normal length should be about the level of noise in the range data. The FastRBF function which performs this is *fastrbf_normalsfromscan*. The third step is to create a distance-to-surface density data for the RBF fitter. The ambiguity in the sign of the normal is resolved by using the position of the camera when the point was measured. The FastRBF function which performs this is *fastrbf_densityfromnormals*. The fourth step is to perform RBF fit for the density data by defining certain accuracy values. The FastRBF function which performs this is *fastrbf_fit*. The Final step is isosurfacing where the fitted RBF data is isosurfaced by creating a mesh. This gives a reconstructed surface for the given point cloud data. The FastRBF function which performs this is *fastrbf_isosurf*.

4.2 Simulation Work

Reconstruction of a three dimensional surface using multiple range images of different views is the aim of this research. Simulation of the range data was carried on as an initial step for this research. The main aim of the simulation work is to generate different range data of a virtual object from different views. This enables reconstruction of precise data without noise and comparison of the reconstruction against the known scene.

4.2.1 Principle

Ray tracing is the basic principle used in this simulation for generating range images. A point ray from predefined image plane is projected on to a finite plane and the intersection point of the ray with the plane is determined. Then the distance between the ray origin point and intersection point on the plane is calculated. The calculated distance or range value is recorded as z coordinate value for the image plane. The same is repeated for other planes as the scene in the simulation is made of planes.

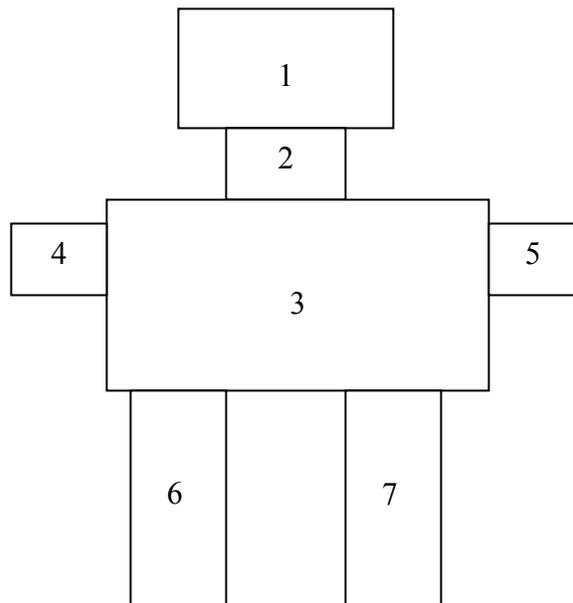


Figure 4.5 : Scene used for simulation work

4.2.2 Setup Description

Scene description

A scene is located at a certain distance from the camera. The scene used in this simulation is a multiple number of cuboids stacked on each other resembling the shape of a robot (see Figure 4.5). The cuboid shape is chosen because of its six finite sized rectangular planes. The centre for the whole scene is chosen exactly at the location where the optical axis passes through. This condition is maintained because of the three dimensional rotation transformations, where axis of rotation and coordinate centre is required (see Figure 4.6). The measurements for defining the scene used in this simulated study are given in the Table 1. The cuboid numbers correspond to those shown in figure 4.5.

Table 1 : Scene coordinate values used in the simulated robot scene

Description	X max (Units)	X min (Units)	Y max (Units)	Y min (Units)	Z max (Units)	Z min (Units)
Cuboid 1	250	-250	400	200	2900	3300
Cuboid 2	150	-150	200	100	3000	3200
Cuboid 3	300	-300	100	-150	2900	3300
Cuboid 4	400	300	100	25	3000	3200
Cuboid 5	-400	-300	100	25	3000	3200
Cuboid 6	250	150	-400	-150	3000	3200
Cuboid 7	-250	-150	-400	-150	3000	3200

Camera Description

An image plane is defined with an $N \times N$ array of points whose origin $(0, 0)$ is located at its centre and lies on the optical axis as shown in figure 4.6. The image plane used in the simulated study consisted of 101×101 and 512×512 pixels. In three dimensional coordinates the image plane can be located as the plane lying on the origin of the z coordinate. The camera assumed in this simulation work is a pinhole camera, with a centre of projection at (x_p, y_p, z_p) , that is, on the optical axis at a distance z_p from the image plane.

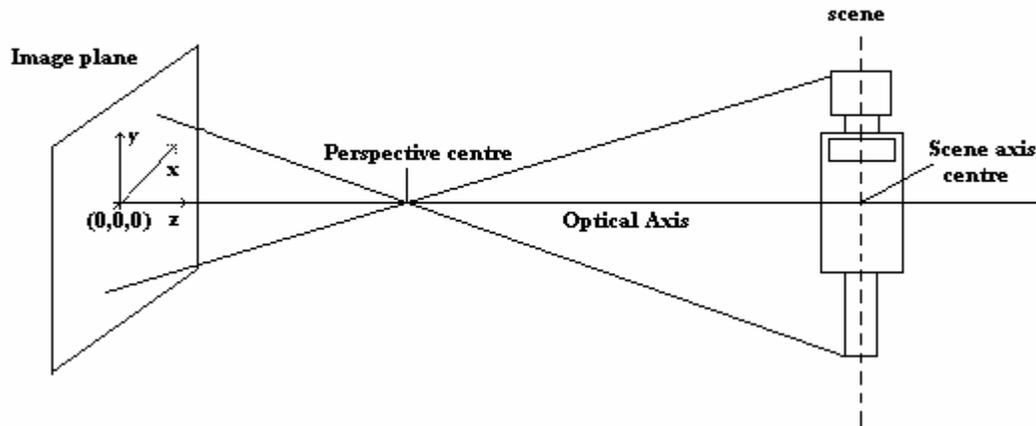


Figure 4.6: Simulation setup showing image plane, centre of projection and the scene.

4.2.3 Procedure

We now describe how the range values are determined for the scene. The first step is finding the intersection point of a ray on to a finite plane of the scene. The second step is to check whether the point of intersection lies on the finite plane and calculate the distance between the ray origin and intersection point. The third step is to check whether the ray intersects any other planes and if so the closest intersection point is taken. This procedure is repeated for each pixel of the image plane to make a complete range view. Multiple range views are required so the scene is rotated by certain angles and new range images are generated.

First step:

Let $\vec{R}_0 = (x_0, y_0, z_0)$ be the point on the image plane (i.e. the location of the pixel) for which a range value is to be calculated and $\vec{R}_p = (x_p, y_p, z_p)$ be the perspective centre. The direction vector \vec{R}_d for the ray is given by $\vec{R}_d = \vec{R}_p - \vec{R}_0$ (see Figure 4.7). Let $\vec{P}_1, \vec{P}_2, \vec{P}_3$ and \vec{P}_4 be the vertices defining a finite plane in the scene

with (x_l, y_l, z_l) the coordinates of \vec{P}_l and so on. The two adjacent vectors for the polygon are given by $\vec{v}_1 = \vec{P}_2 - \vec{P}_1$ and $\vec{v}_2 = \vec{P}_4 - \vec{P}_1$.

There are several ways to find out if a ray intersects with a polygon or not. We use a general method of calculation as follows. The general of a plane

$$\vec{A} \cdot \vec{x} + D = 0 \quad (4.5)$$

where $\vec{A} = (A_x, A_y, A_z)$ is the normal to the plane, D is the distance of the plane to the origin and any \vec{x} satisfying the equation 4.7 must be a point on the infinitely extending plane. It is useful to normalize equation 4.7 to be

$$\frac{1}{D} \vec{A} \cdot \vec{x} = -1. \quad (4.6)$$

The planes that make up the simulation scene are given by the four points $\vec{P}_1, \vec{P}_2, \vec{P}_3$ and \vec{P}_4 . To test whether a ray intersects the plane we need to calculate \vec{A} and D for the plane from the $\vec{P}_1, \vec{P}_2, \vec{P}_3$ and \vec{P}_4 . Substitution of \vec{P}_1, \vec{P}_2 and \vec{P}_3 in to the equation 4.6 gives three constraints which when solved

$$A_x = y_1(z_2 - z_3) + y_2(z_3 - z_1) + y_3(z_1 - z_2) \quad (4.7)$$

$$A_y = z_1(x_2 - x_3) + z_2(x_3 - x_1) + z_3(x_1 - x_2) \quad (4.8)$$

$$A_z = x_1(y_2 - y_3) + x_2(y_3 - y_1) + x_3(y_1 - y_2) \quad (4.9)$$

$$D = -x_1(y_2 z_3 - y_3 z_2) + x_2(y_3 z_1 - y_1 z_3) + x_3(y_1 z_2 - y_2 z_1) \quad (4.10)$$

After calculating \vec{A} and D, point \vec{P}_4 can be substituted in to equation 4.6 to verify that the four points do define the four corners of a plane of finite extent.

The next thing is to determine whether a ray from a particular image plane pixel intersects a finite plane in the scene. A ray parametrically defined by

$$\vec{R} = \vec{R}_0 + (\vec{R}_d t). \quad (4.11)$$

We substitute \vec{R} in to the plane equation (4.5) giving

$$\vec{A} \cdot \vec{R}_0 + t \vec{A} \cdot \vec{R}_d + D = 0, \quad (4.12)$$

and solving for t,

$$t = -(\vec{A} \cdot \vec{R}_0 + D) / (\vec{A} \cdot \vec{R}_d). \quad (4.13)$$

Substitution of t back into equation 4.11 gives \vec{R} , the intersection point of the ray with the plane as,

$$\vec{R} = R_0 - \frac{(\vec{A} \cdot \vec{R}_0 + D)}{\vec{A} \cdot \vec{R}_d} \vec{R}_d. \quad (4.14)$$

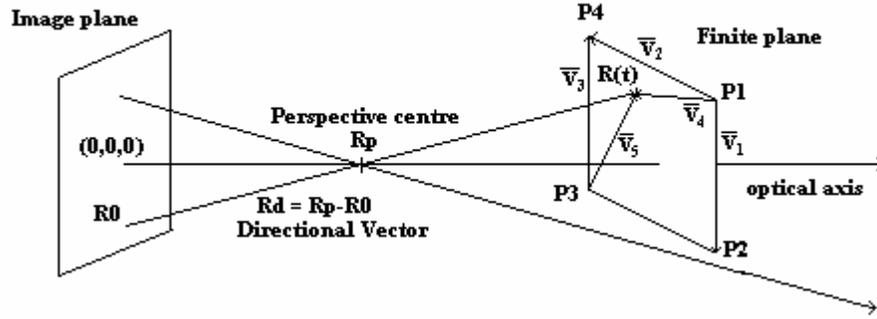


Figure 4.7 : Ray tracing - Intersection point of the ray originating at R_0 with directional vector R_d on the finite plane at $R(t)$

Second step

The second step is to check whether the intersection point \vec{R} is within the finite extent of the plane. Let $\vec{v}_3 = \vec{P}_4 - \vec{P}_3$ be the vector of the boundary of the polygon on the opposite side from \vec{v}_1 . Let \vec{v}_4 and \vec{v}_5 be the vector from \vec{P}_1 and \vec{P}_3 to the point of intersection \vec{R} ,

$$\vec{v}_4 = \vec{P}_1 - \vec{R} \text{ and } \vec{v}_5 = \vec{R} - \vec{P}_3. \quad (4.15)$$

In the following we assume that the vectors $\vec{v}_1, \vec{v}_3, \vec{v}_4$ and \vec{v}_5 have been normalised. The dot products $\hat{v}_1 \cdot \hat{v}_4$ and $\hat{v}_3 \cdot \hat{v}_5$ are calculated. If both the dot products are non-negative, then the point of intersection \vec{R} lies inside the polygon. This works because we use the fact that if the angle between (\vec{v}_1, \vec{v}_4) and the angle between (\vec{v}_3, \vec{v}_5) are

between 0^0 and 90^0 inclusive, then the point is lying inside the polygon. To maintain the accuracy and also if the defined finite plane is of different shape other than rectangle then the above said condition can be checked for other two points \vec{P}_2 and \vec{P}_4 as well. Thus the above condition proves whether the point lies inside the polygon or not.

The next step is to calculate the distance d between the point R and R_0 given by,

$$d = \sqrt{(R_x - x_0)^2 + (R_y - y_0)^2 + (R_z - z_0)^2} \quad (4.16)$$

This is the range for the point on an image plane to the point on the scene. The same can be checked and calculated for all the points on the image plane and range value is recorded for each pixel on the image plane.

Third step

The range value for a single defined plane is calculated. This process is repeated for other every in the scene. As described earlier the scene is made of seven cuboids with six planes in each cuboid. A single ray from a point on the image plane may pass through many scene planes, thus generates many intersection points (see Figure 4.8). The range values obtained from each ray are compared, and the shortest range value is recorded. This then gives the range value that a camera will see for unoccluded objects. The process is repeated for all the rays that intersect the planes and the whole scene which has forty two planes.

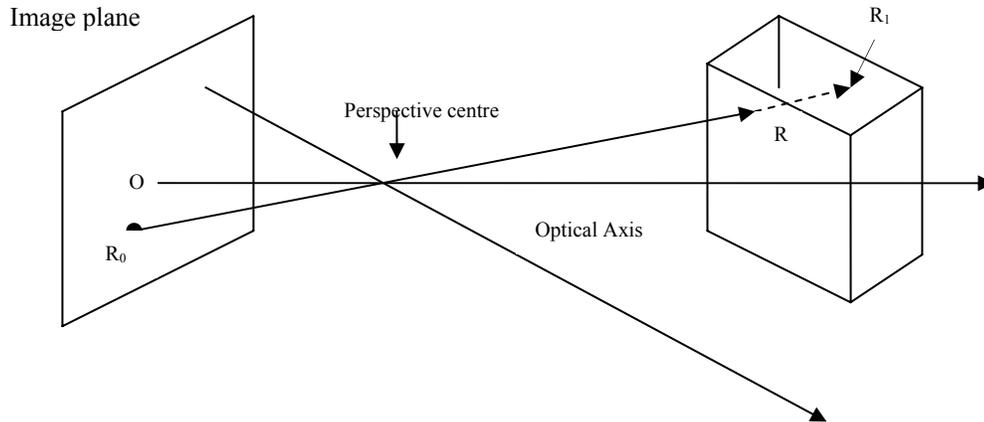


Figure 4.8 : A ray intersecting at two points R and R1. The shortest range value for the is recorded.

Fourth step

After acquiring the range image for the scene by the above steps, the scene is rotated by 30 degrees to generate a new view. Figure 4.9 shows the 0 degree simulated range image. This process of rotating the scene is done in three steps translating the scene to origin, performing rotation and translating it back from origin. The scene is rotated about y axis because in the real experimental work the rotation is carried out about the y axis. The robot scene is located so that the centre of the robot is at $\vec{T} = (0,0,3100)$.

Taking \vec{S} to be a point in the scene and \vec{S}' to be the point after rotation by some angle, then the equation for rotation on the scene has to be rotated by 30 degrees. The equation for rotating a point S on the scene is

$$\vec{S}' = R_m(\vec{S} - \vec{T}) + \vec{T} \quad (4.17)$$

where \vec{T} is the centre point of the scene as given above and R_m is the rotation matrix.

The rotation matrix for a rotation about the y axis is,

$$R_m = \begin{bmatrix} \cos \theta & 0 & \sin \theta \\ 0 & 1 & 0 \\ -\sin \theta & 0 & \cos \theta \end{bmatrix} \quad (4.18)$$

where θ is the rotation angle. The same procedure is carried out for all the points. After the range image was obtained for a 30° rotation the process is repeated for other views. For this simulation thirteen views were chosen, 0° to 360° with an interval of 30° each.

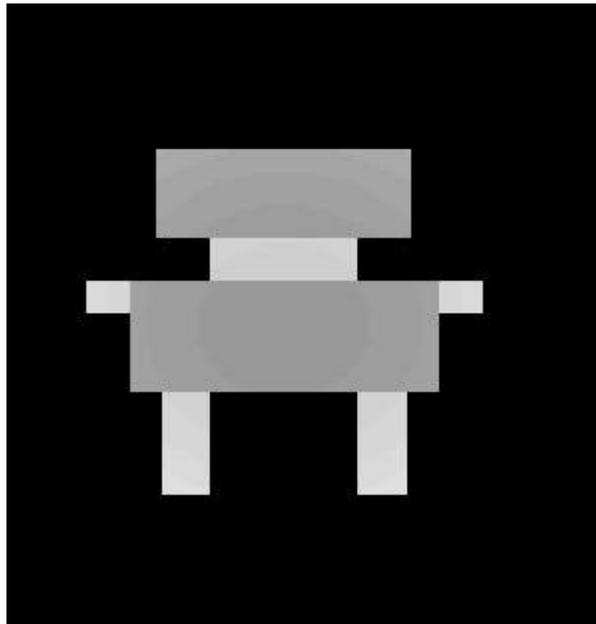


Figure 4.9 : Simulated range image

Fifth step

The fifth step is to establish the exact world coordinate of every pixel in the range image. This is done by taking the cross-section measurement of the simulated object used and comparing it with the number of pixels occupying the image exactly at the same location as that of the scene. This computation is done for the resolution image used in this study (101×101 and 512×512). Since the point of each plane in real world

is predefined (see Table 1), the breadth across the head (cuboid1) of the roboshape is 500 units in real world units. The number of pixels occupying one row in the body area of 0° range image is 46 for 101×101 resolution image plane and 130 for 512×512 . So each pixel width is 10.86 units for 101×101 resolution and 3.85 units for 512×512 resolution image plane. The process of registration and reconstruction for the calculated scene data is explained in section 4.4.

4.2.4 Setup Values used for the simulation study

Table 2 : Showing the values used for simulation

Description	Values
Image plane	101×101 and 512×512
Centre of the image plane	(0,0,0) units
X and Y coordinate for image plane	-50 to 50 and -255 to 255
Perspective centre location	(0,0,250) units and (0,0,600) units
Number of views	13 (0 degree to 360 degree)
Object distance from camera	2800 units (closest range to camera)
Depth of object	Varies from 200 units to 600 units
Number of cuboids used for scene	7
Coordinate axis centre for the scene	(0,0,3100) units

4.3 Experiment using Real Range Data

In this section it is discussed how scenes were captured with the Waikato range imager, converted to range images, extracting real world coordinates and finally finding the scene coordinate centre.

Setup Description:

The experimental setup as a whole, with ranger camera attached to image intensifier, laser light source, signal generator and a scene on a turntable is shown in Figure 4.10. In this experiment the laser light source is located around the lens so that the error produced due geometrical distortion is considerably reduced (see section 2.7.2). The setup values used in this experiment for capturing the views are listed in table 3.

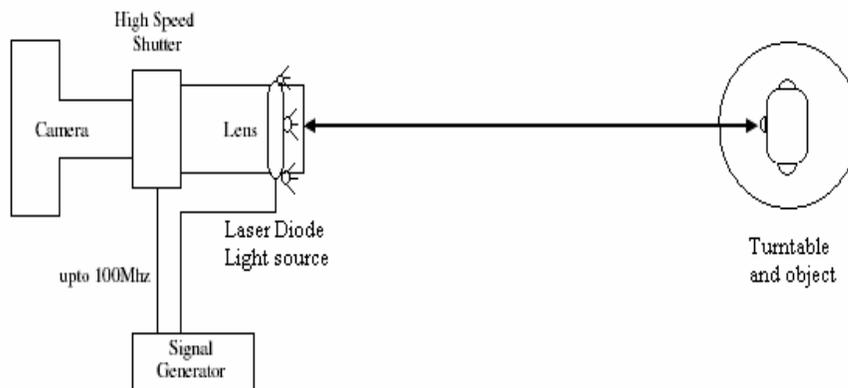


Figure 4.10 : Experimental setup used for capturing the test object.

Table 3 : Experimental Setup values

Description	Values
CCD sensor	512 × 512
Length of video capture	11 s
Number of frames for each capture	330 frames
Frames per second	29 Hz
Modulation frequency	100 MHz
Beat frequency	1 Hz
Gain of image intensifier Minimum	400 V
Gain of image intensifier Maximum	700 V
Cooling time for camera	15 – 20 minutes
Focal length for image lens	80 mm
Distance from camera to the camera	1.096 metre (approximate)
Distance between background and object	20 cm (approximate)
Number of captures	13 (0 degree – 360 degree in steps of 30)

Scene Description

The test object used in this research is a ceramic bear of height 19 cm and a maximum cross-section diameter of 10.9 cm. The object is of uniform colouration. The bear is placed on a turntable which is capable of 1° increments (see Figure 4.11). Thirteen different views of the bear were captured varying from 0° to 360° in steps of 30. The background used in this study is a uniform coloured and low reflective plain cardboard which is located about 20 cm behind the turntable. Using uniform coloured and low reflective background helps to reduce the errors due to poor light returned in black areas of the scene.



Figure 4.11 : Test object bear used in this study as scene is placed on a turntable

Experimental methodology:

The scene is captured as an AVI movie file by the ranger camera. The AVI movie files are processed to give a range data file, intensity file and range image file for the captured scene. The process of extracting the range data is done using a discrete Fourier transform. The Discrete Fourier transform is a procedure for calculating discrete frequency components from sampled time data. Since the frequency domain result is complex, the number of points is equal to half the number of samples. The Discrete Fourier transform is initially used to isolate the fundamental frequency of the beat from the received heterodyne beat signal. The phase of the complex quantity of

the bin corresponding to the fundamental frequency gives the phase shift of the signal. The range to the point of the object is determined by

$$d = \phi c / 4\pi f \quad (4.19)$$

where ϕ is the phase due to time-of-flight of the light, f is the modulation frequency and c is the speed of light. The Discrete Fourier transform is performed down the time axis of the video sequence for each pixel in the camera view. A range value is thereby obtained independently for each pixel. The algorithm uses the above concept to extract range data. The magnitude of the signal received is used to produce the intensity image. Thus the intensity image and range image are extracted initially.



Figure 4.12 : Range image front view of bear captured from ranger camera

The second step is to crop the image. The image is cropped so that only the area of interest on the image is used for the further analysis. Figure 4.12 shows the range image of the test object (the bear) including background object such as the turntable and the aperture due to the field of view of the image intensifier. Care must be taken that the cropping should not affect the other views from different angle, where the bear might be displaced a little bit during rotation. So a buffer of ten pixels on either side of the bear edges is chosen. The original range image or intensity image obtained is of 512×512 pixels and is cropped from image coordinate (100,100) to (310,450) resulting in an image of 351×291 pixels (see Figure 4.13).

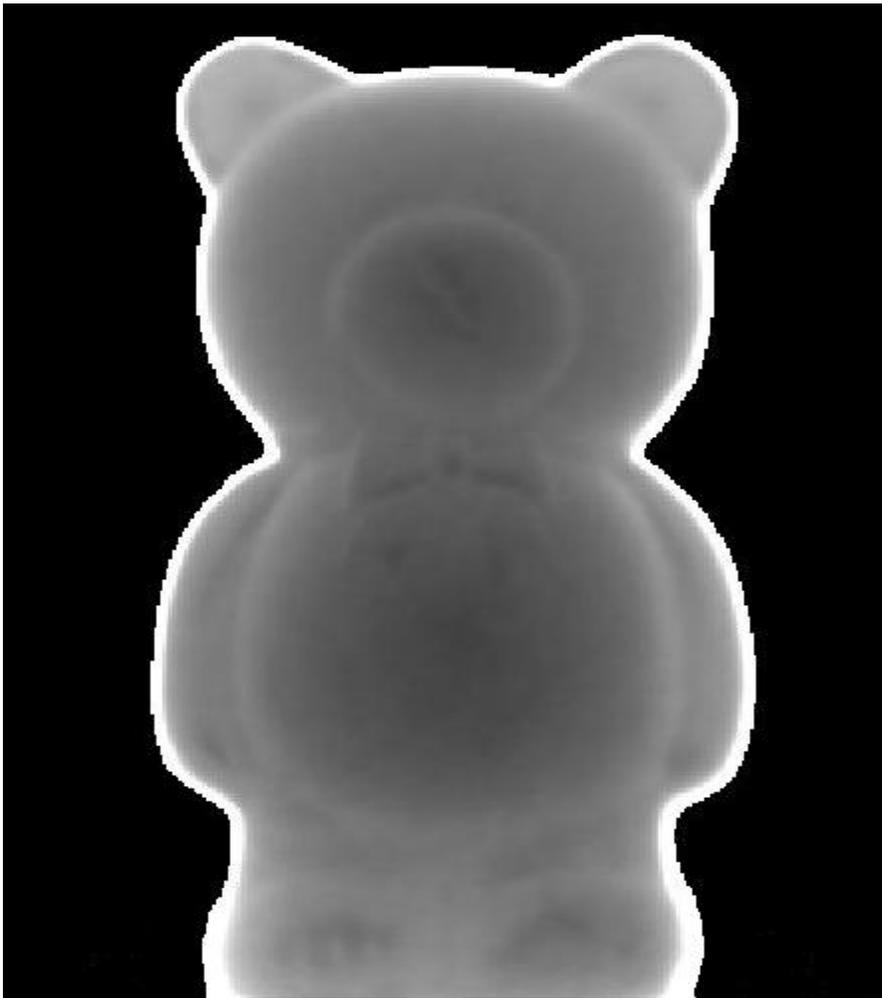


Figure 4.13 : Range image cropped to 351×291 from 512×512

The third step is to establish the exact world coordinate of every pixel in the range image. This is done by taking the cross-section measurement of the bear and comparing it with the number of pixels occupying the image exactly at the same location as that of the scene. This calculation is done at four locations of the test object (Bear) using vernier calipers. Figure 4.14 shows the four cross-section locations of the bear where measurements were made and their corresponding pixel count in the range.

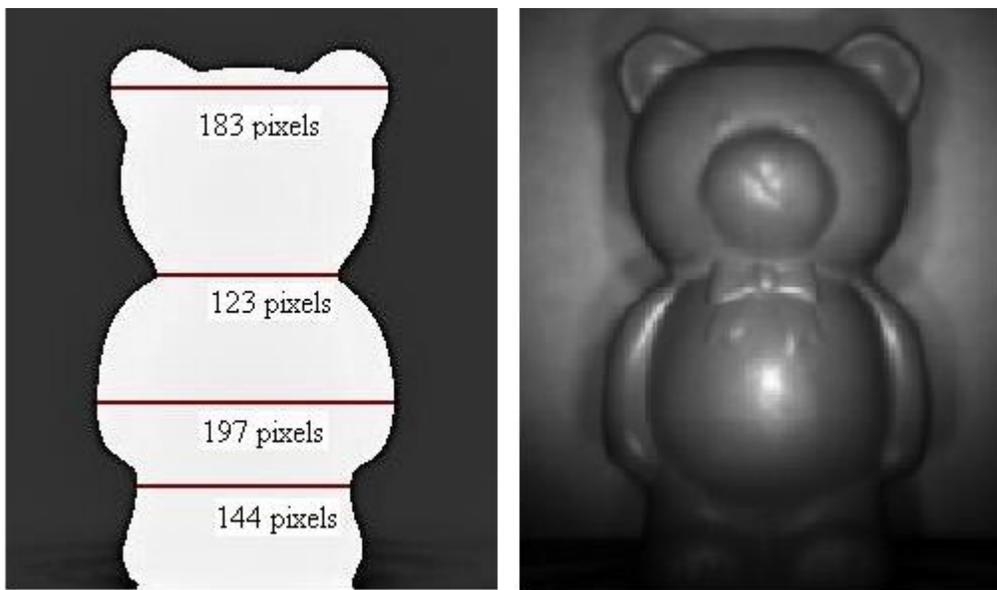


Figure 4.14 : Establishing world coordinates for every pixel in the range image

After obtaining the values across each location averaging them gives a better estimate of the pixel width in desired units. Table 4 shows the measurement of the scene and the pixel count at the same location on the image as shown in the Figure 4.14.

Table 4: Showing the number pixels and corresponding measurement across the range image and the bear

Location	Width (measured on bear) C (metre)	Width (Pixels in range image) P	One pixel width C/P (metres)
Outer ear edges	0.102	183	5.57×10^{-04} m
Across Neck	0.0661	123	5.37×10^{-04} m
Outer arm edges	0.1094	197	5.55×10^{-04} m
Below arm	0.07932	144	5.508×10^{-04} m
Mean			5.5×10^{-04} m

It was found that each pixel in the range image has a width and height of 5.5×10^{-04} m (or 0.55 mm).

The fourth step is the process of defining the image plane in three dimensional scene coordinates. Initially for computational purposes a parallel projection is used in this research. So all the rays projected from the image plane will be parallel to the optical axis (see section 3.2.1). The image plane used in this research has x coordinates ranging from -145 to 145 and y coordinate ranges from -175 to 175 with the image plane centre lying at origin (0, 0). The x and y coordinates of the image plane are multiplied by the pixel width measurement obtained from the previous step to give the scene's x and y coordinates in metres. The range values encoded in the pixels are considered as the z coordinate values (already in metres). Thus an array of three column vector defining x, y and z coordinate values of the scene all having the same units in metres are obtained.

The fifth step is to threshold the unwanted data available in the range image, such as the background and poor signal data. The main aim of this is to remove the noise that is caused due to the background and also the shadows. By thresholding the intensity, errors that are produced near the edges of the bear where the light intensity is weak can be reduced.

The sixth step is to find the coordinate axis centre for the scene so that registration of different views can be performed. Since the rotation of the scene is about the y axis we concentrate on finding the x and z coordinate centres. The calculation described in section 4.2 of the simulated work is applicable here. In simulation work the optical axis centre and depth of the object were known so finding x and z axis centre was not a difficult task. But in the real data the depth of the object is irregular also the optical axis is unknown. Two range images 0° and 180° views were considered initially for finding the x coordinate centre. The edges for one of the views is superimposed on the other image (see Figure 4.15). The dashed dotted line shows the edges for 180° view range image.

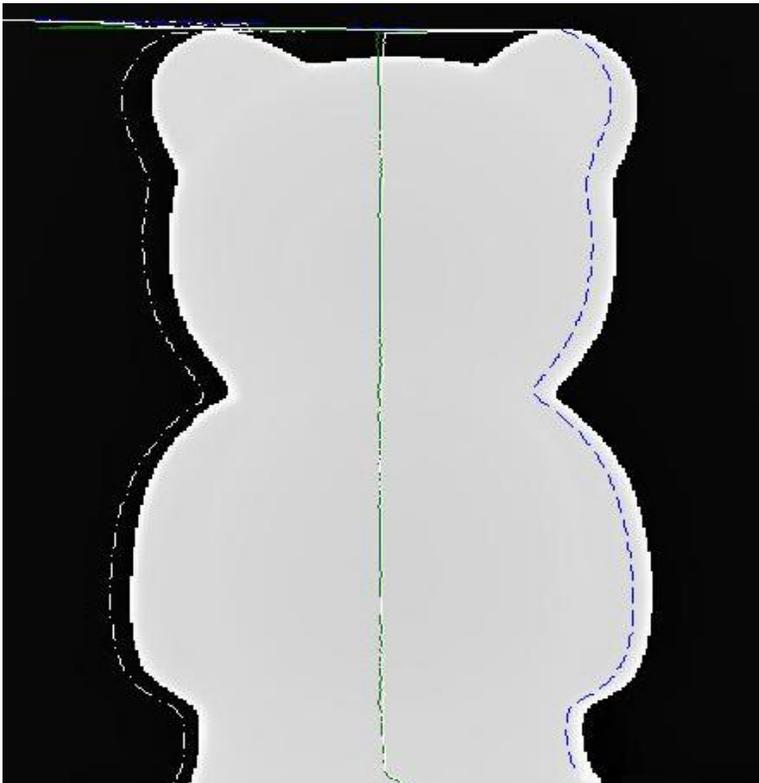


Figure 4.15 : Range image showing 0° overlapped with 180° (dashed line) and the x coordinate centre

The centre x coordinate is calculated for every row involving the two outer edges and other two inner edges. The black and white line in Figure 4.15 shows the two centres obtained by the outer edge and inner edge. The average of these two centre coordinates gives the x axis coordinate centre. The x coordinate centre was found to be located at $-5.4e-4$ m on the image plane.

Finally the z coordinate centre is determined by the following method. Two views considered for finding the z coordinate centre are 0° and 90° . The distance from the x coordinate centre to the edge of the bear is calculated for each row of the range image. The maximum distance d is recorded. In the 0° view, the smallest range value d_r is determined which gives the closest point to the camera. The difference between d and d_r gives the z coordinate centre for the scene. The z coordinate centre determined for the bear has a value 0.81m. Figure 4.16 shows the two views and z coordinate centre. Thus the scene coordinate axis centre calculated is $S_0(-.00054, 0, -.81)$ m.

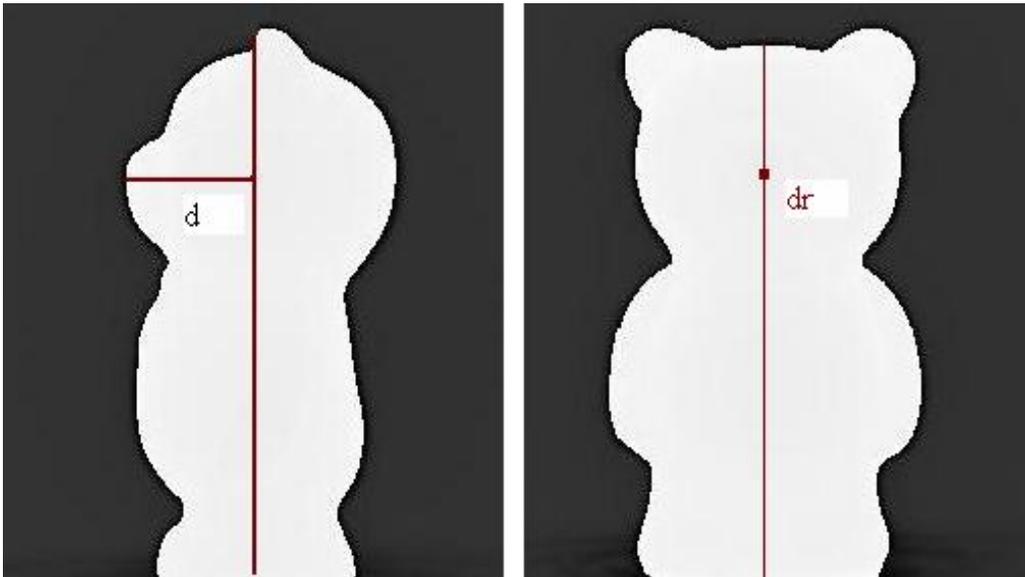


Figure 4.16 : Range image showing calculation of z axis centre (see explanation for detail)

4.4 Registration and Reconstruction

In this section it is described how different range view data is registered together and surface rendered using the FastRBF tool. The process described below is similar for the simulated range data and the real range data.

The first step is to perform registration by rotating different views in to a single coordinate system. For performing transformation requirements are scene coordinate centre and scene coordinate values. These values were determined for the real range data in sixth and fourth steps and in fourth and fifth step for simulated range data. Let $S_\theta(x,y,z)$ be the scene coordinates of a certain θ view which has to be rotated on to a 0° coordinate view. We write the S_w for the S_θ coordinate transformed in to the 0° view coordinate system. Then

$$S_w = \text{Rm}(S_\theta - S_0) + S_0, \quad (4.20)$$

where S_0 is the rotation coordinate centre as found above and R is the three dimensional rotation matrix for the rotations about the y-axis. To rotate the θ view in to the 0° view coordinates an angle $\theta = -\theta$ is used in rotation matrix.

$$\text{Rm} = \begin{bmatrix} \cos(\theta) & 0 & \sin(\theta) \\ 0 & 1 & 0 \\ -\sin(\theta) & 0 & \cos(\theta) \end{bmatrix}. \quad (4.21)$$

This rotation process is repeated for all views on to a single coordinate system.

The final step of experimental methodology is to reconstruct the scene points in to a complete surface. This is done with the use of FastRBF tool (see section 4.1). The input for the FastRBF is given as a structure having the scene coordinates, length of the scene coordinates and origin of the scene coordinate. The x, y and z coordinate value of the scene in *location* variable. The length of the *location* variable is given in *size* variable. The origin of the scene is given in *origin* variable of the scan structure. For example, a scene S with origin at S_0 then the structure scan will be given as

$$\begin{aligned} \text{Scan.Location} &= [x \ y \ z] \\ \text{Scan.Length} &= \text{length}(z) \\ \text{Scan.Origin} &= [x_0 \ y_0 \ z_0] \end{aligned}$$

After inputting the values the initial step will be calculating the normal using the inbuilt FastRBF function “*fastrbf_normalsfromscan*”. Then density from the normal is calculated using “*fastrbf_densityfromnormals*”. This density function creates an off surface points using surface normal and surface point. So a distance for this off surface point has to be defined in this density function. This distance is also called as projection distance. The minimum projection distance used in this study is zero. The maximum projection distance is varied from 10 units to 3100 units for the simulation study and for the real range data the maximum projection distance is varied from 0.0001 m to 0.1 m. The best projection value used in this research for both real and simulated range data is discussed in the result section. Then interpolating the data from the density values is carried on using the function “*fastrbf_fit*”. Accuracies should be specified while using the function to obtain a best fit. These accuracies refer to the tolerance where the interpolated RBF values should be within this amount of the original value at each point. The values of accuracies are varied from 0.1 units to 10 units for simulated range data and 0.001 m to 0.01 m for real range data. Then the fitted data is isosurfaced using mesh generation function “*fastrbf_isosurf*” with defined surface resolution. The surface resolution is varied from 0.1 units to 10 units and 0.0001 m to 0.001 m. The discussion of best parametric values used, results and errors obtained while performing the above is discussed in the following result chapter.

5 Results

Here we present two different outputs of the research: shape reconstruction using simulated output and shape reconstruction for range data captured from the Waikato ranger.

5.1 Simulated Reconstruction results

The results obtained from each stage of the simulated work are discussed in this section. The range images were simulated for two different image resolutions: 101×101 and 512×512 pixels. Low resolution range images were used for faster computational purposes and for testing/debugging the code. Figure 5.1, 5.2, 5.3 and 5.4, shows the range image obtained from the simulated object at the 0° view and 270° view for both the resolutions used in this study.



Figure 5.1 : Simulated 0° Range image for 101×101 resolution image

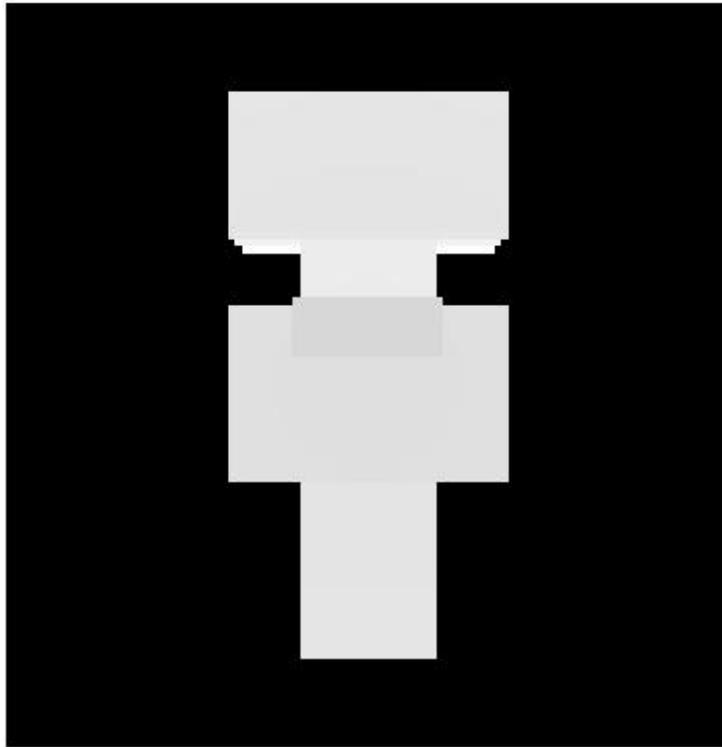


Figure 5.2 : Simulated 270° Range image for 101×101 resolution image

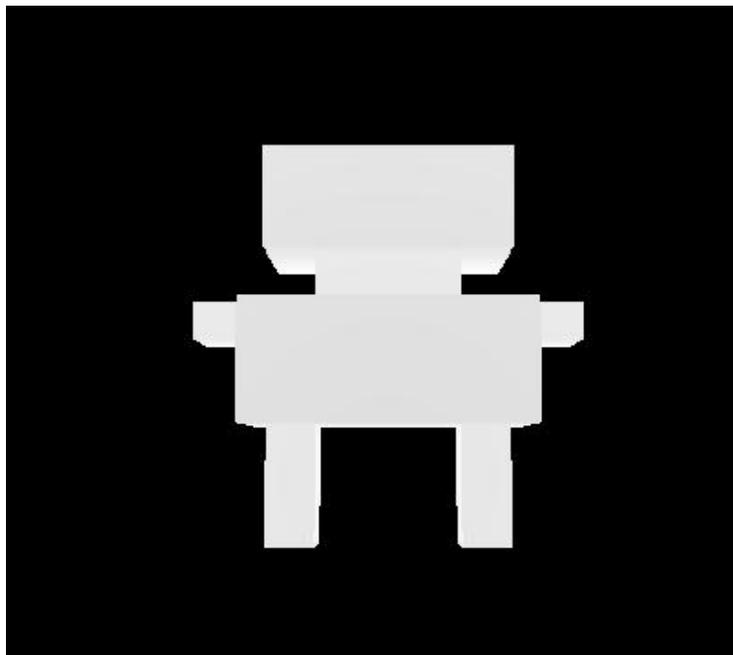


Figure 5.3 : Simulated 0° Range image for 512×512 resolution image

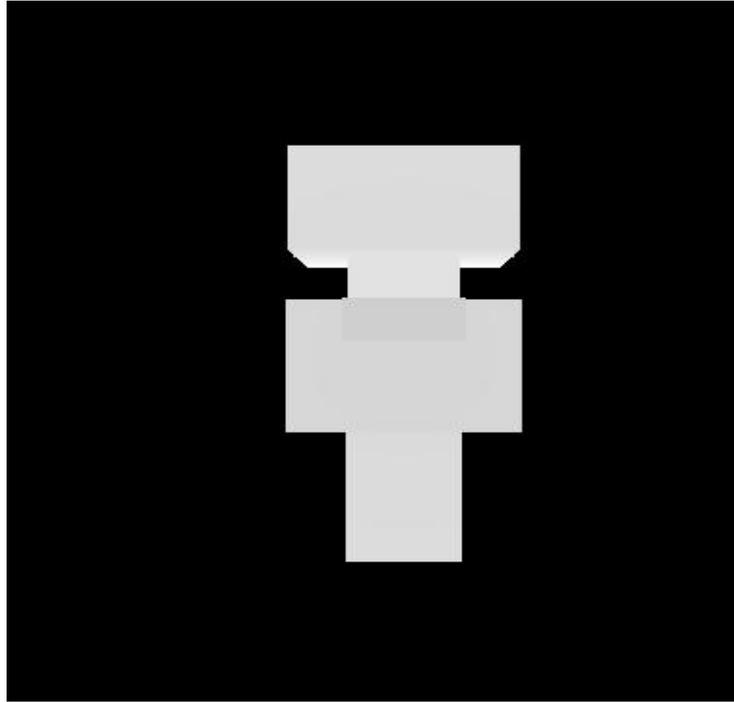


Figure 5.4 : Simulated 270° Range image for 512×512 resolution image

The background is separated from the object of interest by thresholding the range values. The threshold range value used in this simulation work is 4800 units as the background is located at 5000 units. After calculating the real world coordinates as described in section 4.2.3, the data are available as a set of points (point cloud data). The image appears curved at the edges because of the radial distortion (see section 2.7.1). The experiment is carried out with this distortion as the real world range images also have this distortion. Once the scene data for each range images are obtained, merging different views on to a single coordinate system is performed (see section 4.2.3).

Initially reconstruction is performed for 0° and rotated 30° range image on to a 0° coordinate system using the FastRBF tool. The point cloud data with scene origin and length of the point cloud are passed as input for the FastRBF to perform reconstruction. The steps involved in FastRBF reconstruction were explained in section 4.2.3 and 4.1. Therefore the accuracies and parameter values used in reconstructed results are discussed in the following.

The surface normal is calculated for the surface points. Then density function is used which creates off surface points from the surface normal and surface point with some distance specified from the surface point, which is also called as projection distance (see section 4.1.1 for detail). The projection distance is varied from 500 to 3000 units for testing purpose and a best projection distance is determined. It was observed that the reconstruction failed for the values below 500 units though a part of neck and face were visible. For projection distance above 1000 units the reconstruction failed completely. This shows that the projection distance for the off-surface points should not be very close or too far from the given surface point. Then FastRBF interpolation for density data is performed and results obtained are verified for the accuracy value varying from 0.01 units to 100 units. The interpolation failed during reconstruction for accuracy values greater than 10 units and less than 0.1 units. Accuracy values ranging from 1 unit to 5 units had better reconstructed results than any other. The final step for reconstruction using FastRBF is isosurfacing. The resolution values varying from 0.1 to 10 units were tested for isosurfacing. In low resolution isosurfacing, resolution values greater than five units did not produce the surface with all features of the object. Lesser values like 1 unit and 3 units were also considered for high resolution isosurfacing but results obtained using these values produced a very less difference with surface obtained with resolution of 5 units. The disadvantage of using 1 unit and 3 units was it takes lots of computation time to execute and also consumes lots of memory for isosurfacing. Thus the final values used in this study for reconstruction at each stage in the FastRBF reconstruction are; for density function minimum projection distance is zero which is default in FastRBF and maximum projection distance is 500, fitting accuracy of 1 and isosurfacing resolution of 5 units.

The reconstruction is repeated by merging two other views 180° and 330° along with 0° and 30° view rotated on to single coordinate system. Then reconstruction is performed. The reconstructed surface is shown in figures Figure 5.5, 5.6, 5.7 and 5.8.



Figure 5.5 : Reconstructed surface using 0° , 30° , 180° and 330° range data (front on)

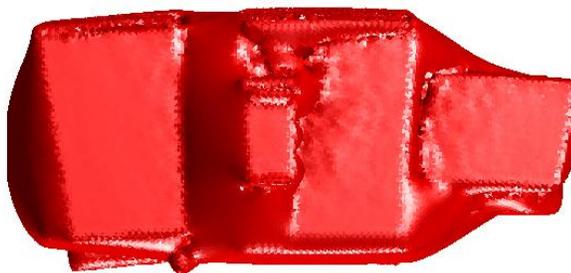


Figure 5.6 : Reconstructed surface using 0° , 30° , 180° and 330° range data (side view)

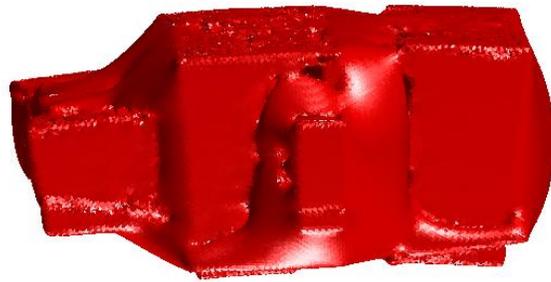


Figure 5.7 : Reconstructed surface using 0° , 30° , 180° and 330° range data (side view)

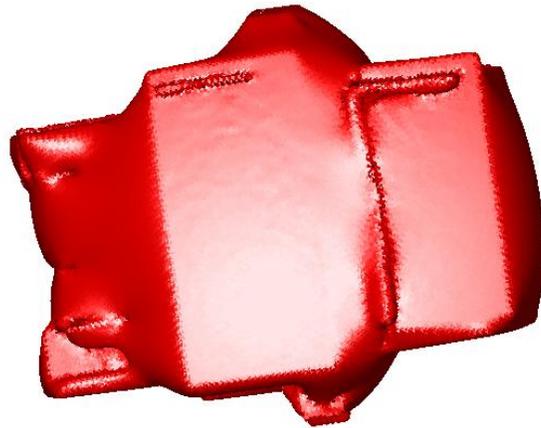


Figure 5.8: Reconstructed surface using 0° , 30° , 180° and 330° range data (rear view)

5.2 Experiment Reconstruction results

The results obtained from each stage of the experimentation using real range data obtained with the Waikato range imager are now presented. The test object “bear” was placed on a turntable and rotated with an increment of 30° from 0° to 360° . Range images of the bear were captured at each 30° degree view. Figure 5.9, 5.10 and 5.11 shows the intensity and range image for the bear at the 0° view, 30° view and 270° , respectively. The range image is encoded so that darker greys are closer. The bright white fringe around the bear in the range image occurs because of mixing of the signal from the background and the edges of the bear itself. This bright fringe is ignored in three dimensional rendering of the bear. This is achieved by the segmentation process using the intensity values. The process of segmentation is explained in the following paragraph.

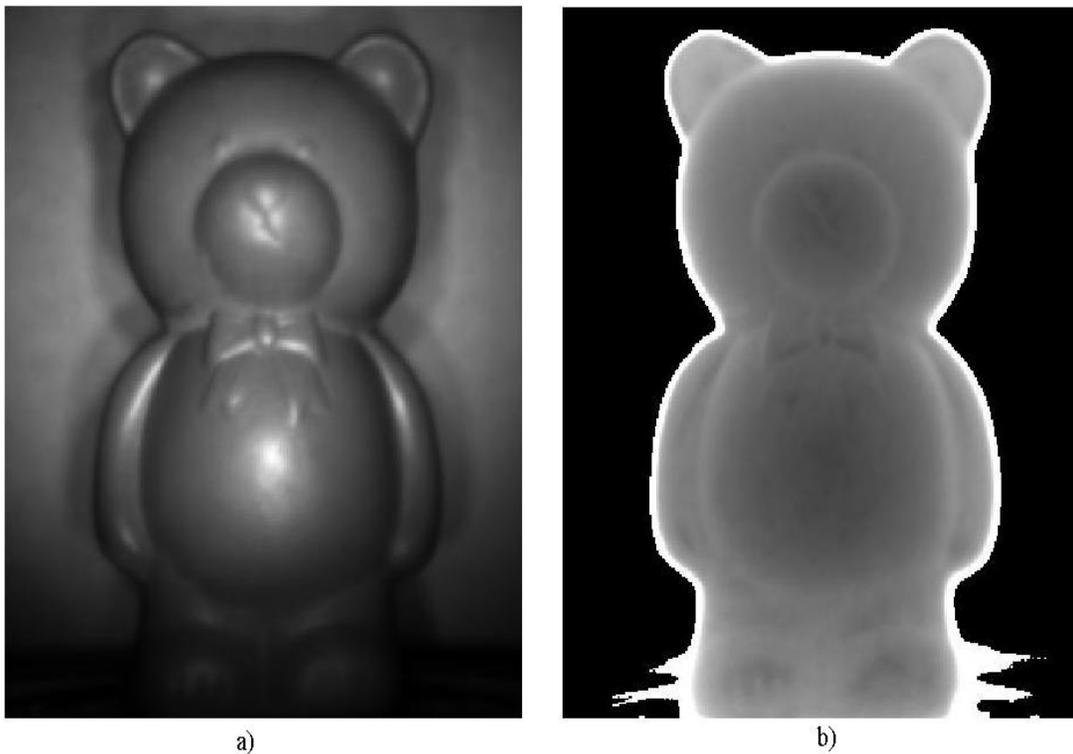


Figure 5.9 : a) Intensity image and b) range image for the 0° view

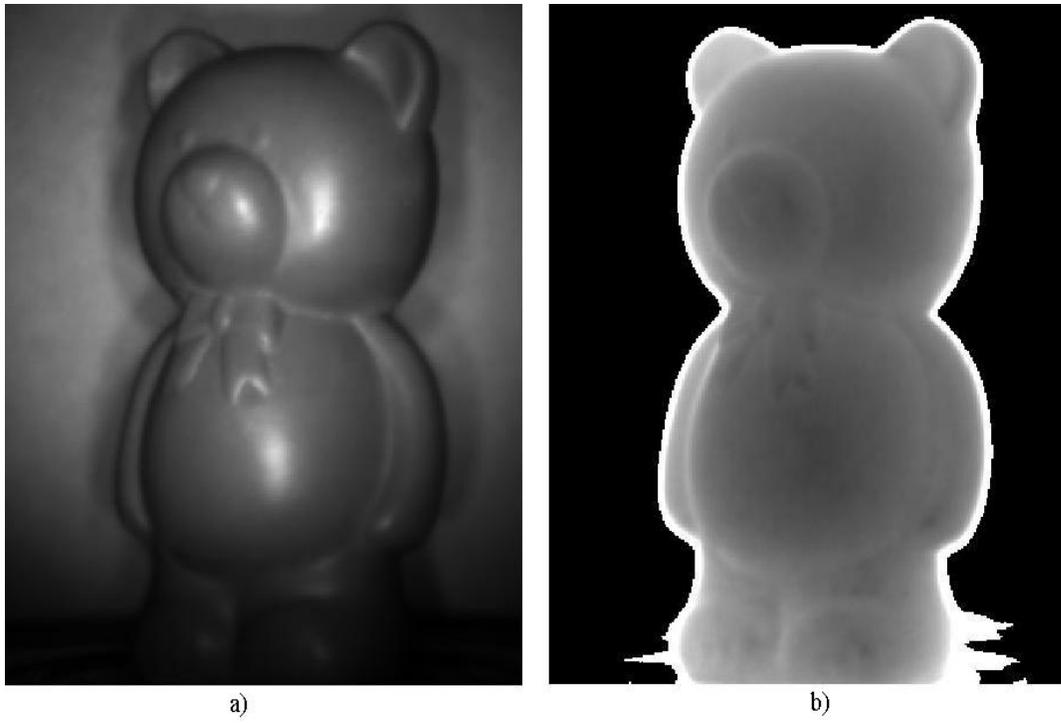


Figure 5.10 : a) Intensity image and b) range image for 30° view

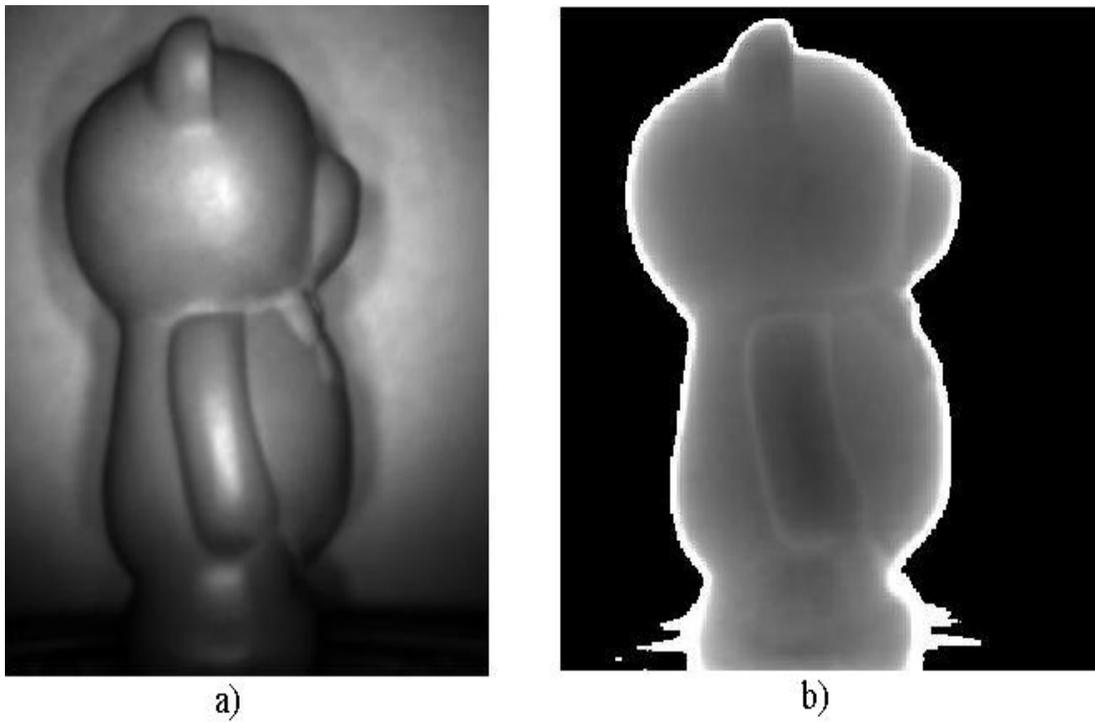


Figure 5.11 : a) Intensity image and b) range image for 270° view

The area of interest on the range image is the bear, so segmentation is done on the image to separate the bear from its background. Two approaches to segmentation were tested and compared. The first is thresholding the range value and the second is thresholding both the range and intensity values. The background is eliminated by accepting only the range values in the range image that is less than 0.87m. This range value is chosen because the approximate distance from camera is known and the approximate distance separating background and bear is also known (see Table 3). Thus by taking the difference between the two distance values we can locate the objects area of interest. The second kind of segmentation is based on intensity value and range values from the scene. By taking only the high intensity values, the errors that are produced where a poor signal is received back, such as on edges can be reduced. The intensity value used for thresholding in this research is 150000. Figure 5.12 and Figure 5.13 and show the results of two segmentation schemes. In figure 5.13 the bright fringe cannot be seen around the edges. Thus the segmentation using both range and intensity value is chosen for further experimental study.

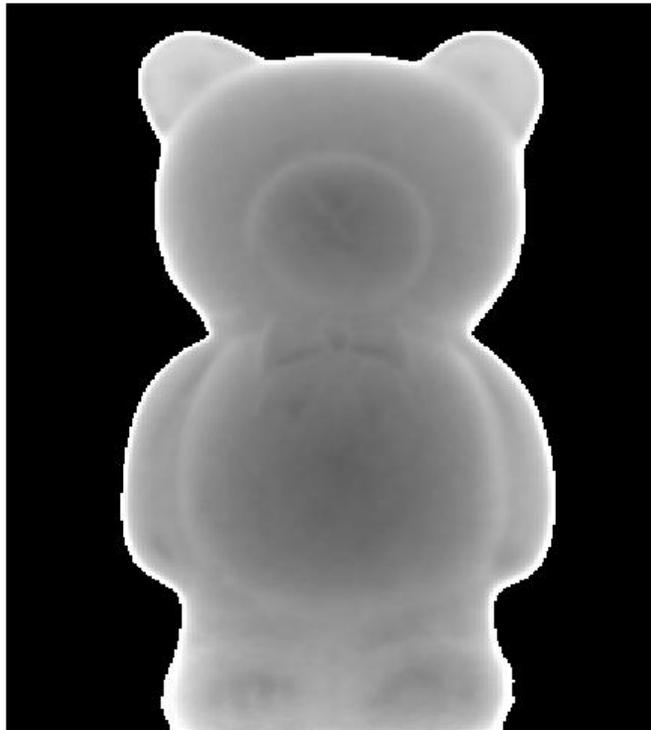


Figure 5.12 : Segmentation of bear based on range values only



Figure 5.13 : Segmentation based on using both range and intensity values

The segmented data are then used for reconstruction with the FastRBF tool. The real world coordinates for the scene were calculated for each range image as discussed in section 4.3. First the real world data obtained for the zero degree view is used for reconstruction using the FastRBF tool (see Figure 5.14). This reconstruction is performed to find the suitable values that can be used in FastRBF later on when merging other views.

The scene coordinates and its coordinate centre is given as input to the FastRBF and their normal is calculated. Then the density data is estimated from normal and surface point, using maximum projection distance and minimum projection distance (see section 4.1.1 for detail). Maximum projection distance values were varied from 0.0001 m to 0.1 m and minimum projection distance is maintained at 0. The best value for maximum projection distance was determined by trail and error and found to be 0.009 m. Then the calculated density data is used for the interpolation. To determine a good accuracy value to use in the interpolation the

accuracy values were varied from 0.001m to 0.01m. The accuracy value was chosen for the best achieved and was found to be 0.005 m. The final stage of reconstruction is isosurfacing from the FastRBF fitted data. The isosurfacing is performed with a resolution value of 0.0008 m from the interpolated data, as this value was found to be efficient and reveals the features of the bear clearly.

To review the final values used in this study for reconstruction at each stage of FastRBF reconstruction are; for density function minimum projection distance is zero which is default in FastRBF and maximum projection distance is .009 m, fitting accuracy of .005 m and isosurfacing resolution of .0008 m.

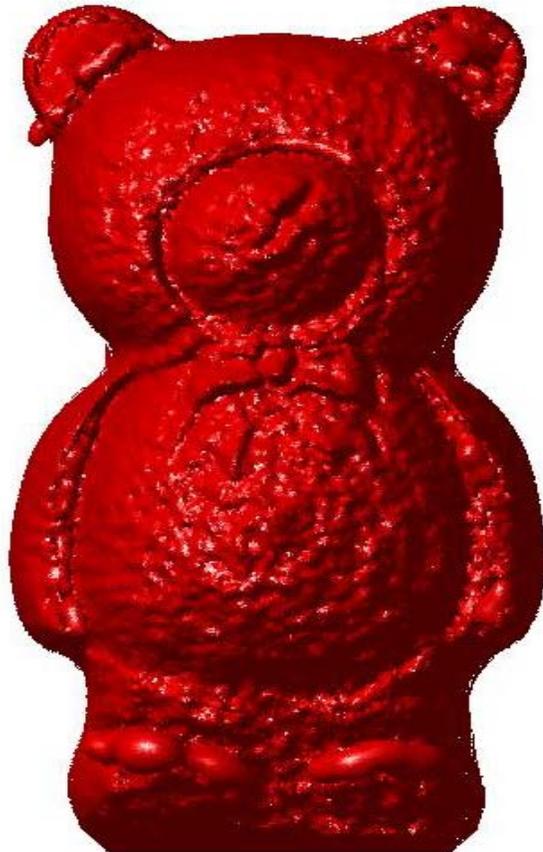


Figure 5.14 : Shape reconstructed 0° degree range image

The next stage is merging the views and their reconstructed outputs. The 30° view range data was rotated on to the 0° coordinate system (see section 4.3). The obtained rotated range data are appended to the 0° range data to give a complete set of points having an extended view with overlap of points. This combined dataset was passed to the FastRBF tool for shape reconstruction with the same FastRBF values used in the 0° range data reconstruction. Figure 5.15 shows that the reconstruction is partly successful for both the segmentations used. However the registration is not acceptable as can be seen by examining the left ear of reconstructed surface: two separate ears are visible here. By visibly observing the reconstruction output, the front part of the bear is reconstructed and areas around the left ear of the bear were not reconstructed.

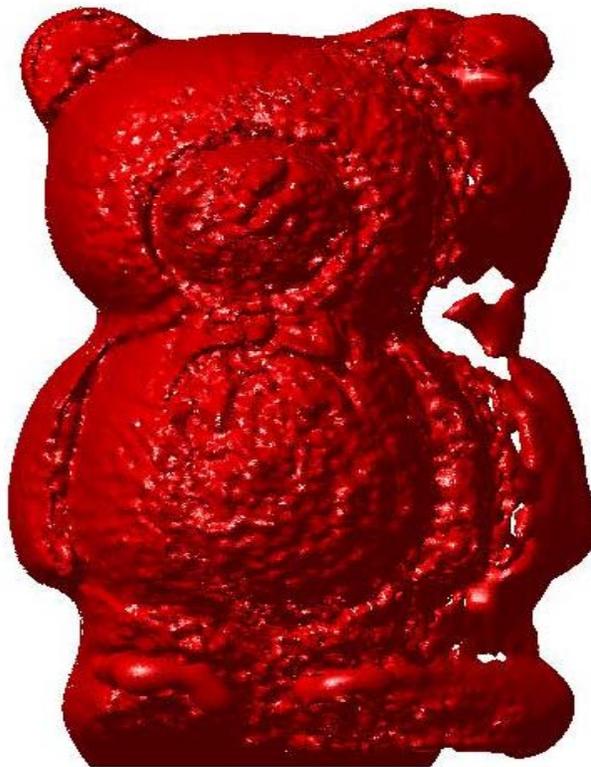


Figure 5.15 : Shape reconstruction using Fast RBF for two set of data (0° and 30°)

The data used for reconstruction were analysed for any errors during the registration process. Figure 5.16 shows the plot for two set of data used for the reconstruction. The analysis is carried on by calculating the deviation between the two set of range data. The rotated 30° data is fitted using “fastrbf_fit” with an accuracy of 5mm. Then the fitted surface is evaluated with the locations of the pixels of the 0° view using “fastrbf_pointeval”. The evaluated value is compared with the range value of the 0° view at each pixel location which gives an estimate of the transformation error. If the difference is zero then the transformation of the 30° data was carried out without any errors. Figure 5.17 shows the error caused due to deviation of two set of range data inputs. The grey colour bar shows the error produced in millimetres. The extreme error values down the right side of the bear are due to the area being obscured in the 30° data.

The root mean square (RMS) value is also calculated for the evaluated data obtained for both 0° and rotated 30° in a suitable region-of-interest of the image. The least value obtained shows that range transformation performed during registration is the best. Let E_0 be the evaluated range for the 0° range image and E_{30} be the evaluated data for the registered 30° range data. The root mean square value E is given by

$$E = \sqrt{\Omega(E_0 - E_{30})^2} \quad (5.1)$$

where Ω is the set of pixels of the region of interest used in the calculation. For the data shown in figures 5.17 an RMS error value of 0.155 was obtained. Given that the range data in the range images has a much lower noise level (precision of 0.005 m). The high RMS value of 0.155 further suggests that the registration is not optimal.

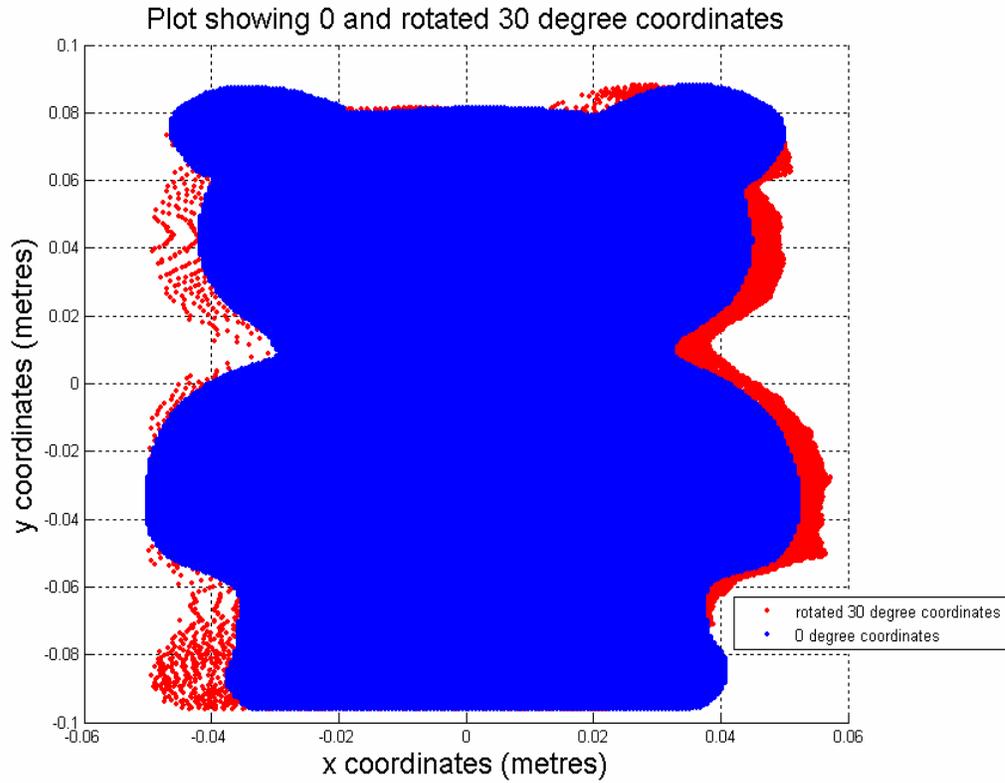


Figure 5.16 : Scatter plot showing two set of range data 0° and rotated 30°

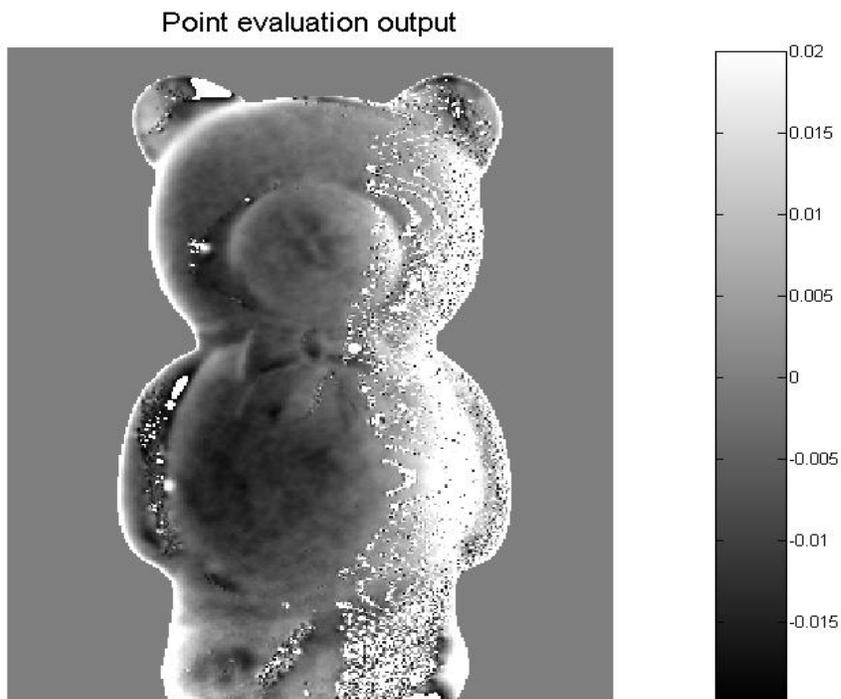


Figure 5.17 : Error between two data sets determined using point evaluation

The angle of deviation is calculated from the evaluated error data by fitting a linear fit for the plot obtained across each row of the image shown in figure 5.17. Figure 5.18 shows the plot for evaluated error values across five rows of the evaluated error image with respect to x coordinates. The linear fit equation is obtained from each plot, which is of the form

$$y = mx + c \quad (5.2)$$

where y is the evaluated deviation error value, x is the x-coordinate value of the image, m is the slope and c is the constant. The angle of deviation θ is calculated using,

$$\theta = \tan^{-1}\left(\frac{\Delta y}{\Delta x}\right). \quad (5.3)$$

By substituting the estimated slope in equation 5.2 the angle of deviation can be calculated. The average angle of deviation across the five rows calculated was found to be 10° . This suggests that there is a substantial error in the rotation of the data and the non-zero c values suggests an error in the translation of equation 4.24.

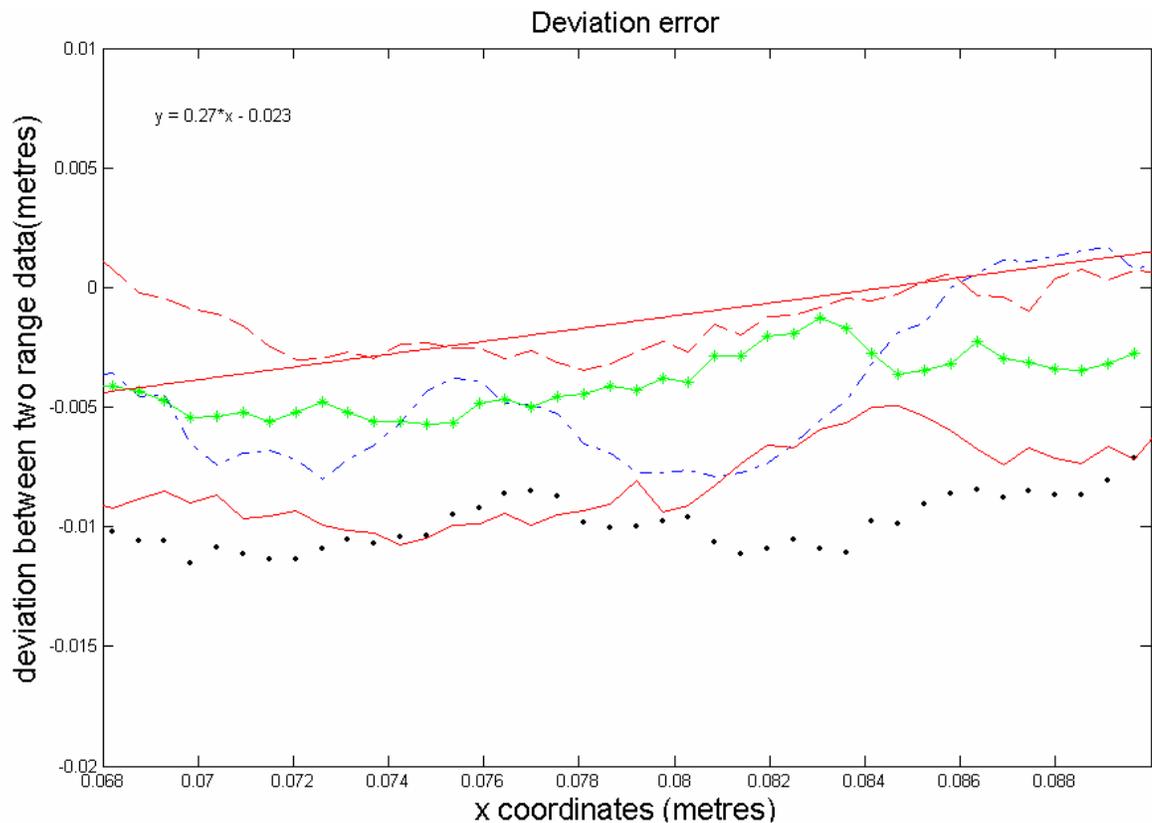


Figure 5.18 : Deviation error between two data sets across the image at five locations

The coordinate value used in the translation and the rotation angle were varied to find the best values for registration. For the registration shown in figure 5.17, the translation vector is $T = (-0.0005, 0, 0.81)$ and a 30° used, By trail and error a translation of $T = (0, 0, 0.805)$ with a 30° rotation was found to be superior. A RMS value of 0.075 was obtained. This is half that of the first trail registration. An angle of deviation of 4° was found which suggests that an even better registration by modifying the rotated angle could be obtained. The new reconstruction is shown in figure 5.19 to 5.21. Figures 5.23 to 5.25 shows the point evaluated image used for calculating angle of deviation and RMS value. By keeping the translation value same, the rotation angle is varied and registration is performed. But the reconstruction failed for the change in rotation angle which suggests that the error lies in the data used for transformation and not in the rotation angle.



Figure 5.19 : Reconstructed bear with corrected range data



Figure 5.20 : Side view of the reconstructed bear



Figure 5.21 : Side view of the reconstructed bear showing the rotated 30° range data

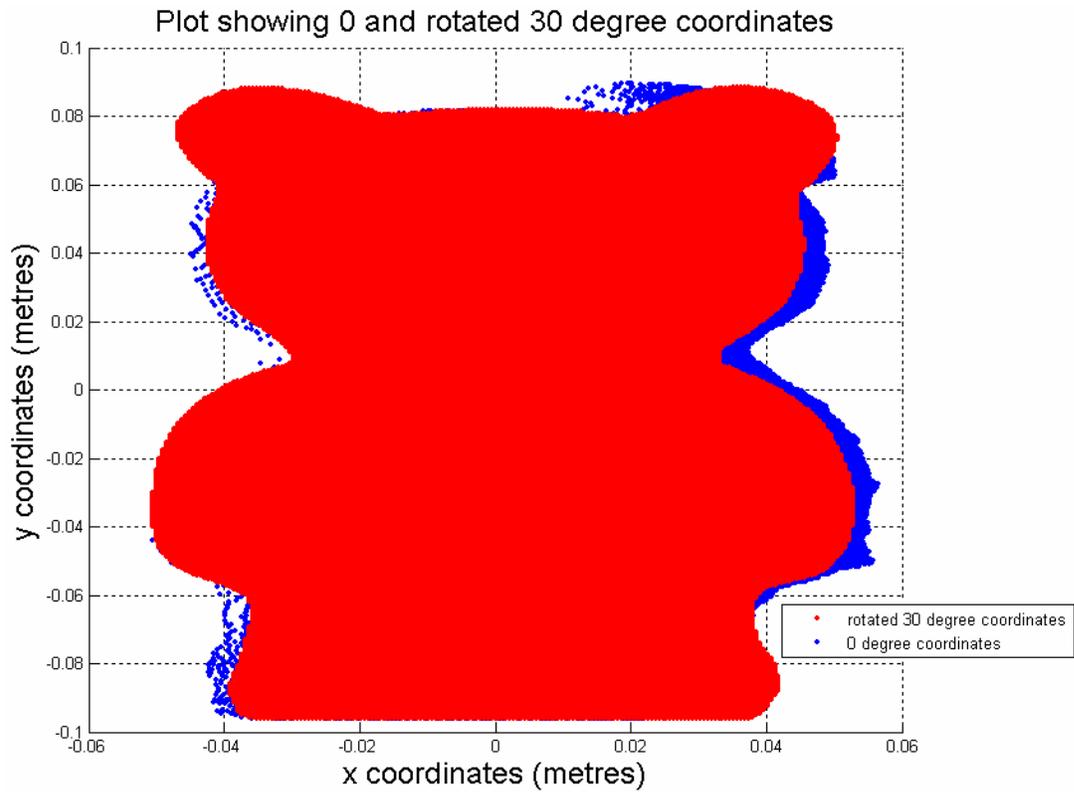


Figure 5.22 : Scatter plot showing two set of range data (0° and new rotated 30°)

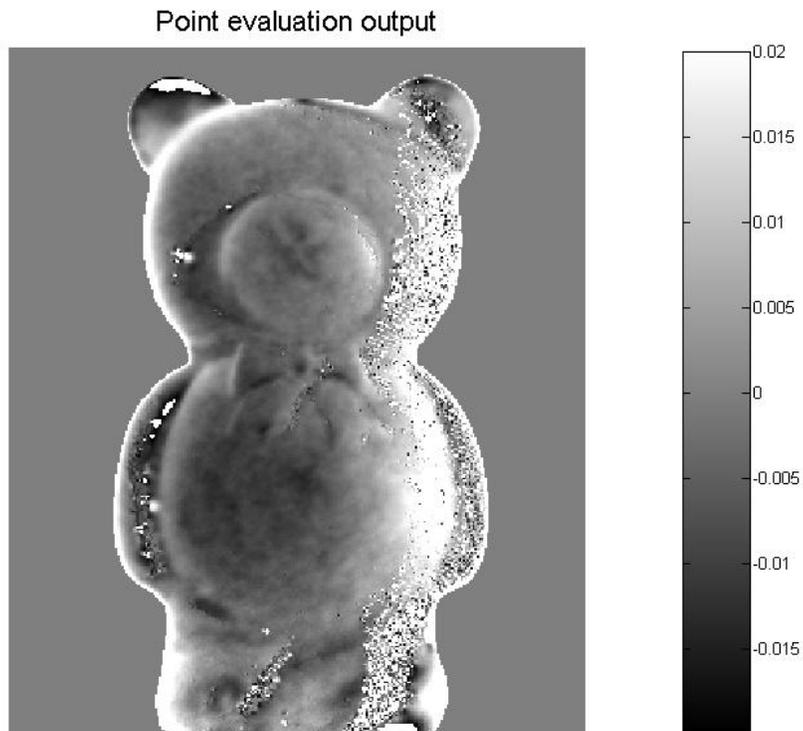


Figure 5.23 : Error between two new data sets determined using point evaluation

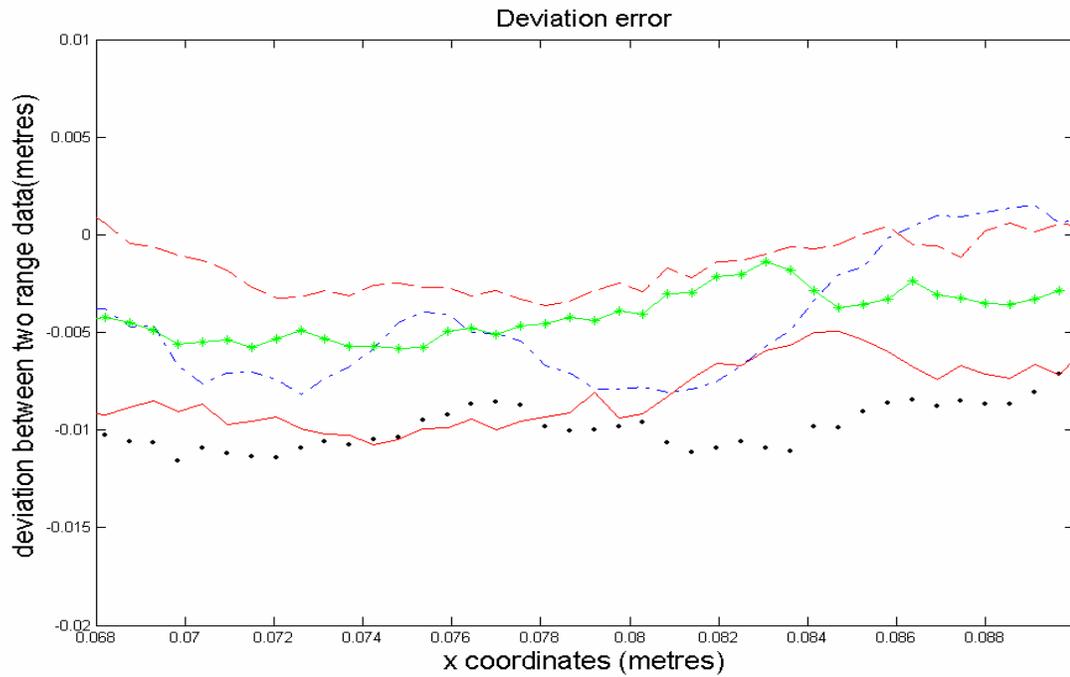


Figure 5.24 : Deviation error between two data sets across the image at five locations for the coordinate registration values

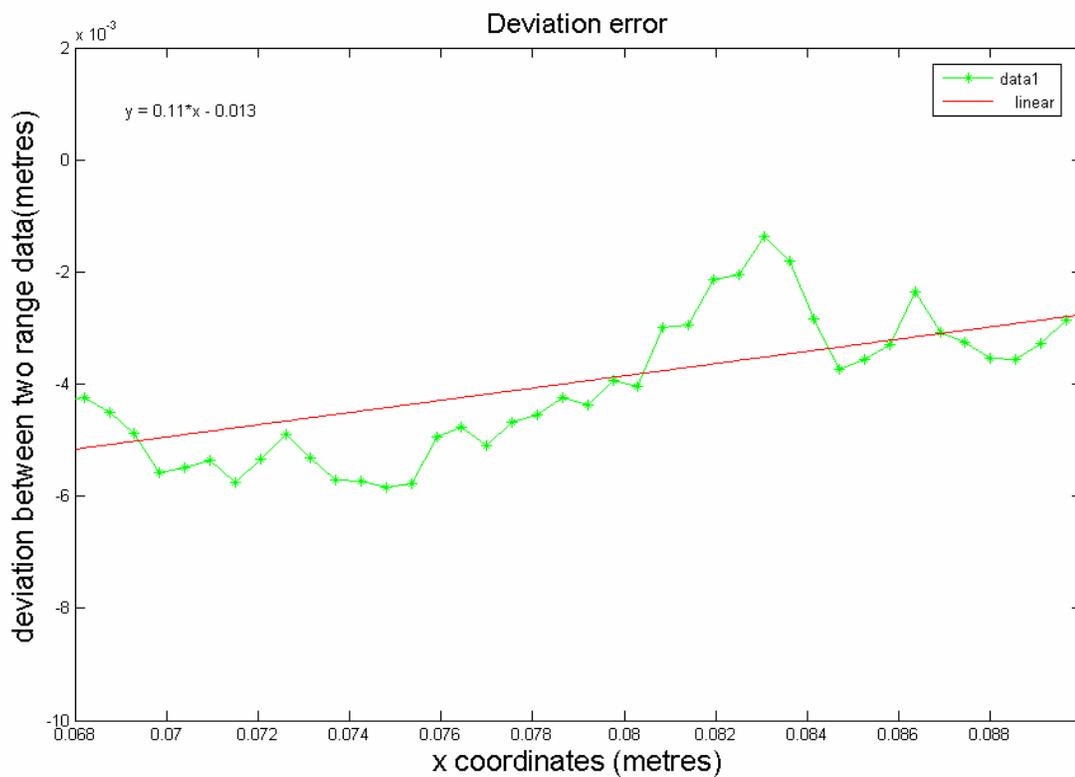


Figure 5.25 : Linear fit for a single row across the deviation error image

6 Conclusion

The initial work of the research was performed to explore the parameters that are used in the FastRBF for reconstruction of computer simulated range image. It was found that the FastRBF software is very sensitive to the accuracy parameters used; even a change of 0.001 units in the accuracy parameter can produce a drastic change in the output. It was also observed that neck of the simulated range image was not reconstructed. This was due to the surfaces that were not visible during the range image capture. So the FastRBF connected the head and body surface leaving the interpolated neck region behind the joined surface.

The FastRBF reconstruction of the bear using the Waikato ranger data was performed with the function parameters used in the simulation work. The registration failed initially but the error was observed to be in the transformation values used. A best fit of two merged data was achieved by trial and error for different scene coordinate centre used during transformation. Initially the translation values were changed by keeping the rotation angle unchanged. The best fit with the changed translation values was observed. Then the angle of deviation was calculated using point evaluation method (see section 5.2), for every change in the two merged data. The determined angle from angle of deviation was substituted in the rotation transformation and registration was performed with the change. But the reconstruction failed for any addition of rotation angle. This proves that there was no

error in rotation angle, instead the error in the original data. The RMS value that was estimated for each change performed also shows the same.

During the process of analysing the registration process it was observed that the radial distortion inherent in the acquisition had a significant effect on the registration process. This was determined when the angle of deviation between two range data was estimated for different transformation values. In figure 5.19 it can be observed that the area near left leg, collar and nose of the bear shows that two datasets are merged with some error in them caused due to distortion. There are other systematic errors in the range image which can considerably affect the reconstruction process. Some of them like the geometrical distortion produced due to the traveling distance of the light and iris effect produced during the shutter operation.

Future works

Many aspects of my study require further attention to improve on reconstruction of surface using multiple images. Areas where improvement can be carried on this research are correcting distortion, estimating exact translation value after removing distortion and finally camera calibration. Performing reconstruction after fixing up the above said systematic error can improve the registration. The improved Waikato ranger that was available during the finish of this study is capable of giving a range data with a precision in sub millimetre range, so performing registration using this data can produce best reconstructed surface even for complex surfaces. The test

object used in this research was a uniform coloured and smooth surface. So performing reconstruction for complicated surfaces like cup with a handle or vase with a hole on the surface will be of valued addition to the current research. In the FastRBF tool global consistency was used for interpolation, so using the local consistency for each pixel interpolation can further be improved giving a better reconstructed output for the surface.

References

A. Aizawa, *Image processing technologies: Algorithms, sensors and applications*, Marcell Dekker, New York, NY, 2004.

G. A. Baxes. *Digital image processing principle and applications*, Wiley, New York, NY, 1994.

D. A. Carnegie, M. J. Cree, A. A. Dorrington, *A High-Resolution Full-Field Imaging System*, Review of Scientific Instruments, 76, 2005.

M. J. Cree, A. A. Dorrington, and D. A. Carnegie, *A Heterodyning Range Imager*, Ninth IAPR Conference on Machine Vision Applications, Tsukuba, Japan, pp.16 – 18, May, 2005.

B. Curles and M. Levoy, *A Volumetric method of building complex models from range images*, Proceedings of SIGGRAPH, pp. 303–312, 1996.

S. Danijel and L. Ales, *Range image acquisition of objects with non-uniform Albedo using structured light range sensor*, Proceedings of the international conference on pattern recognition, pp. 10–46, 2000.

R. Fabio, *From point cloud to surface: The modelling and visualization problem*, International archives of the photogrammetry, remote sensing and spatial information sciences, xxxiv,2003.

FarField Technology, *FastRBF toolbox – Matlab interface*, version 1.4, FarField Technology, 2001.

A. Ghatak, *Optics second edition*, McGrawHill, New Delhi, India, 2002.

R. C. Gonzalez and R. E. Woods. *Digital image processing*, Eddison-wessley, New York, NY, 1992.

Horn *et al.*, *Shape from shading*, MIT, Cambridge, 1989.

D. F. Huber and M. Hebert, *Fully automatic registration of multiple 3D data sets*, image and vision computing, 21, pp. 637–650, 2003.

J.R Kender, *Shape from texture*, Proceedings DARPA IU workshop, 1978

R. Klette, K. Schluns and A. Koschan, *Computer vision three dimensional data from images*, Springer, Singapore, 1998.

X. Li and W. G. Wee, *Range image fusion for object reconstruction and modeling*, Proceedings of the first Canadian conference on computer and robot vision, 2004

Y. Ma, S. Soatto, K. Jana and S. Shankar . *An invitation to 3-D vision from images to geometric models*, springer, New York, NY, 2003.

J Mitra , G Natasha, P Helmut and G Leonidar, *Registration of point cloud data from geometric optimization perspective*, Siggraph symposium on geometry processing, pp. 22–31,2004.

T. Miyasaka, K. Kuroda, M. Hirose and K Araki, *Reconstruction of realistic 3D surface model and 3D animation from range images obtained by real time 3D measurement system*, Proceedings of the international conference on pattern recognition, 2000.

W. M. Newman and R. F. Sproull. *Principles of interactive computer graphics*, second edition, McGrawHill, Singapore, 1979.

P. Neugebauer, *Reconstruction of real world objects via simultaneous registration and robust combination of multiple range images*, International journal of shape modelling, 3, pp. 71 – 90,1997.

A. D. Payne, D. A. Carnegie, A. A. Dorrington, and M. J. Cree, *A Synchronised Direct Digital Synthesiser*, Proceedings of the First International Conference on Sensing Technology, Massey University, pp. 174–179, 2005.

A. D. Payne, D. A. Carnegie, A. A. Dorrington, and M. J. Cree, *Full Field Image ranger Hardware*, The third IEEE International Workshop on Electronic Design, Test & Applications, pp. 263–268. 2006

G. Pujitha, S. Hisanari and S.Yukio, *3-D face modelling with multiple range images*, Proceedings of the international conference on pattern recognition, 2000.

Russ C J, *The image processing handbook*, CRC press, florida, 2002.

URL References (August 2006):

DDS

<http://www.ehb.itu.edu.tr/~eepazarc/ddstutor.html>

Distortion

<http://vision.fe.uni-lj.si/docs/janezp/pers-wwk2002.pdf>

http://groups.google.co.nz/group/sci.image.processing/browse_thread/thread/5dc302b5ef578159/62038c6a25e212cb?lnk=st&q=radial+distortion+image+coordinates&rnum=3&hl=en#62038c6a25e212cb

Grabbers

http://www.coreco.com/Web/home.nsf/MainFrame/MainFrame?OpenDocument&L_FS=Gb&C_FS=PgTopSecondGb&T_FIG=pg&C_FIG=PgNavNewsEventsGb&db_FID=news.nsf&T_FID=doc&C_FID=14462EB50B2A5D8485256A4C0072B5AB?OpenDocument&L_FID=Gb

Image acquisition

<http://zone.ni.com/devzone/devzone.nsf/webcategories/C4FD2B63DB961E7F8625685E00779388>

Image intensifier reference sites:

<http://www.proxitronic.de/prod/bv/eein.htm>

<http://www.hpk.co.jp/eng/products/ETD/iie/iie.htm>

<http://www.answers.com/topic/image-intensifier>

Projection

<http://www.siggraph.org/education/materials/HyperGraph/viewing/view3d/perspect.htm>