ACQUISITION AND MODELING
OF
3D IRREGULAR OBJECTS

A THESIS SUBMITTED

TO

THE GRADUATE SCHOOL

OF

THE CHINESE UNIVERSITY OF HONG KONG

BY

SAI-BUN WONG

(Student ID - 90371280)

IN

THE DEPARTMENT OF ELECTRONIC ENGINEERING

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

MASTER OF PHILOSOPHY

JUNE 1994

Acquisition and Modeling of 3D Irregular Objects

Table of Contents

	Abstract	v
	Acknowledgment	vii
1	**Introduction**	1-8
	1.1 Overview	2
	1.2 Survey	4
	1.3 Objectives	6
	1.4 Thesis Organization	7
2	**Range Sensing**	9-30
	2.1 Alternative Approaches to Range Sensing	9
	2.1.1 Size Constancy	9
	2.1.2 Defocusing	11
	2.1.3 Deconvolution	14
	2.1.4 Binolcular Vision	18
	2.1.5 Active Triangulation	20
	2.1.6 Time-of-Flight	22
	2.2 Transmitter and Detector in Active Sensing	26
	2.2.1 Acoustics	26
	2.2.2 Optics	28
	2.2.3 Microwave	29
	2.3 Conclusion	29
3	**Scanning Mirror**	31-47
	3.1 Scanning Mechanisms	31
	3.2 Advantages of Scanning Mirror	32
	3.3 Feedback of Scanning Mirror	33
	3.4 Scanning Mirror Controller	35
	3.5 Point-to-Point Scanning	39

	3.6 Line Scanning	39
	3.7 Specifications and Measurements	41
4	**The Rangefinder with Reflectance Sensing**	48-58
	4.1 Ambient Noises	48
	4.2 Occlusion/Shadow	49
	4.3 Accuracy and Precision	50
	4.4 Optics	53
	4.5 Range/Reflectance Crosstalk	56
	4.6 Summary	58
5	**Computer Generation of Range Map**	59-75
	5.1 Homogenous Transformation	61
	5.2 From Global to Viewer Coordinate	63
	5.3 Z-buffering	65
	5.4 Generation of Range Map	66
	5.5 Experimental Results	68
6	**Characterization of Range Map**	76-90
	6.1 Mean and Gaussian Curvature	76
	6.2 Methods of Curvature Generation	78
	6.2.1 Convolution	78
	6.2.2 Local Surface Patching	81
	6.3 Feature Extraction	84
	6.4 Conclusion	85
7	**Merging Multiple Characteristic Views**	91-119
	7.1 Rigid Body Model	91
	7.2 Sub-rigid Body Model	94
	7.3 Probabilistic Relaxation Matching	95
	7.4 Merging the Sub-rigid Body Model	99
	7.5 Illustration	101
	7.6 Merging Multiple Characteristic Views	104

7.7 Mislocation of Feature Extraction 105

7.7.1 The Transform Matrix for Perfect Matching 106

7.7.2 Introducing The Errors in Feature Set 108

7.8 Summary 113

8 **Conclusion** 120-126

References 127-131

Appendix A - Projection of Object A1-A2

Appendix B - Performance Analysis on Rangefinder System B1-B16

Appendix C - Matching of Two Characteristic views C1-C3

Abstract

Range image is referred as the two dimensional matrix of distance values in the function of z = f(x,y), where z is the distance, and x,y are the spatial components. Traditional methods for machine vision were extensively dealt with the passively sensed intensities. However, it is advantageous [14] for the machine perception to deal with the actively probing data rather than the passively sampled data. This leads to our interests for the researches in design of rangefinder and processing of range imaging.

The data taken by a typical rangefinder is not a true 3D representation of the original object since it only acquires a partial view of the object. Such range images contain information more than 2D but not enough for 3D representations which are typical examples of 2½D data. Some information may be available in one view but absent when viewing from another angle. Also, self-occlusion and/or mutual occlusion cause the "missing parts" on the range images. The problem leads to the question of whether it is possible to get a global representation of a 3D object from images acquired from a rangefinder? Indeed, this involves the problems of registration of multiple range images, 3D representation, feature characterization and common view matching, which are currently addressed by many researchers.

Range sensing and range image processing become a crucial component of autonomous systems. The problem of range sensing is speed when comparing with intensity sensing. Snapshot of intensity in CCD camera involves just only 20 mili-seconds for CCIR-50Hz. However, range sensing may take from minutes to hours for one single frame. As range image processing requires multiple frames, the analysis and verification would be so time-consuming. With the advancement in range sensing technology, rangefinder can be made faster, more accurate, and compact. These advances in the range sensing technology also alleviate the burden of range imaging processing.

This research project address both range sensing and range image processing. A rangefinder is designed using the active triangulation method with x-y scanning mirrors for capturing both the range data and the intensity. The acquired range images which are just as partial views of 3D objects. Methodology to establish the relationship between each pair of characteristic views and then merge them into a single 3D coordinate system for full representation is investigated.

The thesis presents an active triangulation rangefinder which employs the x-y scanning mirror as the scanning mechanism. We would also present a software tool to simulate such process of generating range images. A rigid body relaxation matching process is designed in order to obtain the mapping function of two characteristic views. Finally, we have also developed the methodology for merging those characteristic views into a single 3D object. To analyze our model and merging algorithms, a visualization system has been developed on the SGI[1] workstation such that :-

a. In order to make the analysis more accurate and faster and to isolate the uncertainties on realistic generation of range data, a set of tools were developed to generate, analyze the range images from global 3D objects.

b. Features can be selected interactively. Also these features could be visualized in the 3D global coordinate system.

c. The mapping implemented by the matching function can be visualized in the 3D global coordinate system. The mapping could also be modified interactively to evaluate the matching accuracy.

d. The results of merging multiple characteristic views can be visualized by utilizing the results of relaxation matching.

[1] SGI - Silicon Graphics, Inc. The tools are described on operational manual separately.

Acknowledgment

I would like to thank my supervisor, Dr Tong Lee, for his helpful advises and the directions of the whole research works. And also, thanks for his patience in correcting the grammatical mistakes and misspelling in my thesis. And also, Dr H T Tsui, thanks for his great concerns about all my progress in the research although he was so busy all the time. Dr John Chiang also gave me helpful comments and supports for the application programs.

I must thank Mr D W Chan, our laboratory technician, for helping me to set up and maintain the equipment. And thanks to our metal workshop, without their helpings, the project cannot be completed. Thanks to Fu in Boston for sending me the reference papers.

By all means, I am truly grateful to my father and mother for their supports and encourages along the days in my study. Finally, I must sincerely thank those not listed in the above.

Chapter 1
Introduction

Traditional computer vision deals with the static 2D intensity images only. However, for many cases, the applications may require information other than just only the 2D intensity. They may want the distances and the 3D coordinates of objects. For example in a vehicle navigation system, the vehicle must be able to recognize its location and navigate to desired locations without running into obstacles. By all means, range images can provide the distance map to the obstacles. Thus, it can plan a specific path which can avoid to collide with the unexpected objects. While, the intensity can only provide the reflectance of obstacles. Similar patterns of reflectance may exist for both paths and obstacles. Intensity images cannot be used to determine the distances and hence not applicable for vehicle navigation system.

We are exploring a new approach for 3D computer vision in the thesis :- range imaging. Range images may be termed by many other names depending on the context : range map, depth map, depth image, range picture, rangepic, 3-D image, 2½-D image, digital terrain map (DTM), topographic map, 2½-D primal sketch, surface profiles, xyz-point list, surface distance matrix, contour map, and surface height map [13]. The range image is defined as the image containing two dimensional array of distance values. It can be considered as contaminated with noise, discrete samples of some surface described by the function of $z=f(x,y)$, where z is the distance values and x, y are the spatial locations [1]. The range image can be obtained by various methods : laser beam scanning, stereo vision, microwave radar, sonic/ultrasonic radar [9]. For many robotic applications, the response must be quick in order to react to the events, for example in vehicle avoidance and ship navigation. It is important therefore to acquire the range image in a quick and accurate manner, and the equipment must be applicable for real-time operations. Many advanced sensors have been developed and used in real-time computer vision and robotics applications such as obstacle avoidance

[5,8,9], autonomous navigation of mobile robots [21], archaeology[22], land survey, human face recognition[20], automatic assembly line, parts molding.

1.1 Passive vs. Active Sensing

Recent researches in computer vision dealt with actively probing data instead of the passively sampled ones [16]. With active sensing, a sensor transmits energy into the environment, and the information is measured from the reflected signals. Since an active sensor is missing in the passive approaches, techniques such as focusing, shading, and stereo vision are required to measure the information from the environment only [16]. Those systems are less reliable because the ambient illumination may be contaminated and disturbed by the third party and they generally required tremendous computational power. Hence, they do not usually find applications in the real-time robotic systems.

Generally, the basic distinction between passive and active systems is that : the active sensing is transmitting energy into environment and detecting the reflected signal from the object, whereas the passive sensing relies on the ambient illumination only. Krotkov [14] stated in a different way, "We prefer a different definition, and believe the use of active sensors is not a necessary condition on active sensing and that active sensing can be performed with passive sensors (that only receive, and do not transmit, energy), which are employed dynamically. Here the term 'active' refers not to an energy-emitting sensor, but to a passive sensor employed in an active fashion, with sensing strategies purposefully changing the sensors' state". Hence, the active sensing is not only strictly defined as active sensor that emitting energy to environment. We take an example for which the passive sensing is changed into active sensing. CCD camera for intensity capturing is one kind of passive sensing since it

does not emit any energy to object. However, we can use a pair of stereo cameras instead of a single camera. The distances of objects can be determined by actively adjusting the view angle until two views on the cameras are matched. Now, the passive sensing is employed in an active fashion with sensing strategies dynamically changed by the sensor's state. We therefore put the passive CCD cameras into the active stereoscopic range sensing.

Range sensing can also be classified into the passive and active sensing. Since the passive sensing does not depend on the external illumination, so an external light source is not required. Passive methods simulate the human visual system to obtain the range data and are usually slow. Examples in passive sensing include the stereo disparity [9,25,26], binocular convergence [9,43], texture gradient [22], focus deconvolution [24], size constancy [22], field of view [9], occlusion and shading [27,28].

The active methods are employed by the systems that project light sources into environment to help the determination of range values. Such systems usually require less computational power and less acquisition time when comparing with those employing the passive methods. However, the field of view and object distance are limited by the optical power. The active range sensing includes the triangulation [16,17], defocusing [14], time-of-flight [10], interferometry [23].

The limitation of active sensing is the requirement on the external source of energy. Generally, the source power is proportional to the distance of measurement

and the field of view. The source may also be contaminated by the ambient noise. So, the power of the source must be large enough for compensating attenuation and the noise. Since the passive method senses only the reflected energy by ambient source which does not require any external source, this method is applicable when the object is at far distance or the external source cannot be used.

With comparison of passive and active sensing, passive sensing generally requires a large amount of computational power and hence the generation of range map is extremely slow. And also, passive sensing usually requires a prior knowledge for analytical determination, e.g. size constancy, pre-defined features and reference image. On the contrast, the active sensing does not require any analytical operations on calculating the range values. Therefore, generation of range map could be achieved very quickly. Consequently, active range sensing is widely used in the real-time robotic applications.

1.2 Survey

Typical examples of active rangefinder were developed by Environmental Research Institute of Michigan (ERIM) [5,15]. ERIM has developed three rangefinders, namely, (1) the Adaptive Suspension Vehicle (ASV, 128X128), (2) the Autonomous Land Vehicle (ALV, 256X64), and (3) the Intelligent Task Automation (ITA, programmable up to 512X512)[13]. Basically, they are amplitude-modulated imaging laser radars. These imaging systems emit a laser-beam to the object and measure the phase difference between source and reflected signal. The source of laser-beam is a laser diode operated at 100mW. These types of rangefinders are mainly used for acquisition of 3D terrain data. where the measuring range is typically long and the accuracy is relatively low which is 61mm over 9.75m for ASV [13].

Perceptron [6] builds a commercial 3D laser radar imaging systems (LADAR). One of their products is the LADAR datacamera (simultaneously range data and intensity image). They have only two models; the 50mW model (which used the AsAlGa laser) and 7mW model with ranges to 70m and 40m respectively. They use the light striping and the Scheimflug condition [22], which makes the object, image, and lens planes all coincide along a single line. The accuracy is 25μm over 45-mm depth of view at a rate of 15 points per second. However, there are a number of problems associated with the LADAR. For examples, the surface reflectivity is a great problem in the sense of peculiarity and non-homogeneity. The reflectivity can create the cross-talk among itself or with other which will be discussed the following chapters in more details.

CUHK [17] developed a rangefinder which used the triangulation method. They adopted the x-y table to scan the object which is known as moving object scanning. The total acquisition speed is very low because the object is sampled in point-by-point basis and the traveling time of x-y table is very long, usually about 2 hours per 128X128 image. Alternatively, the x-y table can be replaced by using the x-y scanning mirror which could complete the scanning quickly. The scanning mechanism will be discussed later in the chapter 3.

The National Research Council of Canada (NRCC), have also developed the X-Y laser scanning systems which use the triangulation approach and give very high precision out of 5m. Their accuracy can be made up to 25μm with sampling rate of 600KHz [11]. Since they are using the x-y scanning mirror as the control, the sampling can be achieved in very high speed and at high accuracy, when comparing with CUHK's which could provide 128x128 points in 10cm X 10cm area for total time of 2 hours.

Cyberware [19] also manufactures their color 3D digitizers. Since they use the low-power laser, they could be used for human body scanning. A digitizer is used to collect the color information via a second video sensor to provide the range image along with the color value. Since they project the stripping light of a line rather than a point, the scanning speed can be 15,000 samples per second. Their scanning mechanism is quit different, which uses the rotary table rather than x-y table. The total acquisition time of a whole scanning is about 16.7s and the scanning volume about 300mm (radius) X 300mm (height).

From the above examples, we can make a summary: firstly, they all actively emit energy to the scenes and sense the reflecting signal rather than only sense the ambient illumination. Such direct sensing increases the attention as a crucible alternative to passive sensing. Secondly, with exception of ERIM, the triangulation seems to be the most common and accurate method in range sensing. Thirdly, NRCC is a successful example for using x-y scanning mirror for increasing their scanning speed. Finally, Cyberware's rangefinder is using the stripping lighting source to speed up their acquisition. These factors come to the conclusion :- our rangefinder should be based on the active triangulation sensing by using the stripping light with x-y scanning mirror as the control mechanism.

1.3 Objectives

The aim of the research is to determine the process of acquisition and modeling of 3D irregular object. Acquisition would concern about building a rangefinder which can capture the range images of real objects. The modeling involves merging the range images for different characteristic views, and representing them into a single global coordinate system.

First of all, the project will investigate the feasibility for building a low-cost rangefinder for both intensity and range sensing. The rangefinder will be built by the

requirement of high speed, high accuracy, and close distance to object. Based on the surveys on section 1.1, the rangefinder would use the x-y scanning mirror to speed-up the scanning process with stripping light source instead of point source.

Secondly, having acquired the range images, the second phase is to extract the 3D features and analyze them. The acquired images are not really true 3D images because they may be shadowed by the object itself or another object. However, the missing features may be revealed after changing the viewing angle or position. Theoretically, if we can locate the common portions from the individual characteristic views, we can apply the homogenous transformation to obtain a single global object. Our aim is to recover the 3D object by multiple characteristic views at different observing positions for the problem of this self-occlusion.

Finally, in order to evaluate how well the process merges multiple views of range image into a single and global coordinate system, we are taking a reverse approach to simulate the process of range and intensity acquisition. A number of views are generated by the computer according to different viewing angles and positions. The viewing angles and positions are also recorded and will be used for verification. We then try to use our developed system to solve for the corresponding homogenous transform and merge the different views into the global coordinate system which will be compared with the original model.

1.4 Thesis Organization

The thesis is mainly divided into two parts, the hardware design and the applications. In chapters 2 to 4, we will discuss the acquisition of range sensing and chapters 5 to 7 describe the modeling and process of the range imaging. In chapter 8, we will conclude on the present works and summarize our contribution for the current research project.

In chapter 2, the theories and operations about different approaches of range sensing are reviewed and selection of the technology of transmitter and detector is considered. In chapter 3, design of the scanning mechanism and controller for the application is discussed. Chapter 4 gives the analysis of system performance, accuracy and precision for the range sensing system. Next is the software portion. In chapter 5, the software to generate the range maps according to the arbitrary viewing coordinates and angles is developed. While in chapter 6, use of differential geometry to extract features from a range map is discussed, chapter 7 illustrates how to use the over-determined system on merging of characteristic views by the homogeneous transform matrix between two different characteristic views. Finally, chapter 8 would conclude the works and contributions in the range imaging system.

Chapter 2

Range Sensing

At the last chapter, we have already discussed the general idea on the new age of computer vision - range imaging. This chapter will discuss and analyze different alternatives to the range sensing in order to verify the most suitable approach for our research. The rangefinding technologies could be classified into size constancy, defocusing, deconvolution, binocular vision, active triangulation and time-of-flight. We will first give a brief descriptions on those approaches for rangefinding. Then, we also examine how different technologies in transmitters and detectors are used for range sensing.

2.1 Alternative Approaches to Range Sensing

2.1.1 Size constancy

One of the methods in passive range sensing is determined by the size constancy [22]. Before getting the distance values, the actual size must be known at a pre-defined distance. The basic principle of the size constancy is that : the object would get smaller as we are moving the object farther, and vice versa.

From figure 2.1, we can obtain the distance by relating : -

$$Z_o = \frac{CZ_i}{X_0}$$

Equation 2.1

where, Z_0, Z_1 = the distance between lens and object;

Z_i = the distance between lens and detector;

C = the object height (constant);

X_0, X_1 = the image height of object at distance Z_0 and Z_1.

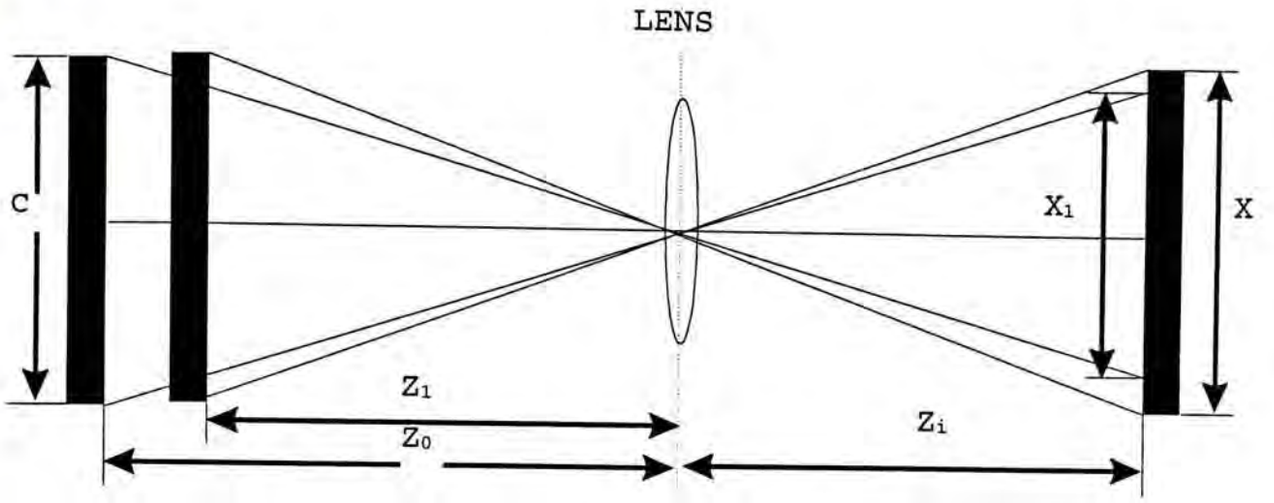


Figure 2.1-Range sensing by size constancy.

Next, we are going to evaluate the accuracy of the system. If the smallest unit on detector is $\delta X = |X_0 - X_1|$, then the corresponding measurement quantity of distance would be $\delta Z = |Z_0 - Z_1|$.

$$Z = \frac{Z_0^2 \delta X}{C Z_i}$$

Equation 2.2

This quantity δZ is hence the resolution of the range value because it means that the measurement will not change if we move the object from Z_0 to Z_1. It happens that the resolution is a function of Z_0^2. Normally, Z_i, δX and C are constants. As the object moves farther, Z_0 becomes larger. δZ will also get larger. In another word, the resolution worsen as we measure the farther range. Size constancy is very similar to the triangulation systems which will be discussed later. The only difference is that, we must first know the size of object before acquiring the distance.

2.1.2 Defocusing

Defocusing is actually some kind of triangulation methods. However, the lens is specially arranged. As shown in figure 2.2, a double slit mask is inserted between two lens and the solid line is showing the ray traces from a base plane (A) to the detector plane (A'). As surface (B) is moved toward the lens, the image should be focused on the plane of dotted line - B', and the image of two points (D'-D') would be formed in the detector plane. Separation between the D'-D' is related to the distance from the camera lens to the target surface. The detector can be typically a camera, a line scanner, or a matrix of CCD, etc., which should be capable of digitizing the separation D'-D'.

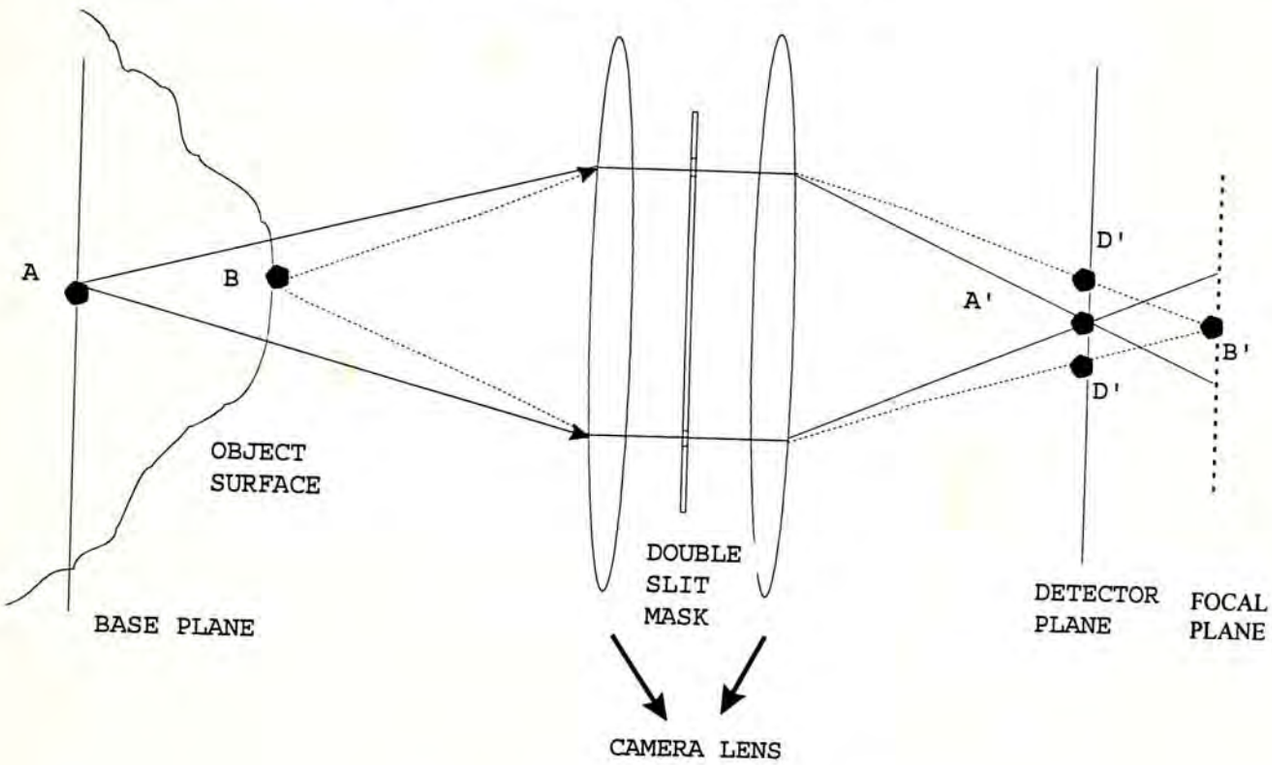


Figure 2.2-Double slit mask lens camera.

To analyze the system, we can express the two equations according to reference plane and target surface to the lens :-

$$\frac{1}{f} = \frac{1}{B} + \frac{1}{B'}$$

Equation 2.3

$$\frac{1}{f} = \frac{1}{A} + \frac{1}{A'}$$

Equation 2.4

where, f = focal length.

From equation 2.3 & 2.4, we can calculate B as :-

$$B = \frac{f(AD - A'S)}{AD - A'S - fS}$$

Equation 2.5

where D = distance of two spots (D'-D'); B = object distance;

S = slit separation in double slit mask; f = focal length;

A = base plane distance.

We can plot the relationship for D against B in figure 2.3 which is a function of 1/D. For some range of D, say greater than 1mm, the separation of two slit changes greatly when the distance changes for only a little. Thus, the resolution of measurement would be the best in this range. Meanwhile, the D changes only a little when the distance to measure is farther than 200mm. Therefore, as indicated in figure 2.3, the range values for the defocusing system is good for some range of 0.2m to 1m and is the finest in the range greater than 1m. However, we could not measure it when distance is smaller than 0.2m.

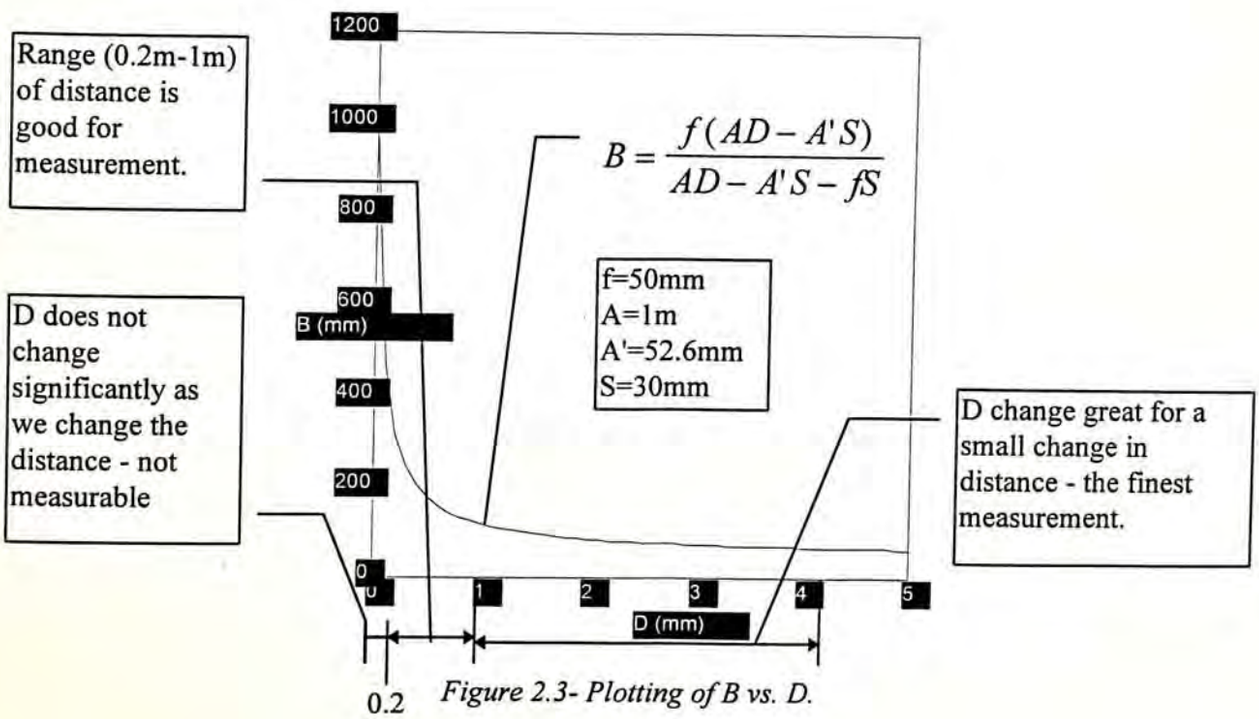


Figure 2.3- Plotting of B vs. D.

The separation S in the mask may also alter the valid range of range sensing. Figure 2.4(a) and 2.5(b) are showing the systems with two different masks.

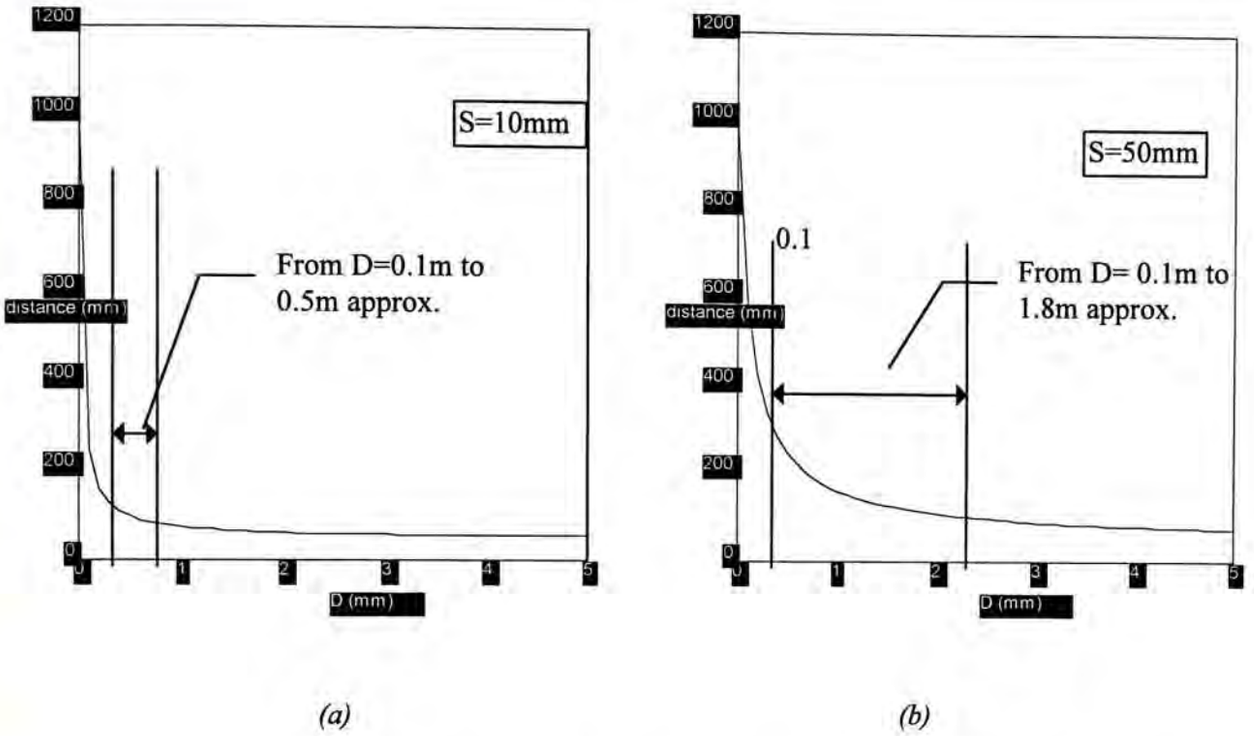


Figure 2.4- Comparison of valid range for different separation in mask.

From figure 2.4, greater S should give a better result of beam splitting and hence better precision on range value. However, greater S may attenuate the light source intensity. The energy is not homogeneously distributed on the lens which is concentrated near the center of lens. Therefore, the light on the outer bound may be very weak.

The double split mask is actually used in NRCC laboratory [12]. The system used the laser projections of stripes rather than a point to acquire the range values in a high speed manner. The data rates are proportional to the number of projected lines times the TV line rate (15.75KHz). The laboratory prototype acquires and processes at very high speed of eight lines per frame which is providing more than 10^5 3-D coordinates per

second. Resolution of this type of range sensing is quit good which is in the order of one milli-meter.

2.1.3 Deconvolution

Deconvolution is a passive technique of range sensing. The concept of deconvolution is very similar to the method of defocusing by taking two images, one in-focused and the other one misfocused to get minimum and maximum contrast respectively. Since the contrast of the features is related to the distance of the object, we can therefore compare the images with maximum and minimum contrast to compute the distance. This system works with the objects of full feature but cannot work on the objects without any feature. This approach is sometimes known as *depth from focus* [24].

Let us consider the system of defocusing in figure 2.5. It is showing a light source projecting at infinity to the camera and is misfocused on the detector plane. The defocusing operator h(x,y), where x and y are the coordinates in the detector plane, is defined as :-

$$h_r(x,y) = \begin{cases} \dfrac{1}{\pi r^2} & , \text{if } x^2 + y^2 \le r^2 \\ 0 & , \text{if } x^2 + y^2 \le r^2 \end{cases}$$

where

$$r = \frac{f d_{det} + fD - d_{det} D}{2F d_{det}}$$

Equation 2.6

where D = distance to the object; f = focal length;

F = F-number of lens; d_{det} = distance from detector plane to lens.

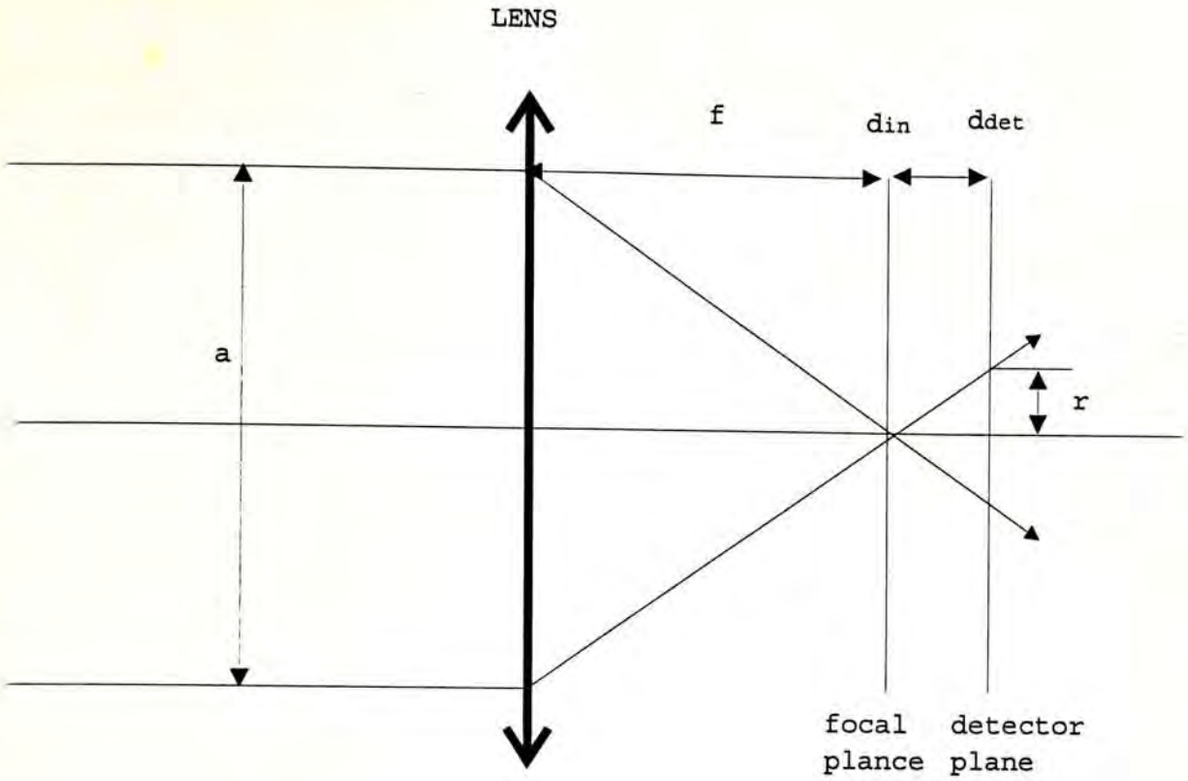


Figure 2.5-Geometry of a defocused imaging system.

The blurring effect due to the defocusing can be modeled by the convolution with the circular pillbox :-

$$h_r(x,y) * I(x,y)$$

Equation 2.7

where $h_r(x,y)$ = circular pillbox; $I(x,y)$ = the ideal image.

The deconvolution system uses two camera lens, with one large area of pillbox and one small area. In figure 2.6, the system is taking two images from two defocusing operators, i.e. $h_1(x,y)$ and $h_2(x,y)$. The acquired images on two cameras can be equated :-

$$i_1(x,y) = i(x,y) * h_1(x,y)$$
$$i_2(x,y) = i(x,y) * h_2(x,y)$$

Equation 2.8

And, the goal of this analysis is trying to find a convolution $h_3(x,y)$ so that :-

$$i_1(x,y) * h_3(x,y) = i_2(x,y)$$

Equation 2.9

And this goal can be achieved by substituting equation 2.9 into 2.8 :-

$$h_3(x,y) = \frac{h_2(x,y)}{h_1(x,y)}$$

Equation 2.10

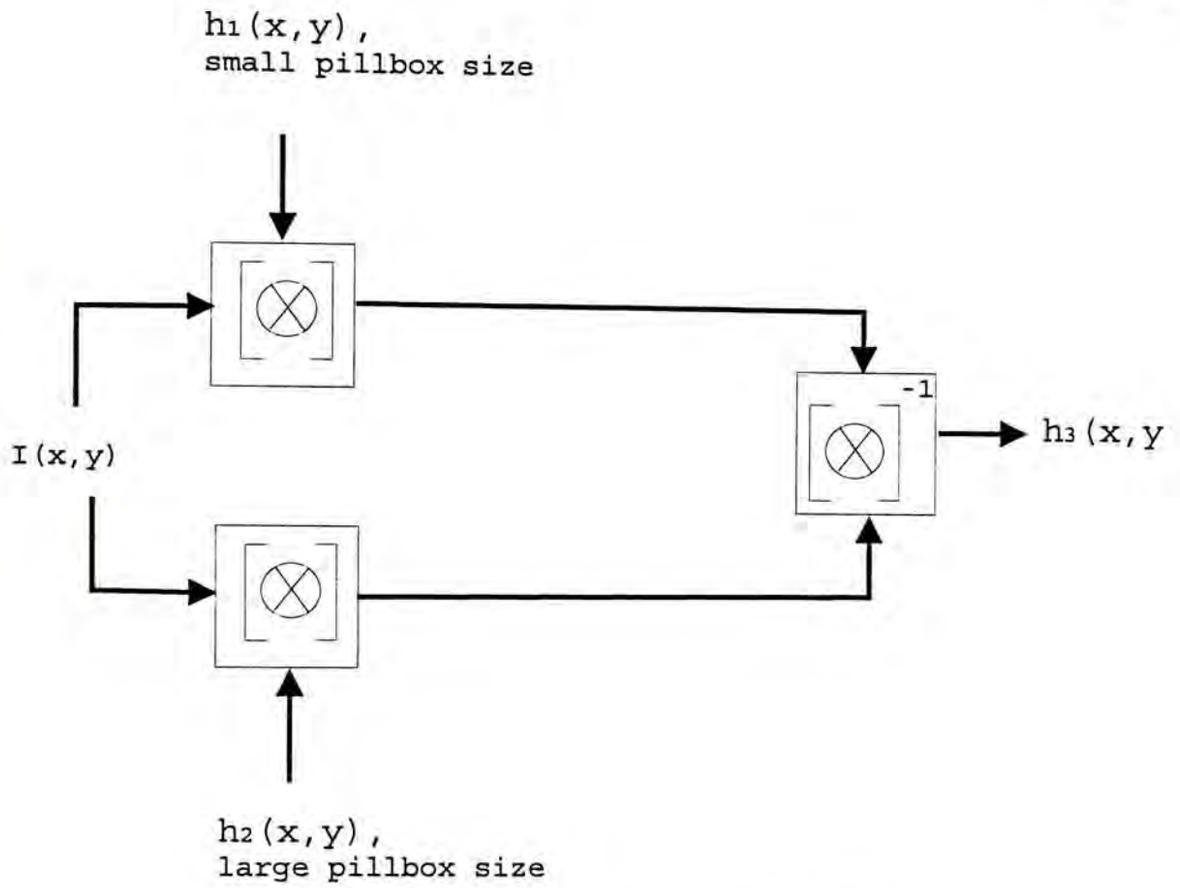
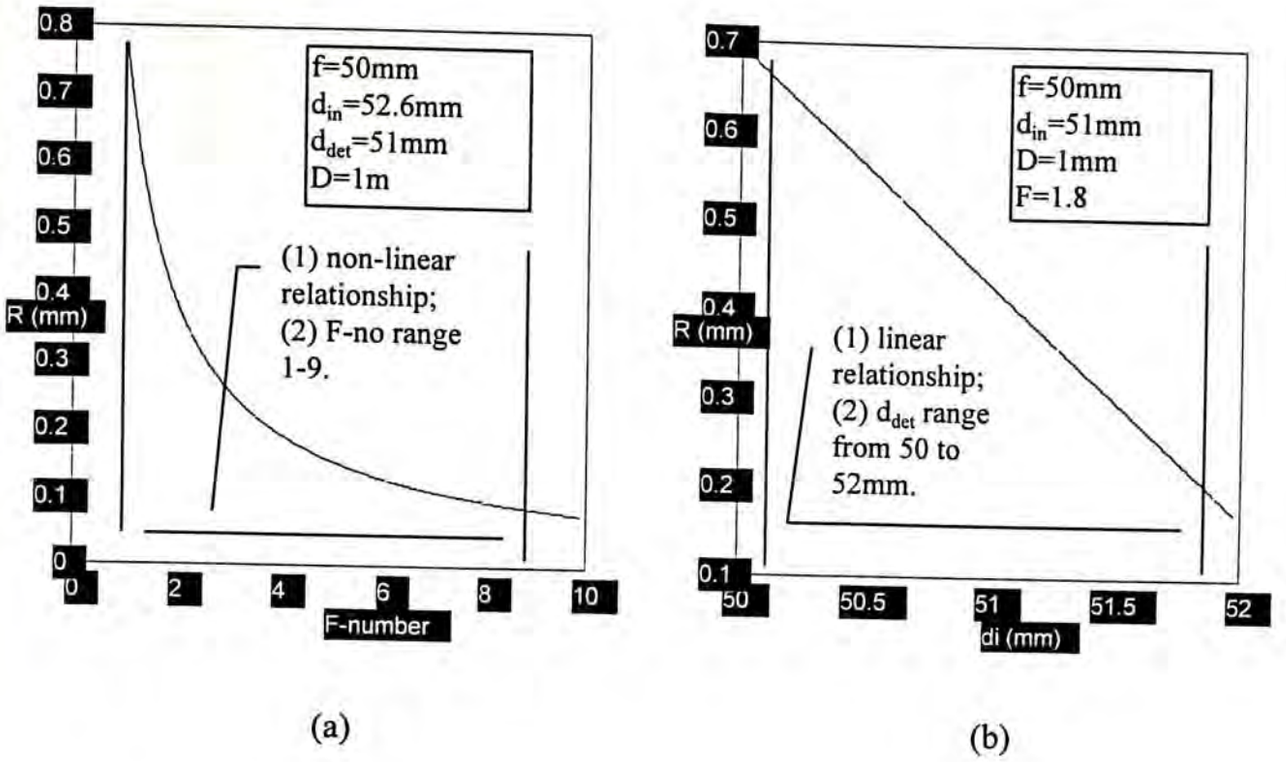


Figure 2.6-Arrangement of deconvolution.

The function $h_3(x,y)$ is so called the convolution ratio of defocusing. For any focal region, $h_3(x,y)$ is uniquely related to the distance D from the camera to that region in the scene. The convolution ratio can be determined by the matrix based method or inverse filtering [24].



(a) (b)

Figure 2.7-Comparison for deconvolution of adjusting F-number and detector plane.

Since the defocusing operator is related to the circular pillbox of radius R which can be controlled by adjusting the detector plane or the F-number according to equation 2.6. By plotting the radius of circular pillbox vs. F-number and d_{det} respectively as in figure 2.7, we can see that R is inversely proportional to F-number. Therefore, a large pillbox will be resulted for a small F-number. However, the radius will not be changed significantly as the F-number is very large, say F=8-10. For the case of adjusting the detector plane, the function is linearly proportional to d_{det} but with a negative slope. However, adjusting d_{det} is small, say within the range of 50-52mm in the figure. Consequently, the circular pillbox would be adjusted by the F-number, i.e. the aperture, rather than adjusting the detector plane.

The typical arrangement of such system is shown in figure 2.8. The half-mirror is used to split the images into both cameras. Camera 1 is adjusted to have the defocusing of $h_1(x,y)$ and camera 2 of $h_2(x,y)$ so on. Two images are then digitized and analyzed to get the deconvolution ratio, and hence the distances.

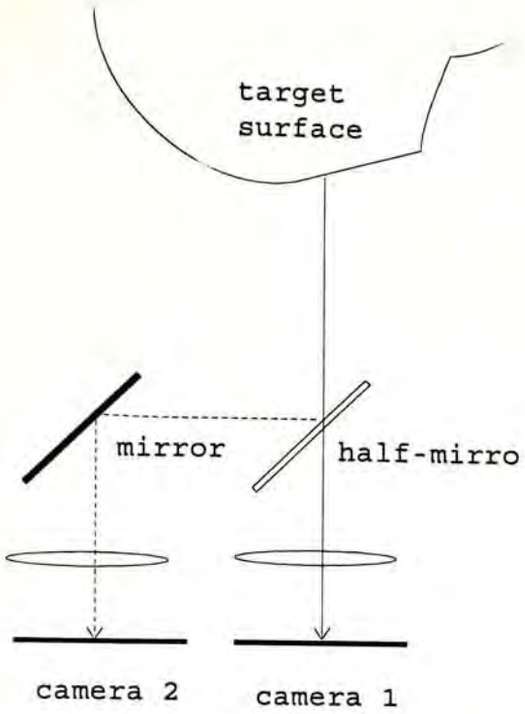


Figure 2.8-Typical deconvolution system.

2.1.4 Binocular Vision

The binocular vision is sometimes referred as stereo disparities, stereo vision or stereopsis which is a passive ranging technique by the theory of human vision system [29]. The system is capable of capturing ranges of unknown objects. Figure 2.9 is showing the system of binocular vision. A single feature on the object will be formed on the left and right camera. The lateral disparities X_1 and X_2 are two parameters determined by the distance in between the cameras and the object. Similarly, we can extract every feature point on the object and then match them from left to right images. The technique of ranging system using the lateral disparities is so called the stereo disparity.

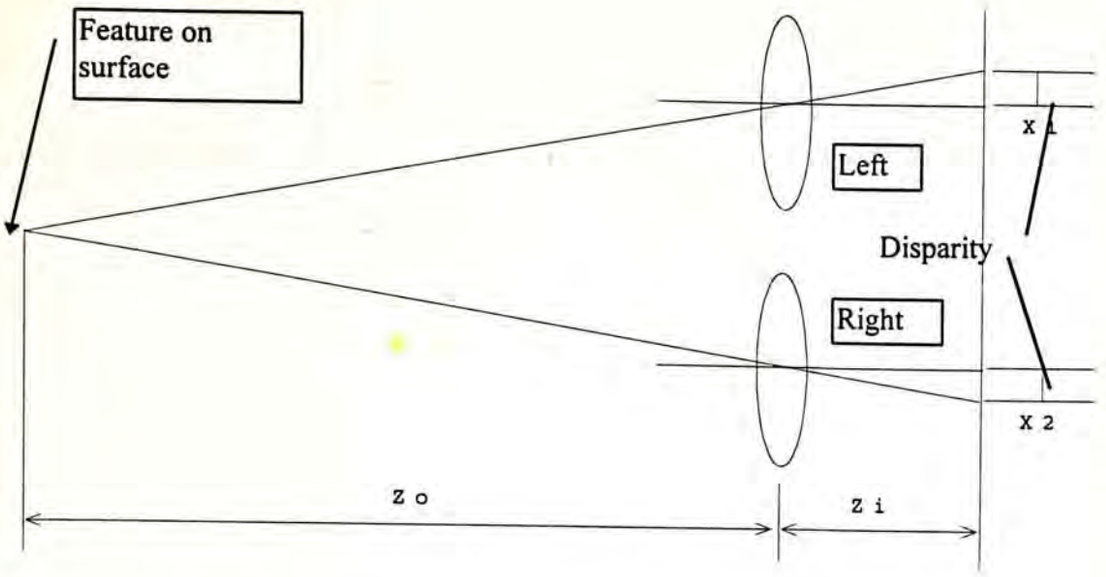


Figure 2.9-Stereo disparity system.

Likewisely, the two cameras can be rotated in the axes of viewing in figure 2.10. Therefore, a maximum number of features in two images will coincide. The two rotating angles θ_1 and θ_2 can also be used as the parameters to calculate the distances. The technique of ranging system using these two angles is so called the binocular convergence.

For some cases, the stereo disparity is only valid for some range of distances. So, practical system may use both stereo disparity and binocular convergence for obtaining the specified range of distances. The problem in binocular vision is the feature matching from one camera to another. The matching depends on the extracted features on the left and right images. The system cannot apply into the object that has no feature, for example a white wall, similarly for object with a periodic features, for example a stripped wall. Therefore, for some applications, the system would project a random patterns onto the surface. And then, the two cameras will try to match the patterns on both sides. So, the passive sensing now change towards active fashion. For such system, it may be referred as active stereopsis [32].

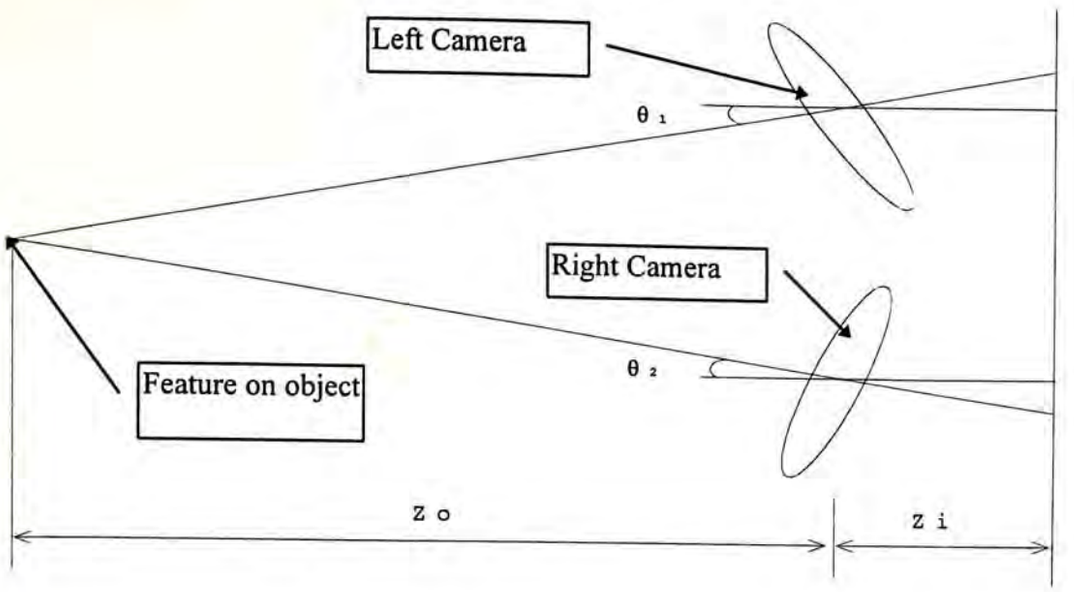


Figure 2.10-Binocular convergence system.

2.1.5 Active Triangulation

The active triangulation is the oldest yet and the most common method in range sensing. Ancient Greek and Egyptian origins have historically used this method for ship navigation, surveying, and civil engineering applications. Triangulation is based on the trigonometric method to calculate the distance and angle of the object local. In figure 2.11, for given the length of one side and two angles of the triangle, it is possible to calculate the length of other sides and determine the remaining angle.

$$B = A\frac{\sin\phi}{\sin\varphi} = A\frac{\sin\phi}{\sin(\theta + \phi)}$$

Equation 2.11

where B = the distance to the object at angle θ.

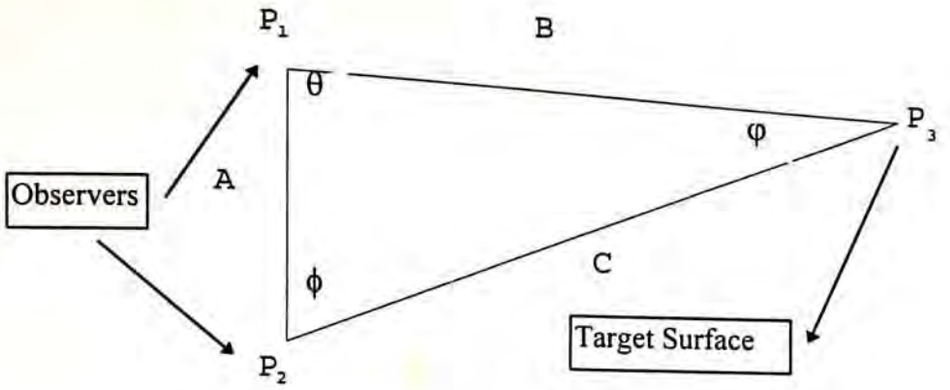


Figure 2.11-Active triangulation.

Active triangulation system would use the controlled light source in position either P_1 or P_2, such as diode laser, which is directed to the target surface P_3. And, at the opposite sides for P_1 or P_2, would place 2D array of detectors, such as CCD camera, line scanner, batch of photodetectors. A lens would be installed to focus the light source image to the detectors. Figure 2.12 shows the image formation on the detector.

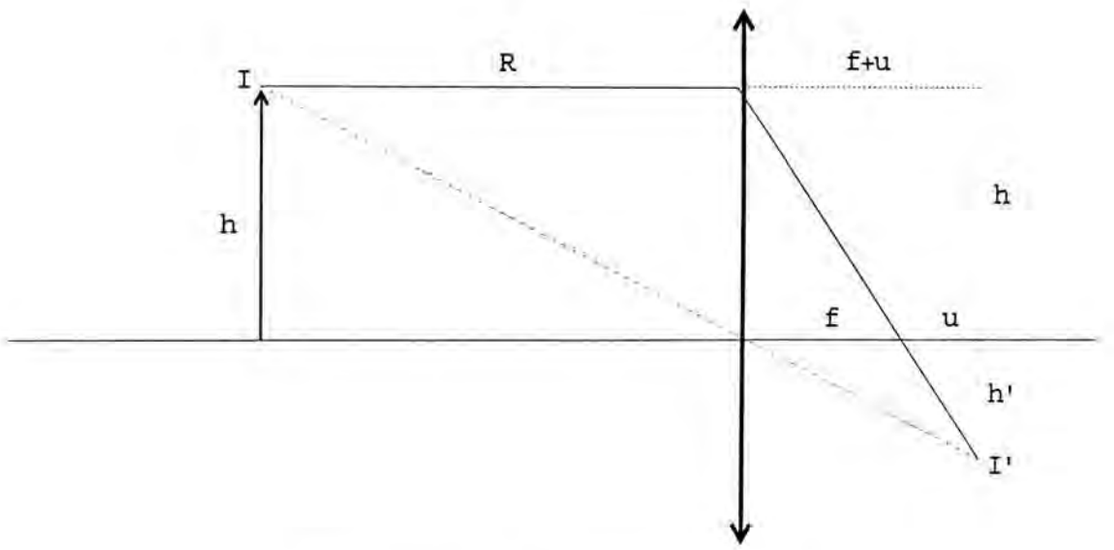


Figure 2.12-Active triangulation system.

The object I is focused as image I' on the plane at the distance f+u behind the lens. As there is an angle in between the detector and the incident path of the light, the location of the light spot may be expressed as a function of the distance :-

$$R = \frac{h}{h'}(f + u)$$

Equation 2.12

Certain limitations exist for the active triangulation system. First of all, sometimes no strike is recorded at surface with shiny reflectance or absorbance of the light. A particular portions of a scene would be observed by only one viewing location (P_1 or P_2) is known as the self-occlusion. Secondly, a trade-off exists for optimizing accuracy and area of self-occlusion. That is because the accuracy of sensing is good when distance h in figure 2.12 is long. However, the area of self-occlusion is large when h is long. Finally, according to equation 2.12, as the range R is an inverse function to spot location h', the detectable change of value in h' becomes small when R is large. The resolution diminishes when the distance of the object increases.

To obtain the three dimensional information for an entire scene, the above system can be scanned in both azimuth and elevation. The scanning can be achieved in fixing the emitter and the detector and moving the object mechanically, viz. the moving object scanning. Alternatively, the system is configured with movable optics, i.e. the scanning mirrors, adjustable camera head, etc., in order to move the laser beam, viz. the moving beam scanning.

2.1.6 Time-of-Flight

An active technique of range sensing is the time-of-flight (TOF) which corresponds to the travel time for a signal of energy emitted from a source and back to the detector. The type of energy can be sonic, ultrasonic, microwave, but usually laser diode is used for many applications in range sensing. Since the travel time is related to the distance between the emitter and the object, the range values can be detected. The operation for laser type of TOF is very similar to that of the radar. Usually, the TOF technique is applied for the environmental survey, terrain mapping, collision avoidance, etc.

Pulse is the simplest signal for range measurement. The transmitter emits the pulse to the object which strikes back to the detector. The detector would sense the time difference for the transmitting and striked back signal. Then we can apply the relation of *distance = speed X traveling time* to calculate the distance. The measurable range of distances are restricted by the speed of the energy source. If the speed is very fast, the measuring distances must be far for long enough travel time for measurement. For example when comparing with sound and light, we could see how the speed of source alters the resolution The digital circuit is generally limited up to several nanoseconds, e.g. we take the best case of 1ns. The speed of sound is 0.3 meter/ms which gives the measurable resolution of the range would be about 0.3 micro-meter. However, if the source is light which speed is $3x10^8$ meter/s, the measurable distance is 3 milli-meter.

Rather than emitting the short pulse to the object and waiting for the echo, the source of energy can be modulated by the other signal. For the laser diode, we can bias the drive current to adjust the strength of laser. The modulation signal can be amplitude-modulation (AM) or frequency-modulation (FM). f_{AM} is the frequency of a continuous wave (CW) for AM modulation. The detector is measuring the phase shift for returned signal. The range value in AM system is :-

$$d = \frac{\lambda_{AM}}{4\pi} \Delta\phi$$

Equation 2.13

where $\Delta\phi$ = the phase difference for transmitting and received signal;

λ_{AM} = wavelength for the modulating signal.

From equation 2.13, the measurable distance must be within the wavelength since the phase difference is modolo of 2π. Certainly, the CW detection requires only very narrow bandwidth for measurement. The phase detection may only require a single frequency of f_{AM} which makes the emitter and detector much simpler. However, the pulse detection requires a wide bandwidth in the emitter and also the detector. Sometimes CW detection may not be applicable since noises exist in the environment while the signal of pulse detection could be changed to suit for the environmental characteristics.

An alternative to AM scheme is the frequency-modulation. At this time, the CW is varied with the frequencies. Usually, similar to the radar system, the frequency is modulated linearly above and below a mean frequency f_o as shown in figure 2.13.

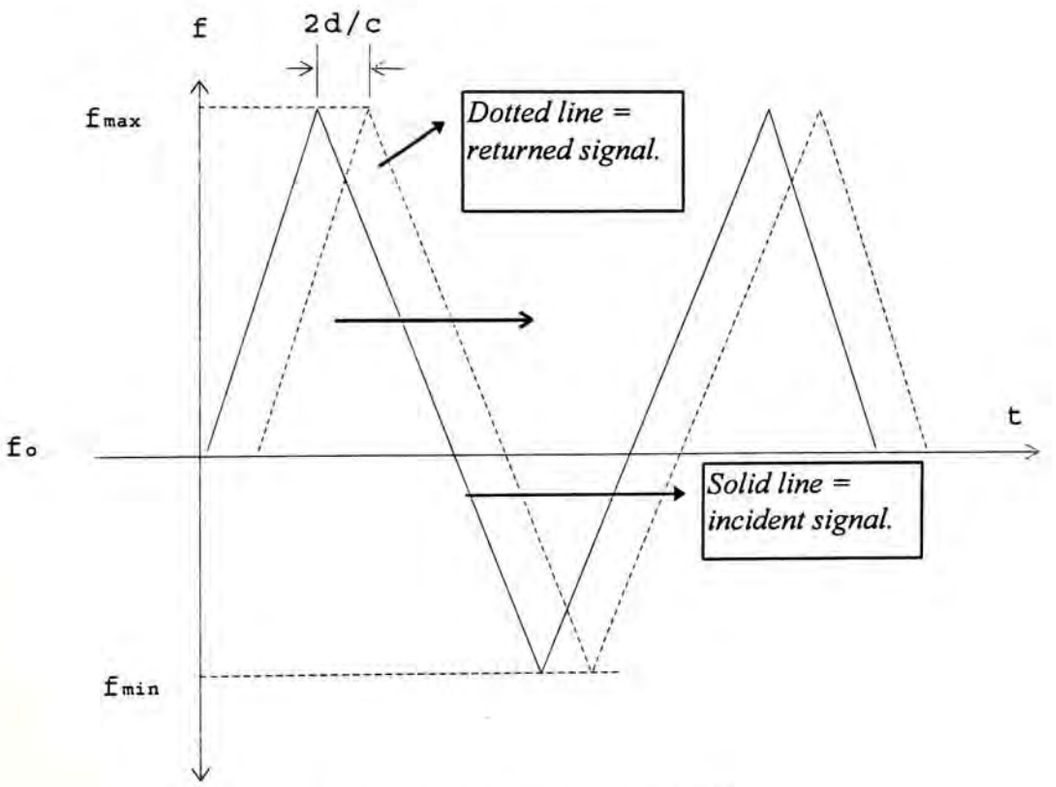


Figure 2.13-Mixer output of FM

The transmitter emits the signal that varies in the frequency as a linear function of time :-

$$f(t) = f_o + at$$

Equation 2.14

This signal is reflected from a target and then arrives at a travel time t+T.

$$T = 2d / c$$

Equation 2.15

where T = the travel time of the trip; d = distance to the target;

c = speed of the source.

There is a displacement of the received signal and the transmitting signal due to Doppler effect. These two frequencies, when combined in the mixer, produce a beat frequency :-

$$f_B = f(t) - f(t+T) = aT$$

Equation 2.16

This beat frequency can then be used to calculate the distance to the target :-

$$d = \frac{f_B \ c}{2a}$$

Equation 2.17

The advantages of TOF system arise from the direct nature of their active sensing. The emitter and detector are in the straight-line manner to transmit and receive the signal. They are generally located coaxially or with close proximity. Practically, the emitter and detector can be located on the same device. The absolute range is directly available as output with no other complicated analysis required. And, the technique is not based on any assumptions concerning the planar properties of objects when comparing to triangulation. The TOF sensors maintain the range accuracy as long as reliable signal is detected, while triangulation schemes suffer diminishing accuracy as range increases.

The drawbacks of TOF primarily relate to the properties of the emitted energy, which vary across the spectrum. The returned signal is only a small portion of the emitting energy. The returned energy on the scattered surface will be absorbed depending on the properties of the object surface and the incident angle. Also, the returned signal can be reflected from the secondary objects. These result in error of measurements of distance. The striking angle of the object surface is limited by the wavelength of the energy and the topographic characteristic of the object. For a certain angle, no signal would be returned to the detector. The above factors could also be happened to the active triangulation. Finally, the circuit for detecting the short pulse in TOF at close range and high accuracy is very expensive since the propagation delay must be short. And, the propagation delay may introduce the error for calculating the distance.

2.2 Transmitter and Detector Technologies in Active Sensing

2.2.1 Acoustics

Usually, the acoustical ranging uses the ultrasonic energy as the transmission and detection media. The performance of the system is significantly affected by environmental phenomenon and sensor design characteristic. The main factor of the sensor performance is the attenuation over distance which approximately -6dB per distance double due to the inverse square law for spherical divergence. The absorption of air also attributes towards attenuating the returned energy. The degree of absorption varies with the humidity and dust content of the air as well as the frequency of operation.

The maximum detection range for the acoustic sensor is dependent on the emitted power and frequency of operation. The lower the frequency, the longer the distance that system can operate. For example with 20 KHz operation, the absorption factor in air is approximately 0.02 dB/foot, while with 40 KHz operation is about 0.06 to 0.09 dB/foot. However, the resolution of the range sensing is good when the bandwidth of the transmitted signal is wide. Meanwhile, wider bandwidth can be achieved at higher frequency. The required operating frequency is high when minimum distance of measurement is short. Thus, wider bandwidths and greater frequencies are required as the distance between sensor and object shortens, and high precision is required.

Also, the propagation speed in the air for acoustic signal is a function of the air temperature and the wind direction. Also, the humidity may also affect the propagation speed. The actual range can be converted by a ratio of the square root of ratio of equivalent temperature T_e to calibration temperature T_c.

$$R_a = R_m \sqrt{\frac{T_e}{T_c}}$$

Equation 2.18

where R_a = actual range; R_m = measured range.

Since the temperature is usually non-uniformly distributed, the equivalent temperature is used and defined as the average temperature so that the traveling speed is constant on the air. Meanwhile, the beam dispersion is another factor for consideration when choosing the sensor. The dispersion angle is defined as :-

$$\theta = 1.22 \frac{\lambda}{D}$$

Equation 2.19

where θ = dispersion angle;

λ = wavelength of acoustic angle;

D = transmitter diameter.

As a result, the dispersion angle is very significant because the wavelength of acoustic is very long. For example, for a 30^{o} dispersion at a distance of 4.5 meters, the covering area of the beam is 0.25 meter square. This may introduce significant error factors to the spatial resolution of the measurement.

2.2.2 Optics

The active sensors in rangefinding include the broadband incandescent, narrowband LEDs, and coherent lasers. The coherent laser is commonly employed in many applications since it may be considered as the most accurate and quickest among the different types of energy.

The advantages for coherent laser are :- (1) it produces a bright, intense, output which is very important for the long-distance range sensing, and distinguishing it from the background. (2) The laser beams are narrow and collimated since it can be corrected by the optics with no divergence. This property allows the source to be highly directive and spatially selective, and hence the X-Y planar resolution precise. (3) The laser beams usually contain a single wavelength with the same phase (highly coherent). Thus, the source is purely transmitting signal without any noise. So that, we can place a narrowband optical filter for the detector to remove the unwanted striking back signal. And hence, maximizes the signal-to-noise ratio.

However, there are still disadvantages for the laser optics. First of all, since the laser beam is highly coherent source, it might cause hazards to human eyes. So, for human body scanning, extra safety measures must be made when acquiring data. Secondly, although the diode laser is easily available for the market, the gas laser can supply a large amount of optical power. However, it may require an extra high driving voltage, whereas, diode laser gives only low optical power which is for long range sensing. Finally, the material of laser is usually unstable and posses short lifetime, which results in the reliability problem.

2.2.3 Microwave

Microwave energy is ideal for the long military range sensing [9]. This is because absorbance is minimal, resolution is sufficient in the atmosphere and low-mode guiding can be constructed. The relatively long microwave wavelength provides radar systems with the capability to overcome the absorption and scattering effect of the air, weather, and other obscurants. However, they are susceptible to specular reflections at the target surface. Usually, the long microwave radar antenna occupies a large area for transmission. However, high resolution of range sensing is impossible for the shorter wave microwave. Therefore, the microwave radar is not so often applicable for vehicle avoidance, or such kind of high-resolution range sensing.

2.3 Conclusion

Our rangefinder would be operating in a closed environment which is capable of acquiring small piece of specimen range from 5cm X 5cm to 30cm X 30cm. The specimen should be placed statically during acquisition. In order to isolate the ambient noises, the rangefinder would be operating in a dark chamber.

As we discussed on the above, the active triangulation sensing have many advantages over the passive methods. First of all, the active triangulation is a direct nature of measurement although the illumination is supplied by an external light source. Secondly, the signal-to-noise ratio of active sensing is usually better. Thirdly, the active sensing does not require a large amount of computations. So, it can be used on the real-time robotics. Finally, the active sensing can give a better precision for the measurement. Hence, the rangefinder will be based on the active triangulation.

Then, the scanning mirror would be adopted such that the resolution of measurement is adjustable. Since the scanning is controlled by deflecting the incident angle of light source, the stepping size on the object would be dependent on distance to light source. Hence, the resolution becomes variable. The total number of angular steps in acquisition are fixed, say 128x128. However, the object would be allowed at a certain range of size. In another word, the spacing is adjustable when acquiring data.

Diode laser seems to be the most suitable light source. First of all, the diode laser can be operated by a low DC voltage, unlike the gas laser, where we must apply a high voltage for normal operation also the optical power is about 0.5mW to 1mW which is suitable for scanning. However, it is still hazardous to human eyes. So, safety precaution of laser product Class II should be taken for the experiments. Also, detector and emitter can be fabricated on the same device for diode laser. Thus, the transmission and detection can be achieved coaxially. Finally, no or little divergence needs to be corrected by simple optics. And therefore, the spatial resolution of the measurement can be achieved directionally. For the above reasons for diode laser source, those may also lead to be our considerations in the emitter for building our own rangefinder later on.

Chapter 3

Scanning Mirror

In the last chapter, different approaches for range sensing have been discussed. We have also chosen the active triangulation with diode laser as our design framework. In this chapter, the scanning mechanism to obtain an entire scene is considered and the design and construction of a rangefinder using the x-y scanning mirror approach are discussed.

3.1 Scanning Mechanisms

We have already discussed the basic theory about active triangulation in chapter 2 by single point basis. In order to obtain the entire scene, the object must be scanned through in both azimuth and elevation angles, or in longitudinal and transversal translations. The coordinate systems for these two types of scanning are different to each other. Figure 3.1(a) shows a system for adjusting azimuth and elevation angles. The coordinate is basically a spherical system. Obviously, we can interchange the coordinate from figure 3.1(a) to figure 3.2(b) with the transformation :-

$$
\begin{aligned}
x_s &= r_s \cos\phi_s \sin\theta_s \\
y_s &= r_s \sin\phi_s \\
z_s &= r_s \cos\phi_s \cos\theta_s
\end{aligned}
$$

Equation 3.1

From figure 3.1(a), the laser beam is controlled with the azimuth and elevation angles and moved around on the surface. The object can be scanned in the same place for the entire surface. This is referred as the moving beam scanning because we are

moving the laser beam. On the contrary, as shown in figure 3.1 (b) the object could be moved for entire scanning without changing the direction and position of the laser beam. Therefore, this mechanism is the moving object scanning.

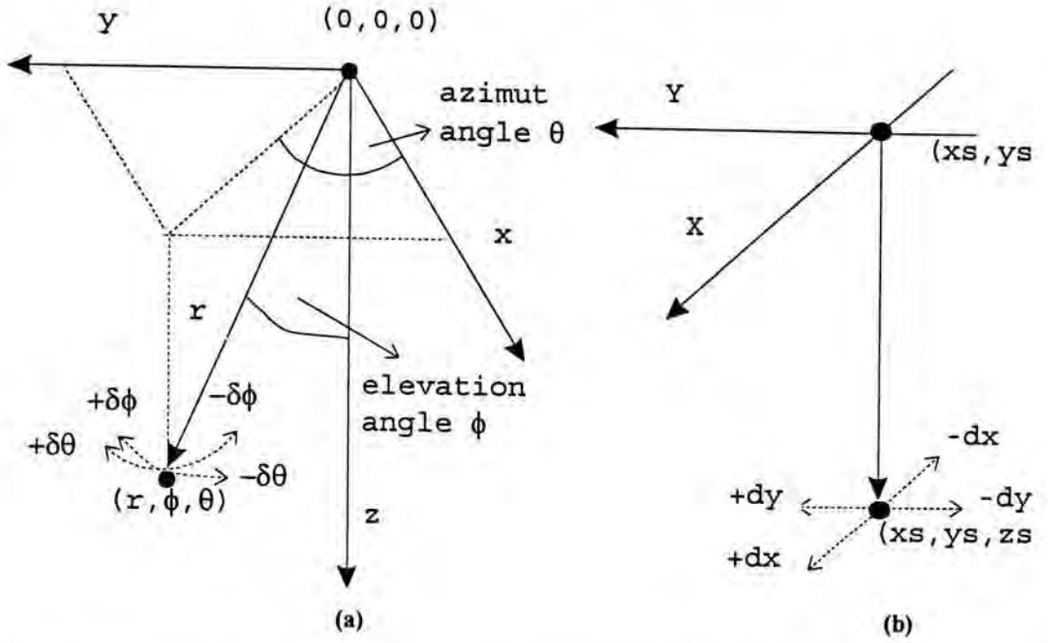


Figure 3.1-(a) Adjusting azimuth angle and elevation angles. (b) Translating longitudinal and transversal positions.

Controllable stage table is one of the common scanning mechanisms for range sensing. The stage table is driven by two stepper motors for the x and y directions. Certainly, the resolution is defined by the angular step of the motors and also the gearing. The tolerance of stepper motor is accumulative since the next step is an increment or decrement of the previous step. Therefore, an object staying on the stage table can be translated to the designating position for each point of sampling. For example of 128 x 128 points on the surface, we should translate the stage table in 16384 positions. Therefore, the errors would be accumulated for total 16384 steps.

3.2 Advantages of Scanning Mirror

Movable optics is an alternative to stage table. Scanning mirror is one kind of movable optics which is commonly used for laboratory applications [33]. The

principle for scanning mirror is to control the reflecting angle of laser beam. Travel speed of moving object is usually slow for object with large inertia or long traveling distance. On the contrary, the optical accessories are usually lighter than the object itself and only small angular adjustment is needed. Thus, the total acquisition with moving beam scanning could be achieved very quickly.

The advantages of scanning mirror are as follows : at first, since the laser beam can be randomly positioned, the accumulated error due to the last steps can be avoided. Hence, the spatial tolerance can be achieved with better precision. Secondly, a small angular change of laser beam can provide in a large position change for a limited field of view. Therefore, the scanning head can be made compact rather than a large area of table. However, as the small change in angular position can produce a great change in the x-y position, a good angular precision control circuit should be considered. Finally, since the position of projection is dependent on angle of incident and also the distance to object. The size of scanning would be adjustable by varying the distance to object. We would scan a large object with far distance and small object with close distance.

3.3 Feedback of Scanning Mirror

Scanning mirror as shown in figure 3.2 is basically consisted of two scanners head A & B which are placed orthogonally. The lower scanner - B has a mirror of smaller size and produces the x axis scan. The upper scanner - A has a mirror of larger size that reflects the x-axis in the y direction, producing the y axis scan. The laser beam may travel onto the target, or it may travel from the target into a camera as in vision applications [35].

The mirrors are driven by low inertia galvanometers with closed loop feedback. A galvanometer is a rotary servo motor with limited angle of rotation. The galvanometer can be either driven by the moving iron or moving coil. The moving iron is referred as the magnet is placed on the rotor and the coil on stator, and vice versa for moving coil. Our scanning mirror is using the type of moving coil since it is specially designed for highly linear torque over a relatively large rotation angle for a single pole motor. The rotating angle is proportional to the driving current for the this type of galvanometer.

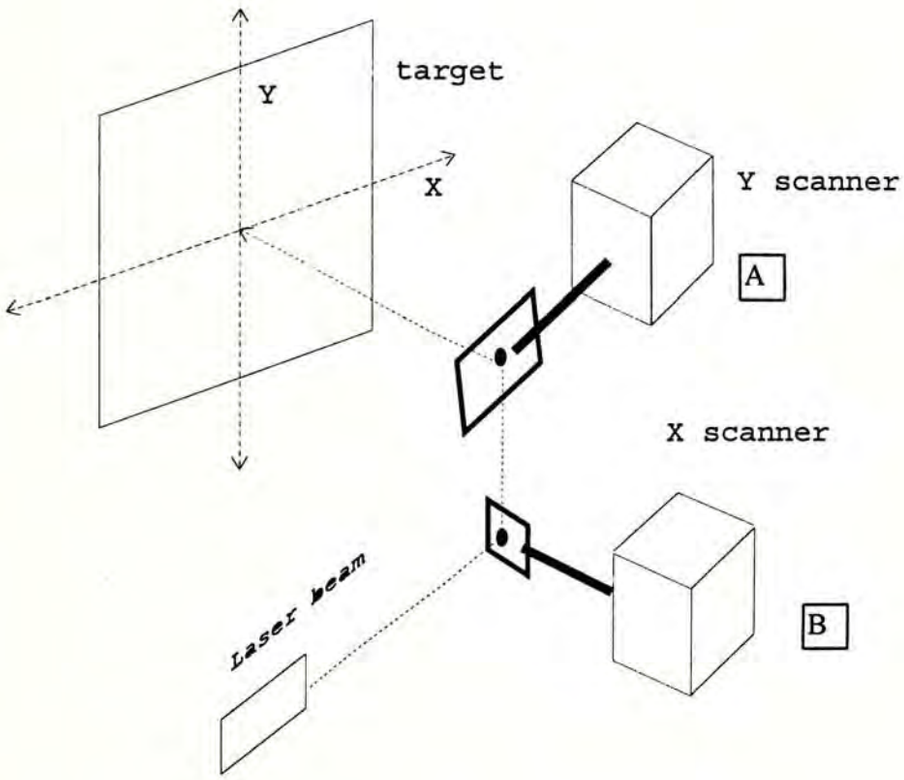


Figure 3.2-X-Y scanning head.

It is a common practice to design a negative feedback to obtain high precision and linearity in a control circuit. In order to achieve this, the scanning mirror is built-in a capacitive sensor as shown in figure 3.3. The stator and rotor are acting like the variable capacitor. Since the capacitance is proportional to the coupled area, and the

area is also proportional to the mirror position, we can therefore measure the coupling capacitance for the angular difference.

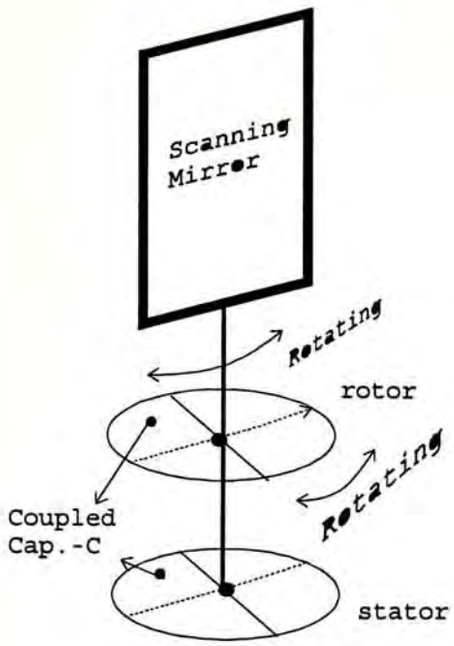


Figure 3.3-Coupling capacitance of scanning head.

3.4 Scanning Mirror Controller

Next, we are going to construct a scanning mirror controller which adopts the closed loop control because the closed loop control can provide a better linearity and stability. The research project is using the scanning head of General Scanning Inc. [35]. The specifications of the scanning head are listed on table 3.1.

Angular Sensitivity :	16μA/degree
Linearity :	±0.3% P-T-P
Excursion Range :	±40° P-T-P max.
Armature Inertia :	0.028g-cm^2
Torque Constant :	0.006Nm/A
Coil Resistance :	3.8Ω
Coil Inductance :	2.2mH
Supply Voltage :	±12Vdc

Table 3.1-Specification of scanning head.

The scanning head provides the feedback by the capacitive coupling which is then converted to a current output by a built-in transducer. However, the electronic comparators usually accept voltage rather than current reference. Therefore, we should convert the current to voltage by a simple circuit as shown in figure 3.4.

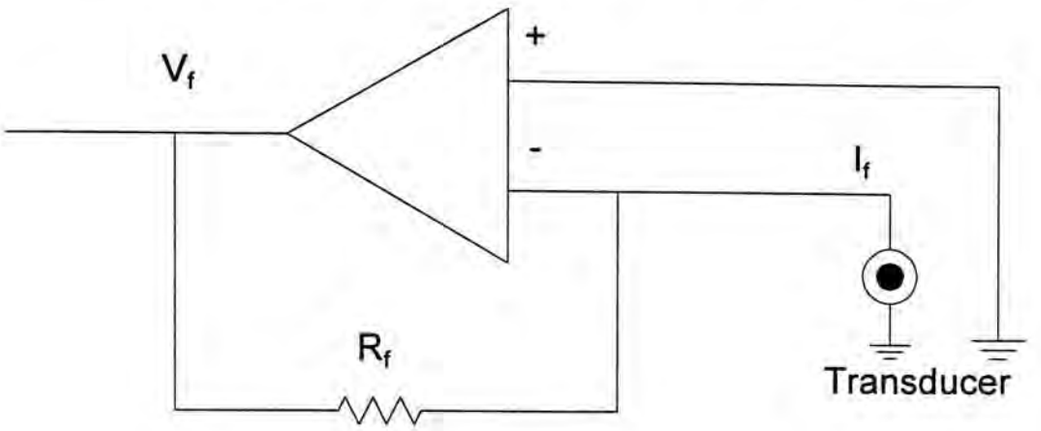


Figure 3.4-Current to voltage converter.

The voltage output can be calculated as :-

$$V_f = -I_f * R_f$$

Equation 3.2

where R_f = feedback transresistance.

The controller consists of 4 parts, i.e. (1) bus interface, (2) D/A converter, (3) galvanometer driver, and (4) feedback converter as shown in figure 3.5. The RHS of the dotted line indicates the devices outside the circuit, and the LHS indicates the circuit inserted in the computer expansion slot.

The operations of the circuit can be summarized in this way :- (1) The computer will issue a command to latch a data into the D/A converter. For our purpose, a 8-bit D/A converter seems to be enough which can be easily implemented on IBM XT bus. (2) The D/A converter converts the latched 8-bit data with the settings of scale and offset trimmers to a set-point voltage V_s. (3) The summing amplifier will subtract V_s with feedback signal V_f and produce driving voltage $V_d = V_s$-

V_f (4) The galvanometer driver will then drive the scanning head by converting V_d to the drive current I_d. (5) The scanning head is rotated with an angle θ by a magnetic force which is driven by I_d. (6) The sensor would produce a capacitance C_c due to the coupling in between rotor and stator. (7) The internal transducer of scanning head would produce a feedback current I_f due to C_c. (8) Finally, the feedback converter will convert the current feedback I_f into voltage feedback V_f as mentioned above.

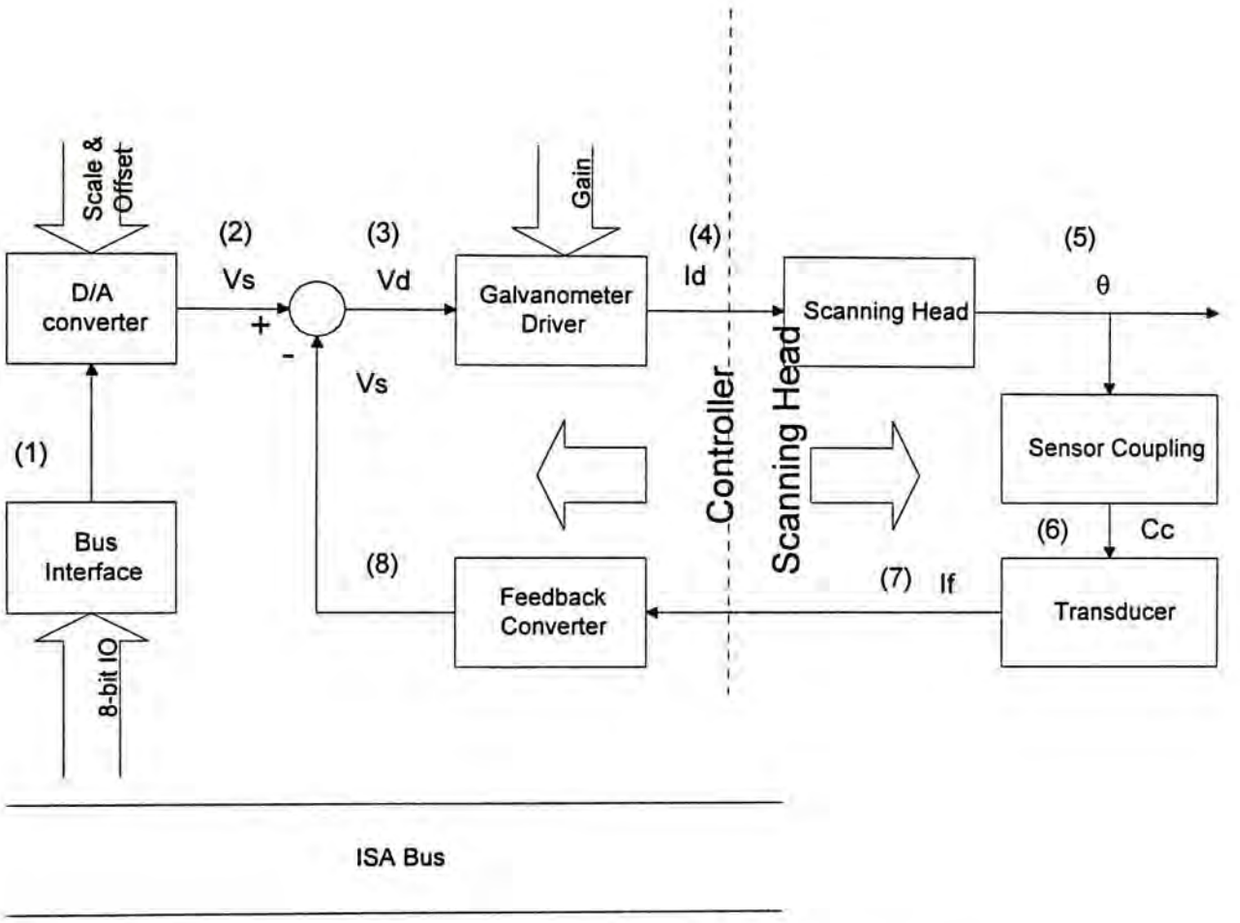


Figure 3.5-Block diagram of scanning mirror controller.

The circuit could be analyzed by an open loop control with a single pole as shown in figure 3.6. We are trying to adjust the feedback transresistance R_f to have an optimum performance. The settling time of the circuit is related to the overall gain and hence R_f. Larger the gain gives faster the settling time. However, if the gain is too large, the circuit will become unstable or oscillating.

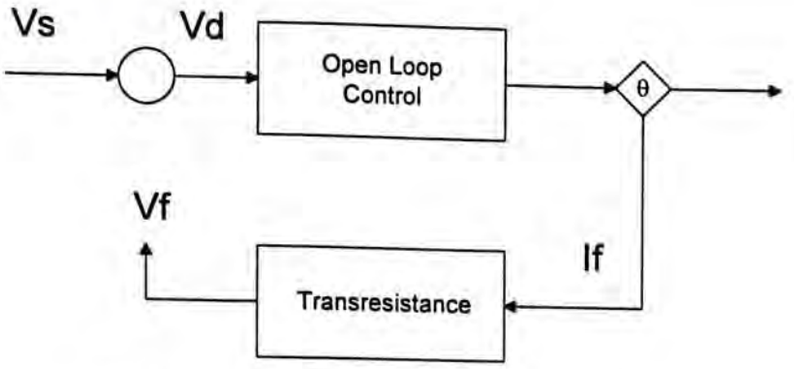


Figure 3.6-Open loop control of scanning head.

The open loop control could be estimated as first order function :-

$$H(s) = \frac{V_s}{V_f} = \frac{G}{T+s}$$

Equation 3.3

where T = propagation delay; G = overall gain.

The propagation delay is the system characteristic which could not be altered. However, the overall gain can be adjusted by tuning the resistance of R_f. Since the oscillation would happen with 180° negative feedback, the experiments were carried out to measure the frequency when 180° phase difference is obtained, which is about 1.2KHz. The experiment was then carrying on at the frequency (1.2khz) to select R_f such that the overall gain would not excess unity. And then, we determined the value of R_f by using a variable resistor, which is about 8.6KΩ, such that the control is highly stable with a maximum gain which is about 0.85 for our setting.

Two identical circuits are built for the x and y directions. And, we are using two I/O registers to control the corresponding directions in 256 steps. The resolution is dependent on the maximum swing and offset setting on D/A converters. For example, if +/- 10° of swing, the resolution should be divided by 256 steps. Therefore, the resolution would be 20° / 256 = 4.7'.

3.5 Point-to-Point Scanning

As for the application of range sensing, the scanning mirror is used to deflect and reflect the laser beam as shown in figure 3.7. The diode laser and detector are placed in close proximity. The mirror is used to deflect the laser beam onto the target surface, and the size is supposed to be large enough to reflect the excited spot also. As a result, the incident and reflecting path would form a triangulation. We can therefore use the property of triangulation to calculate the distance.

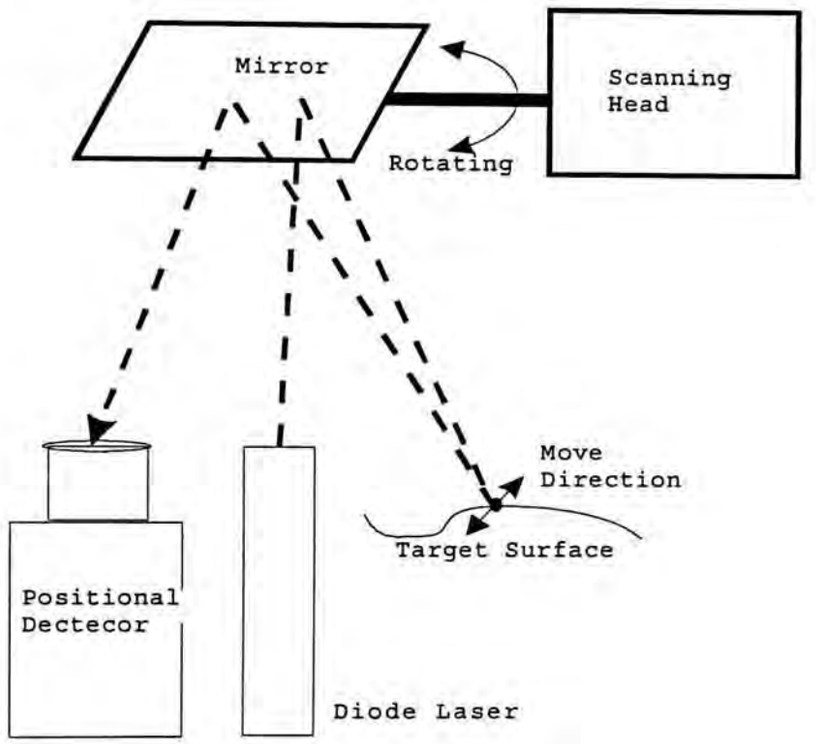


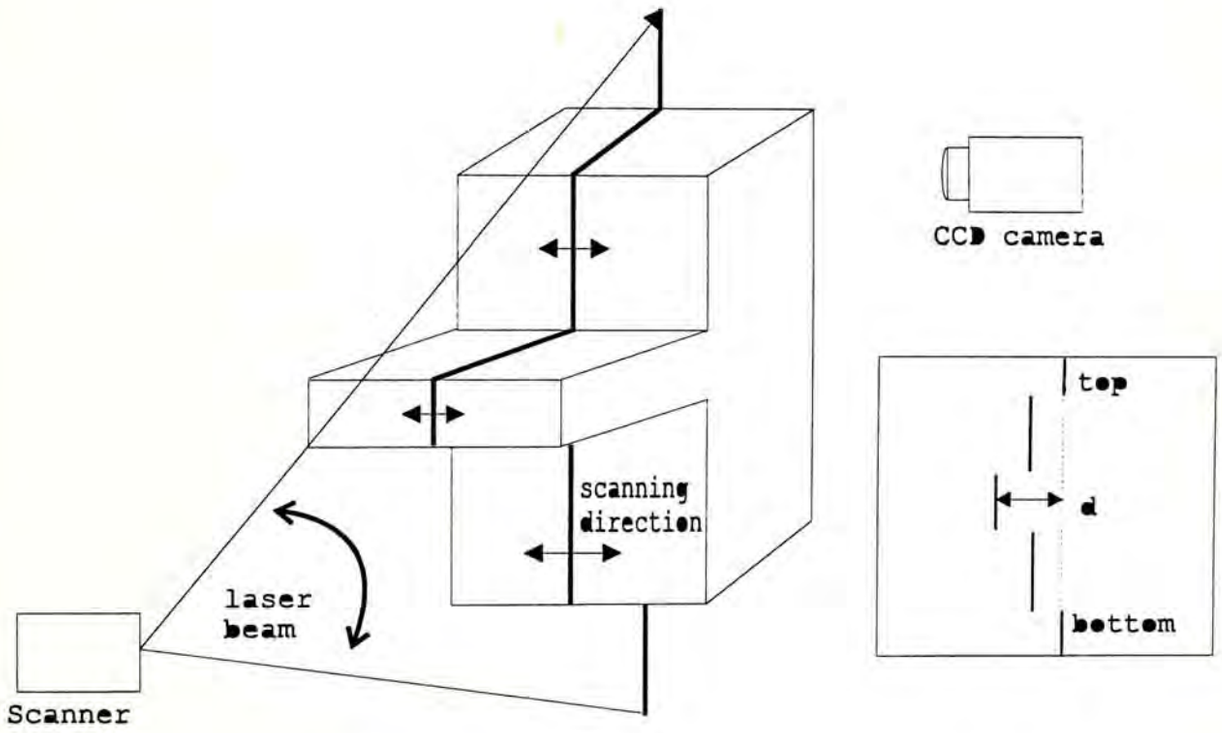
Figure 3.7-Point-to-Point scanning in scanning mirror.

Assuming that the scanner is scanning in the y-direction, another scanner could be placed orthogonally for the x-direction. Hence, by controlling these two scanners, the entire target surface could be scanned.

3.6 Line Scanning

Although we have chosen the high speed scanning mirror, the acquisition speed is still limited by the exposure period on the detector. Typically, the CCD

camera must be exposed for 20ms for enough potential build-up. This puts a restriction on the speed of range sensing. In order to achieve a higher speed of acquisition, it is so often to scan the entire object by a structural lighting of frame or line basis, rather than point-to-point measurement [5,11,12,13,]. For example, if a 128x128 points have to be sampled line scanning would speed-up the acquisition process by 128 times and a frame scanning only takes a single exposure which can save the scanning in 16384 times theoretically. However, structural scanning will involve analytical computations which slow-down the acquisition. As the design trade-off, line scanning is adopted.



(a) System diagram. *(b) Image.*

Figure 3.8-Line scanning using scanning mirror.

One of the scanner is then changed to oscillate for producing the line pattern. We then use the other scanner to select the line positions. The arrangement is shown on figure 3.8(a). The mirrors are now used as deflecting the laser beam only. And also, the detector must be chosen as 2-D array in order to capture a line on the screen.

Figure 3.8 (b) shows that a line of image is captured on the camera (Appendix A). We can apply a simple line searching algorithm to trace the line :-

(1) Find a point of (x_t,y_t) on the top horizontal line.

(2) Find a point of (x_b,y_b) on the bottom line.

(3) Calculate the slope and offset of line y=ax+c.

(4) For y =0, start x = y/a - c.

(5) Search left for a point larger than the threshold.

(6) Record d[y] = number of steps.

(7) Repeat (4), (5) and (6) for y=1 to height of image.

According to that line, we can then calculate the distance *d* with reference to the dotted line. Please note that *d* is strictly related to the height of the object (Appendix A). Thus, we could take these *d* values as an alternative to acquire the range values. Calculating *d* is much faster than the actual range values which can be converted by an off-line batch job.

3.7 Specifications and Measurements

We have implemented both experimental setup with point and line scanning. Figure 3.13 shows the scanning mirror controller which was built for the experiments. Figure 3.14 shows the scanning head from General Scanning [35]. Figure 3.15 and 3.16 show the experimental setup for both point scanning and line scanning respectively.

As a conclusion, the specification of scanning head is listed on table 3.2 :-

Maximum swing :	+/- 20 degree
D/A resolution :	256 steps
Settling time :	10ms (max.)
Supply voltage :	+/- 12V DC
Laser power :	0.5mW

Laser voltage :	-18V DC
Range :	10 to 30 cm (f=20mm)
CCD camera :	CCIR, 512 x 512
Point scanning :	20 min., 256 x 256
Line scanning :	4 min., 128 x 256

Table 3.2-Specification of scanning head.

Figures 3.10 (a)-(d) show the laser spot for the object at different heights. The sizes and locations of the spots are also tabulated in table 3.3. As the height of object increased, the diameter and the location of light spot also increased. A horizontal line drawn on the images is indicating the light spot must be formed on the same plane of light path. The intensities of the spot were calculated by averaging over a window of 16 pixels with the center of spot which increased as the height of object increased. The experiment showed that the image did not depend on the reflecting angle in x or y mirrors. As a result, we can scan the object by adjusting the angles in x,y mirrors. The location and the size of image should not be altered, and which should be changed by only the height of the object.

height (mm)	*diameter (pixels)*	*location (pixels)*	*intensity (0-255)*
0	64	76	20
45	213	141	68
90	247	217	75
135	271	316	87

Table 3.3-Diameter/location vs height of laser spot.

Figures 3.11 (a)-(d) shows also the scanned object for the point scanning. The height of object is represented by the gray level. While the range values were taken as the spot location, the intensities were also captured by average intensity of the light spot at the same time. As the problem of self-occlusion which will be discussed one the later chapters, some portions on the images are displayed in dark because no return signal could be found. The occlusion area would be saved as negative values in file for distinguishing.

The experiments were also done with line scanning. As the lighting is in a line basis, we projected the line to an object with known height and calculated the error by least square fitting in the type of $y = mx + c$. The results are tabulated in table 3.4.

Height (mm)	*m(X 10^{-3})*	*c (pixels)*	*Error (%)*
0	-9.12	455	2.36
10	-8.46	387	3.00
14	-7.56	354	2.02
24	-9.01	282	2.14
34	-10.03	203	1.85
39	-8.56	159	3.16
49	-7.70	71	2.66
54	-7.59	26	1.75

Table 3.4-Least Square Fit Error of line scanning.

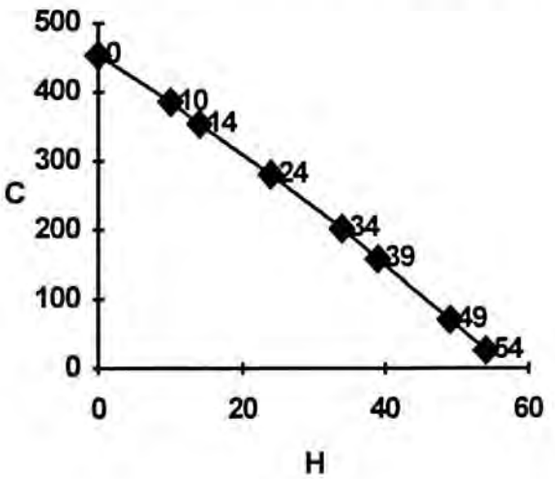


Figure 3.9-Plot object height vs. image offset

Figure 3.9 plots the offset C vs. object height which is approximately a linear relationship. Accordingly, we could calculate the distance by this graph. Figures 3.12 (a)-(d) are showing two sample objects obtained by line scanning. The occlusion area which saved as negative values in file are filled by bi-linear interpolation in an off-line batch job, whereas figure 3.11 was not filled by this process.

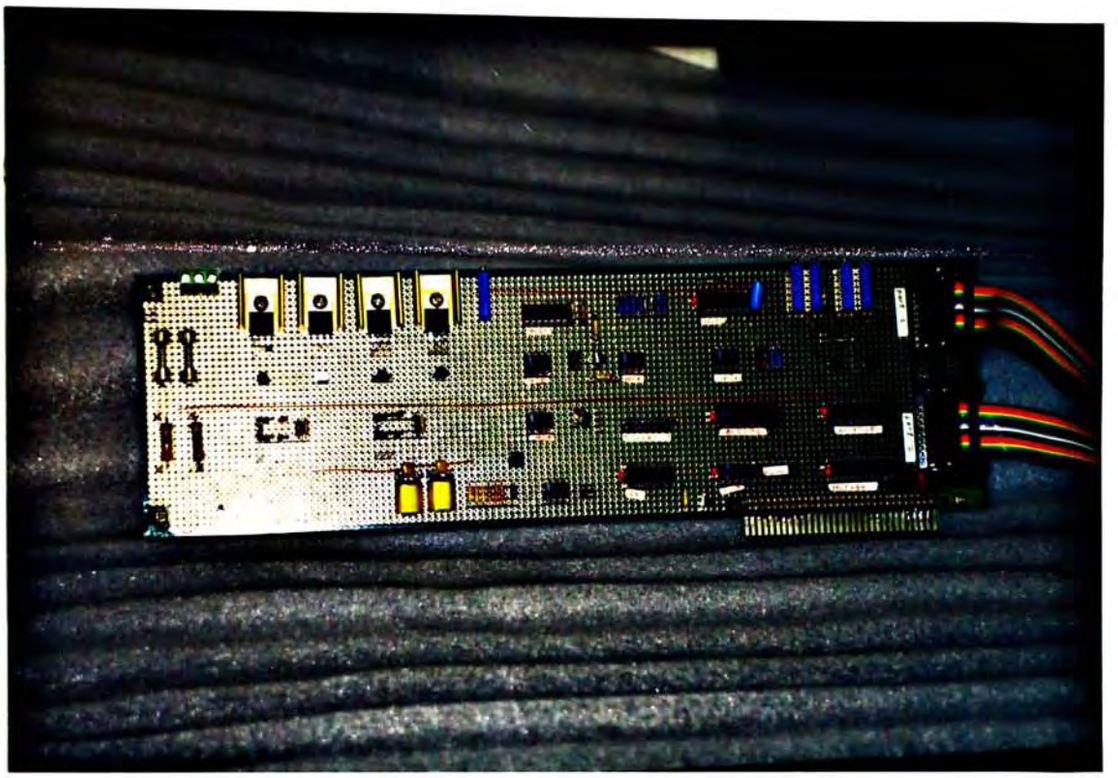


Figure 3.13 - Scanning mirror controller

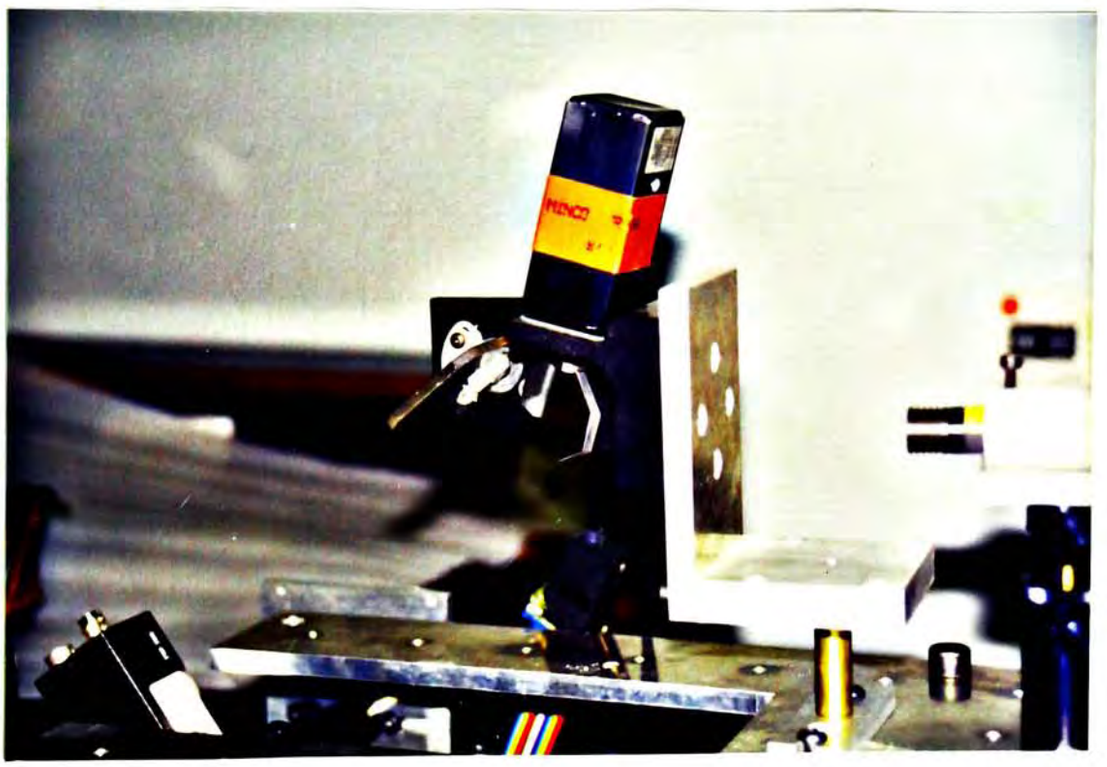


Figure 3.14 - Scanning Head

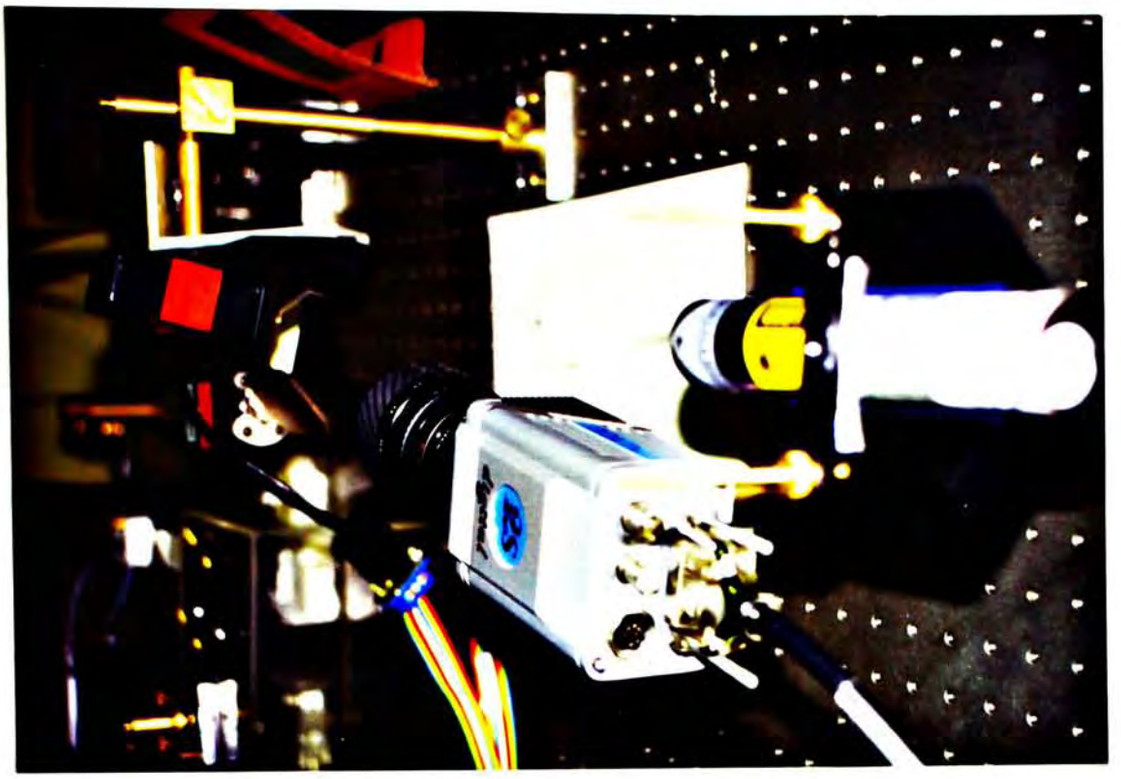
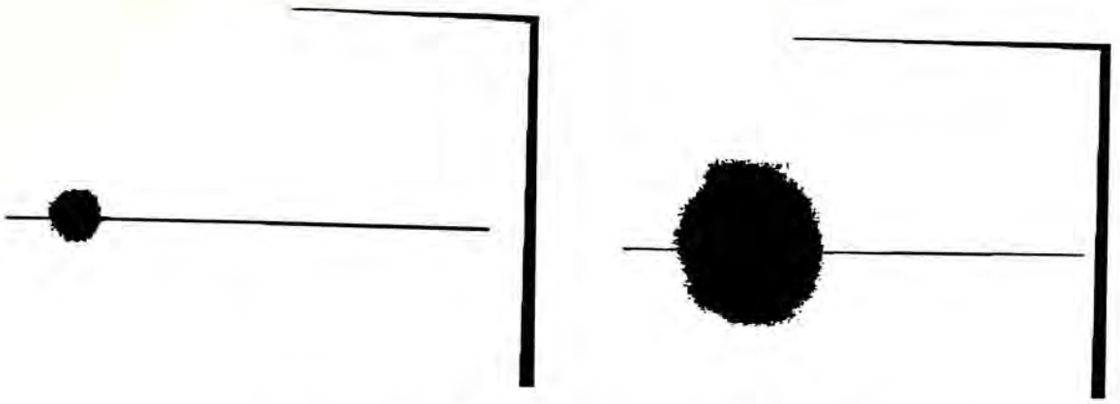


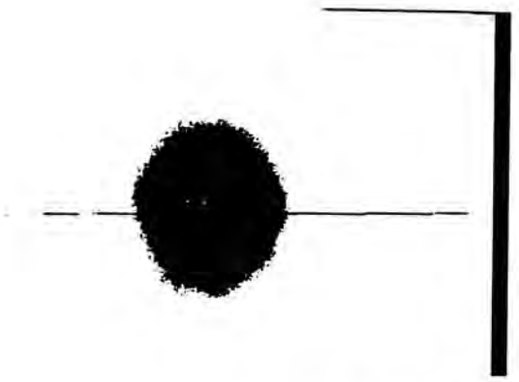
Figure 3.15 - Experimental setup - point scanning.

Figure 3.16 - Experimental setup - line scanning.

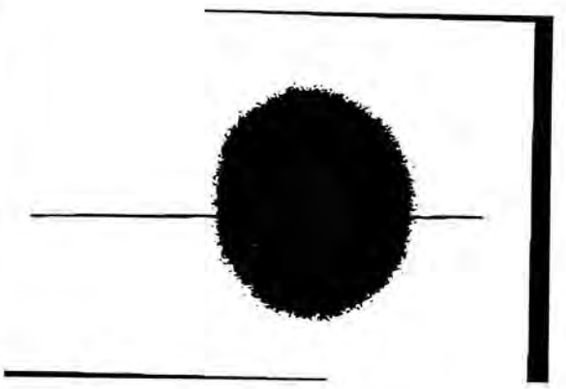


(a) height = 0mm

(b) height = 45mm

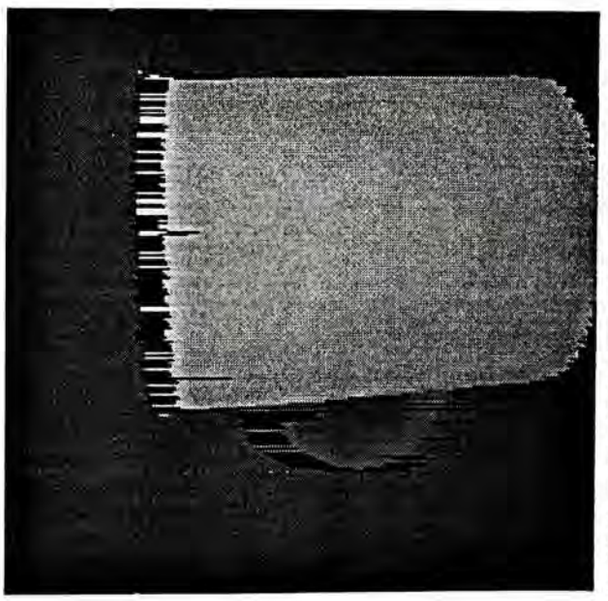


(c) height = 90mm

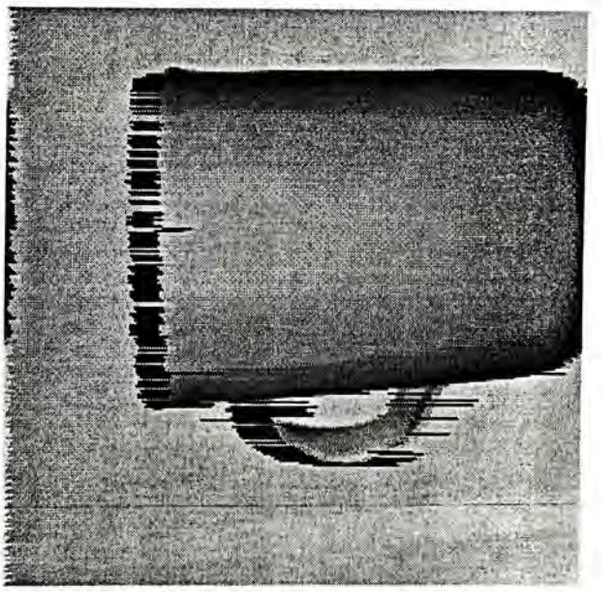


(d) height = 135mm

Figure 3.10 Laser spot of point scanning.

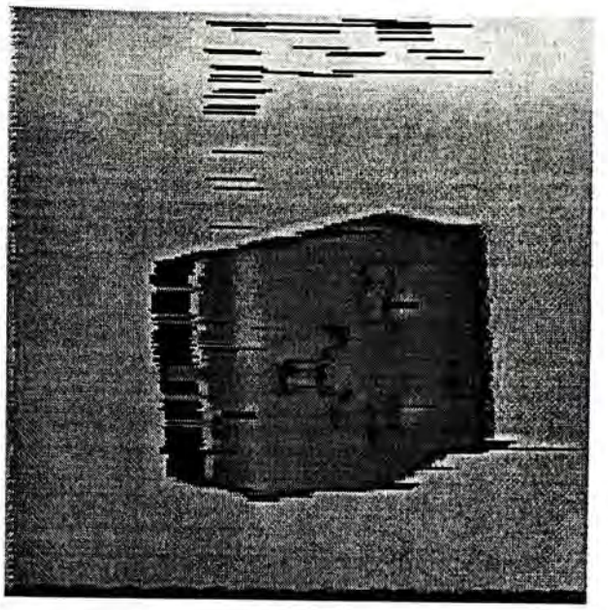


(a) Cup-Depth



(d) Cup-Intensity

(c) Plug - depth.



(d) Plug - intensity

Figure 3.11 Point scanning objects - Cup & Plug.

(a) Object view 1-depth.

(b) Object view 1-intensity.

(c) Object view 2-depth.

(d) Object view 2-intensity.

Figure 3.12 Line scanning-Irregular Object at Different Views.

Chapter 4

The Rangefinder with Reflectance Sensing

We have chosen the movable optics as our scanning mechanism for faster and more accurate range sensing. From the experiments, we could see that there were still many handicaps for our rangefinder, for example shadows and missing points. In this chapter, we will examine the factors which affect the performance of the range sensing.

4.1 Ambient Noises

Noises are the unexpected signals or disturbances affecting the measurement in range value. Our rangefinder is based on the active acquisition which is dependent on the presence of a known illumination pattern to determine the distance of object. Since the signal travels from the emitter to object and back to the detector, there may be ambient illumination to disturb the returned signal. This false signal may lead to mislocating the peak on detector. Since the distance is measured by the peak location, an error may be resulted.

First of all, the solution to overcome the ambient noise is to isolate the system from outside during acquisition. Secondly, we could use a light source with larger optical power in order to distinguish the ambient lighting. However, the power of light source must be very high to do so. ERIM [18] used their laser power of 100mW to overcome this problem. Such kind of high power laser is hazardous to human eyes. So

extra care must be taken during acquisition of the range data. Meanwhile, the laser of large power (e.g. >10mW) is only available for the gas type operating laser which may require high voltage bias (e.g. 1750Vdc), which is not suitable for the autonomous applications. Thirdly, signal processing techniques may be able to overcome noise under ambient illuminations and low power laser. We could consider the ambient illumination as random signals which could be filtered out by using special signalling. Finally, rather than using signal processing filter, an optical bandpass filter could be used early in the detection process. Since our system uses monochromatic source, an optical bandpass filter at the appropriate frequency may be used to pass most of the signal while rejecting most of the ambient illumination. As a conclusion, isolation from the ambient lighting seems to be the most acceptable method to overcome the problem of ambient illumination which does not require special signalling and large optical power. However, the acquisition must be operated in a dark room which is applicable for autonomous robots.

4.2 Occlusion/Shadow

Common problem for triangulation in range sensing is shadowing. There is some region from which no signal will be returned for the detector. As the active triangulation is based on the angle of incident and returned light path, and for the reason that the two light paths are different, some portions in the object cannot be seen by the detector or cannot be illuminated by the source. Figure 4.1 illustrates the conditions for self-occlusions and shadows. Self-occlusion is defined as the area which could be illuminated but is not invisible from the detector position. Shadow is defined as the area which could be visible to detector but not illuminated by the light source.

Since occlusion and shadow are due to the nature of triangulation, we cannot eliminate this kind of error by a single acquisition. Multiple acquisitions at different views may be a solution because the occluded/shadow area may be visible/illuminated at other viewing angle. We can then try to match surfaces into a single global object by merging different views. However, the concave features are still a critical problem for the triangulation. Interferometry [23] is a method of making range measurements based on the difference in phase or pattern between two light fields. Since the light path could be traveled perpendicular to object surface, it seems to be the best alternative solution to measure the depth of concave object.

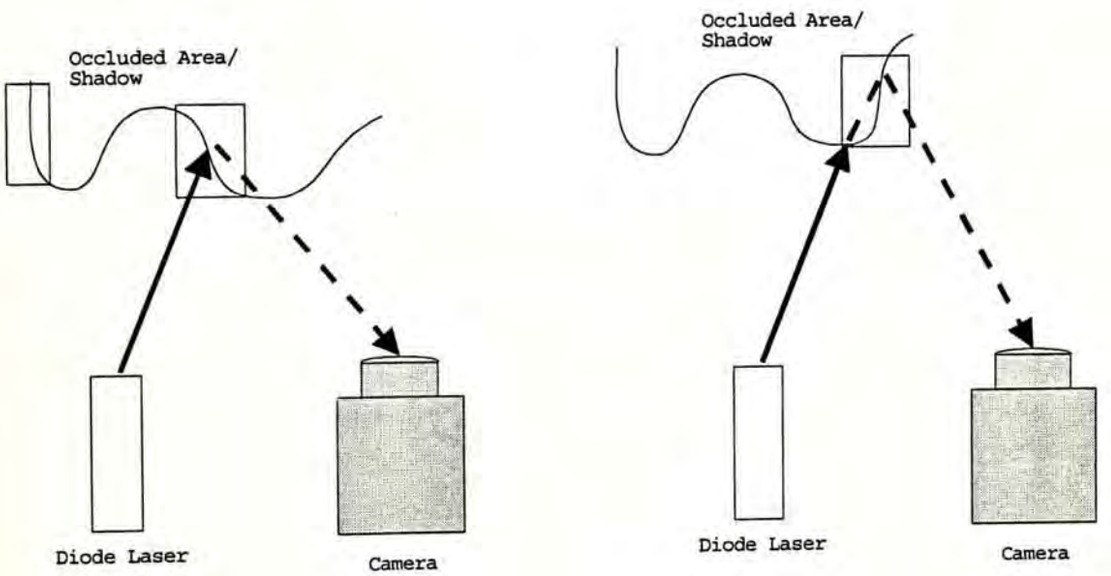


Figure 4.1, Left - Occluded Area; Right - Shadows Area.

4.3 Accuracy and Precision

In chapter 2, we have briefly described that the active triangulation must be measuring a distance value of limited range. We are going to discuss this factor for our specific application. First of all, the detector can be considered as consisting of finite elements of photodetectors. Each element is located on the rectangular grid with

fixed separation. Thus, the resolution of detector is constant within the spatial domain. Recall that the distance is defined in chapter 2 :-

$$R = \frac{h}{h'}(f+u),$$

or,

$$h' = \frac{h(f+u)}{R}$$

Equation 4.1

where h' = peak location;

f+u = distance from image plane to lens;

h = baseline separation.

If we differentiate equation 4.1 with respect to R,

$$\frac{dh'}{dR} = -\frac{h(f+u)}{R^2}$$

Equation 4.2

The derivative $\left|\frac{dh'}{dR}\right|$ is defined as the measurement sensitivity due to change in range value δR. Let δh' is the pixel size in detector. The measurable range shift is simply obtained by dividing δh' by $\left|\frac{dh'}{dR}\right|$.

$$\delta R = \frac{R^2 \delta h'}{h(f+u)}$$

Equation 4.3

This quantity δR is range resolution in determining the accuracy of a rangefinder system. Usually, the pixel size δh' is kept as a constant in detector. According to equation 4.3, the range resolution is coarse as the object moves farther which worsen quadratically as distance increases. In another word, the resolution in triangulation system is a variable to the object distance. We can maximize the baseline separation h of equation 4.3 in order to enhance the range resolution. However, the

area of occlusion/shadows will get larger with increasing h and the size of mirror puts a limit to this separation.

Another quantity of interest is the spatial resolution on the object. Since we are adjusting the angle of incident rather than translating the object. Spatial resolution for the positional adjustment system, e.g. x-y table, is simply equivalent to the stepping sizes of stepper motors in x and y direction, i.e. δx and δy respectively. However, the resolution of angular adjustment system would be a variable to the object distance.

Figure 4.2 shows the angular adjustment system, where R_x is the range value in x-direction and x' is the baseline separation in x-direction. A similar diagram could be constructed for y-direction. Thus, we take the analysis in x-direction for both cases.

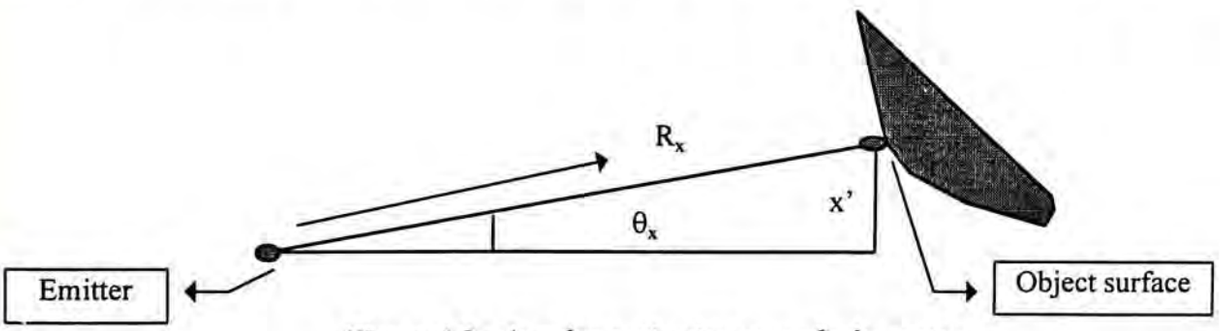


Figure 4.2 - Angular projection rangefinder system.

Generally, the angle of total swing in scanning would be small amount ($\pm 10^{o}$, approx. [35]). Thus, we could take an approximation to get R_x :-

$$x' \cong R_x \theta_x$$

Equation 4.4

We can therefore define the angular sensitivity due to x-direction :-

$$\frac{d\theta_x}{dx'} = \frac{1}{R_x}$$

Equation 4.5

Accordingly, we could obtain δx' by dividing the angular step $\delta\theta_x$ by $\left|\frac{d\theta_x}{dx'}\right|$:-

$$\delta x' = R_x \delta\theta_x$$

Equation 4.6

From equation 4.6, the spatial resolution becomes coarse when the distance to object increases. However, this leads to an advantage that the resolution is adjustable by varying the object distance. When comparing with x-y table, the spatial resolution must be fixed in size.

4.4 Optics

According to equation 4.6, the projective coordinates in angular adjustment system would be changed with the object distance. For line scanning, our rangefinder is emitting a sheet of light into target for a line projection and therefore the distances change within the line. Supposed that the light diverges from a point $(x_s,0,z_s)$. We could like to take an exact computation on the image point focused with y coordinate [22] :-

$$y_i = \frac{f \tan\theta_y (z_o - z_s)}{z_o - f}$$

Equation 4.7

where f = focal length;

θ_y = light projection angle in y-direction;

z_o = object distance of object to target;

z_s= distance of detector to lens.

Object distance z_o must be greater than focal length f; otherwise, the image cannot be formed on detector. Obviously, according to equation 4.4, the denominator determines the projective coordinate on the image plane. So, as the object moves toward the lens, the image will move vertically. This could been illustrated as figure 3.8 which was neglected the projection factor. The actual image should be formed as shown in figure 4.3.

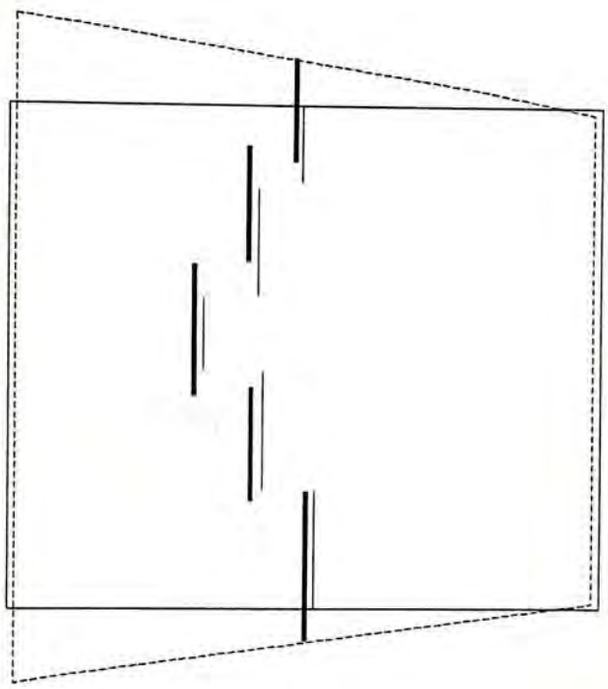


Figure 4.3-Projective Magnification .

We can see that the segments on the left have been magnified. The magnification factor relates to the object height. Therefore, the higher the target, the longer the segments would be. These also result in the overlapping of two segments.

The choice of lens would determine the resolution and the range of distance measurement. The range of distance measurement relates to the depth of field of lens. The image is sharp only in a valid range of distances, i.e. the depth of field.

Sometimes the term depth of field is used for the combination of the distance from the closest to the farthest point that is sharp. The front depth and rear depth can be equated as [36] :-

$$D_{front} = \frac{cN_e}{M^2(1+N\frac{c}{fM})}$$

$$D_{rear} = \frac{cN_e}{M^2(1-N\frac{c}{fM})}$$

Equation 4.8

where f = focal length; N = F-number or F-stop;

M = Magnification factor = $\frac{z_s - f}{f}$;

N_e = Effective F-number (corrected by bellow) = $\frac{N}{1-M}$;

c = diameter of largest acceptable defocusing pillbox, i.e. the resolution on detector.

By all means, small value of F-number can provide a large depth of view and hence good range of distance measurement. However, since the required resolution on detector is directly related to the defocusing pillbox. High density detector requires a large diameter on the lens and hence large F-number. The pillbox size also affects the accuracy of locating the peak because the searching cover a large area. Sometimes the pillbox is used for determining the distance as described in chapter 2. For these measurements, a large F-number of the lens is necessary in order to acquire an accurate range value.

4.5 The Range/Reflectance Crosstalk

Herbert [18] showed that the range measurement could be interfered by the surface reflectance. The experiments taken by him would be used for our illustrations. Figure 4.4 shows the experiment for evaluating the range/intensity crosstalk. The experiment was taken in which a target with low reflectance was observed against a background of higher reflectance. 100 samples for line scanning of intensity and range were taken and plotted out as shown in figure 4.5.

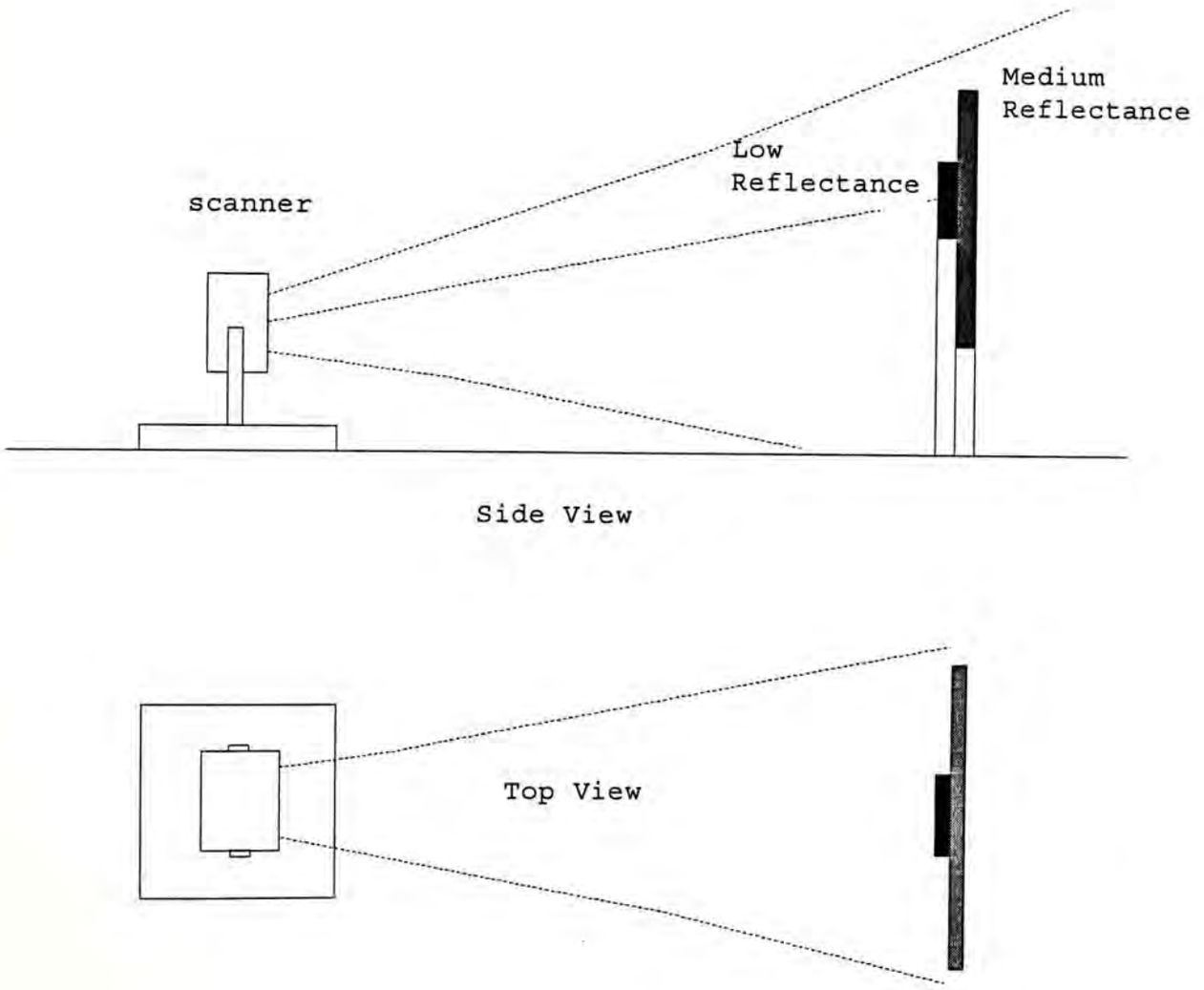


Figure 4.4-An experiment done by Herbert.

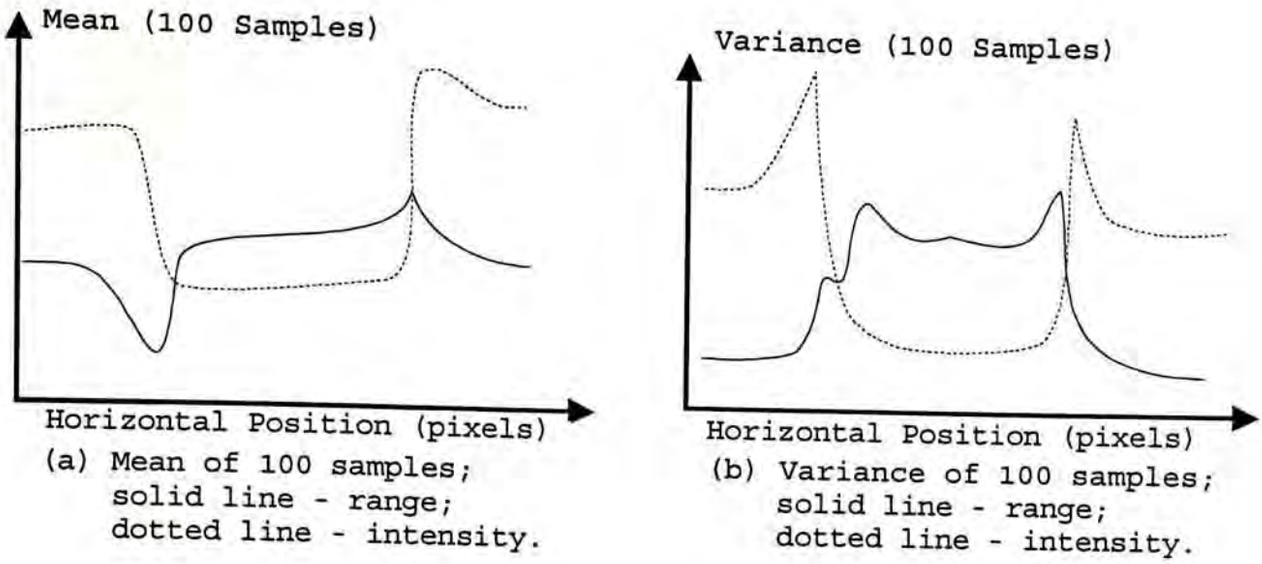


Figure 4.5-Results of Herbert's experiments.

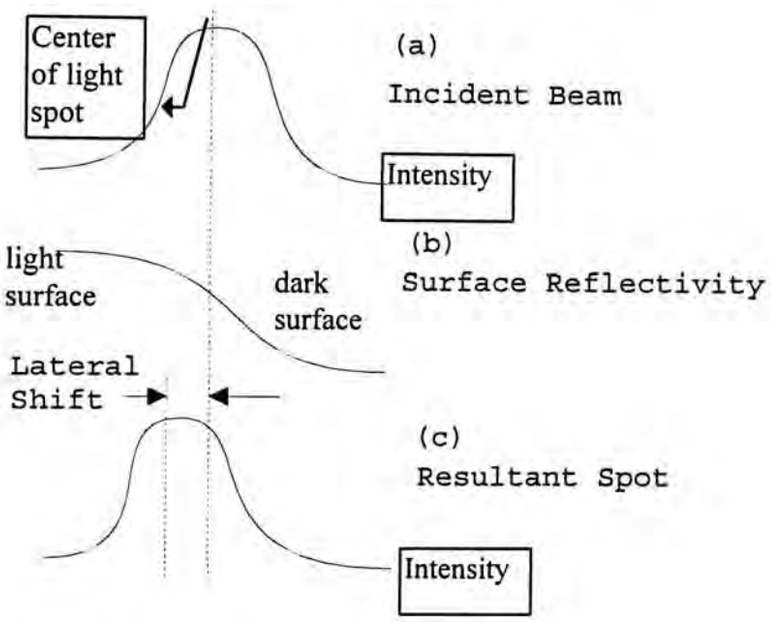


Figure 4.6-Range/Intensity Crosstalk

The experiment showed that mean of the range values remains roughly constant except for a sharp discontinuity at each edge. Also, the variance is high in the low reflectance region. In order to explain the effect of reflectivity variations on the accuracy of range finder, consider figure 4.5. (a) shows that the incident beam profile is of the Gaussian type. Meanwhile, (b) shows the reflectivity on the target surface, which surface is light on the left and dark on the right. Since returned signal is directly

reflected from the target surface, the signal will then be modulated as those depicted in (c). Therefore, we can see that the final result is equivalent to a lateral shift on the detector. Hence, the measured distance is also altered with this shift.

4.6 Summary

We have examined different factors which affect the performance of the range sensing. First of all, ambient noise is the common problem for all the rangefinder system. If the object could be measured in an isolated environment, this problem could be avoided. The solution suits our target and is the most simplest one. Secondly, the design trade-off exists in between occlusion/shadow and range accuracy. The range accuracy is controlled by the baseline separation between emitter and detector which gives higher accuracy as it increases. However, the area of occlusion/shadow is worsened as the separation increases. Thirdly, an advantage of the scanning mirror system is adjustable of spatial resolution by varying the object distance. We can have a coarse resolution if the size of object is large while a fine resolution for small piece of object. Finally, the scanning mirror system results in the projective magnification in the object spaces which must be taken into account for the measurement.

Chapter 5

Computer Generation of Range Map

In previous chapters, we have discussed the development of our rangefinder. Starting this chapter, we continue the discussion with data processing. As described in previous chapters, the captured images from the rangefinder are the partial views of the objects, which contain only the range values from the observer to the target. Some range values behind the surface, caused by either mutual occlusion or shadowing, cannot be obtained in the images. However it is essential to obtain global object model which have to be resolved through registration of multiple range images at different views [47].

Supposed that the object itself is a rigid body which does not change for different viewing coordinate. We could capture the multiple views of object which contain the common overlapping portions. Hence, the individual view could be resolved into the complete model of a physical object by recognizing the overlapping portion. This leads to the problem of feature extraction and matching. Supposed that the object is a rigid body, the features coordinates are simply transformed by a homogenous matrix as we change the viewing angle and position. Since features on the overlapping region remained the same, we can therefore merge them and create a relational graph. From the relational graph, the transformation matrix can be estimated from solving the matching function. Once the corresponding homogenous transform matrix can be found for each view, the global model can then be obtained by merging

image of different views as taking the inverse transformation. The goal of our research is to identify this transform matrix for matching/merging.

Instead of using realistic data to test the merging process, we propose to generate range data for this purpose because the acquisition process would inevitably be associated with noises. In order to put these uncertainties under control, we would like to use the computer simulated results instead of the realistic one. Additional advantages are (1) we could avoid the degradation during the noise removal processes. (2) the object can be constructed in any shape which may not be easily obtained in the laboratory. (3) the range/reflectance crosstalk problem can be avoided. (4) the viewer can be located at any position, or inside the body of object. (5) the acquisition is quick and accurate which can speed-up our experimental cycle. (6) different topologies of range sensing can be simulated in one application. Many computer-aided design tools, e.g. Iris Inventor[1] could satisfy our requirement. We have chosen SGI[2] workstation as the development platform since it is designed with the most advanced hardware to cooperate the manipulation in 3D graphics. SGI has built-in hardware with the 3D graphics engines, vector processor, z-buffering, lighting modeler and 24-bit image color planes which can much reduce our development efforts. Before going into actual implementation, let us first review the theory for computer graphics which may help us to understand the works easily.

---

[1] Iris Inventor is a registered trademark for Silicon Graphics. Inc.
[2] SGI is a registered trademark for Silicon Graphics, Inc.

5.1 Homogenous Transformation

It is so often to express a point in homogenous coordinates. This means that we represent a three-dimensional coordinate position *(x,y,z)* as quadruple $[x_h\ y_h\ z_h\ w]$ which is expressed in equation 5.1. The parameter *w* is assigned with a non-zero value in accordance with the class of transformations to be represented. Usually, for many transformations, we can set *w=1*.

$$x_h = xw$$
$$y_h = yw$$
$$z_h = zw$$

Equation 5.1

Supposed that we have a point (x,y,z) in one coordinate system which can be transformed into (x',y',z') in another coordinate system by a linear relation as equation 5.2.

$$x' = t_{11}x + t_{21}y + t_{31}z + t_{41}$$
$$y' = t_{12}x + t_{22}y + t_{32}z + t_{42}$$
$$z' = t_{13}x + t_{23}y + t_{33}z + t_{43}$$

Equation 5.2

From equation 5.2, we can express the transformation in matrix form :-

$$\begin{bmatrix} x' & y' & z' & 1 \end{bmatrix} = \begin{bmatrix} x & y & z & 1 \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} & t_{13} & 0 \\ t_{21} & t_{22} & t_{23} & 0 \\ t_{31} & t_{32} & t_{33} & 0 \\ t_{41} & t_{42} & t_{43} & 1 \end{bmatrix}$$

$$\begin{bmatrix} x' & y' & z' & 1 \end{bmatrix} = \begin{bmatrix} x & y & z & 1 \end{bmatrix} [T]$$

Equation 5.3

where [T] = homogenous transformation matrix.

We can simply multiply (x',y',z') by inverse of [T] to get the original coordinate (x,y,z) as equation 5.4, provided that $\|[T]\|$ is not zero.

$$[x \quad y \quad z] = [x' \quad y' \quad z'] \; [T]^{-1}$$

Equation 5.4

We can also define the different operations of transformation [46] :-

(1) Translation (T_x,T_y,T_z) =

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ T_x & T_y & T_z & 1 \end{bmatrix}$$

Equation 5.5

(2) Scale (S_x,S_y,S_z) =

$$\begin{bmatrix} S_x & 0 & 0 & 0 \\ 0 & S_y & 0 & 0 \\ 0 & 0 & S_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 5.6

(3) Rotation

$$\text{About x-axis, } Rot_x(\theta) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos\theta & \sin\theta & 0 \\ 0 & -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 5.7

$$\text{About y-axis, } Rot_y(\theta) = \begin{bmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 5.8

$$\text{About z-axis, } Rot_z(\theta) = \begin{bmatrix} \cos\theta & \sin\theta & 0 & 0 \\ -\sin\theta & \cos\theta & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Equation 5.9

5.2 From Global to Viewport Coordinate

The pixels on the target are defined as the global coordinates. When the object is scanned by the viewer (or the rangefinder), the global coordinates are then transformed into the viewer's viewport. The viewer's viewport is a limited size of rectangular box which is confined by the clipping planes. The clipping planes limit the global coordinates for mapping into the viewport such that the points mapped outside the viewport cannot be seen. Viewport is defined by 6 clipping planes, i.e. the far/near planes, the top/bottom planes and the left/right planes. Obviously, the far/near planes define the shortest and longest distances with reference to the viewer. The top/bottom and left/right planes corresponds to the field of view for the viewer.

We can see from figure 5.1 (a) the object is displayed as the global coordinate systems. The white lines show the viewer's coordinate and the direction. The viewport is indicated as the red and yellow rectangular box with the same direction of viewing. The yellow lines are used to distinguish the near plane, whereas red line are the far plane. Figure 5.1 (b) shows the projected image of the object. The size of the image is adjusted as we change the left, right, top, bottom clipping plane. Figure 5.1 (c) and (d) show the resultant images when we adjust the top and left plane respectively.

The transformation from the global system to the viewport involves two steps of homogeneous transformation. At first, the global coordinates are mapped into the viewer's coordinate system where the viewer is defined as the origin. The viewer transformation matrix is defined as equation 5.10.

$$view(v_x, v_y, v_z, p_x, p_y, p_z, twist) = tranlate(-v_x, -v_y, -v_z) \cdot rot_y(\theta) \cdot rot_x(\varphi) \cdot rot_z(-twist)$$

Equation 5.10

where v_x, v_y, v_z = the viewer's coordinate;

p_x, p_y, p_z = the viewer's pointing coordinate;

twist = twist angle for viewport;

$$\sin\theta = \frac{p_x - v_x}{\sqrt{(p_x - v_x)^2 + (p_z - v_z)^2}};$$

$$\sin\varphi = \frac{p_y - v_y}{\sqrt{(p_x - v_x)^2 + (p_y - v_y)^2 + (p_z - v_z)^2}};$$

Rot_x, Rot_y, Rot_z, Translate are defined in equation 5.5 - 5.9.

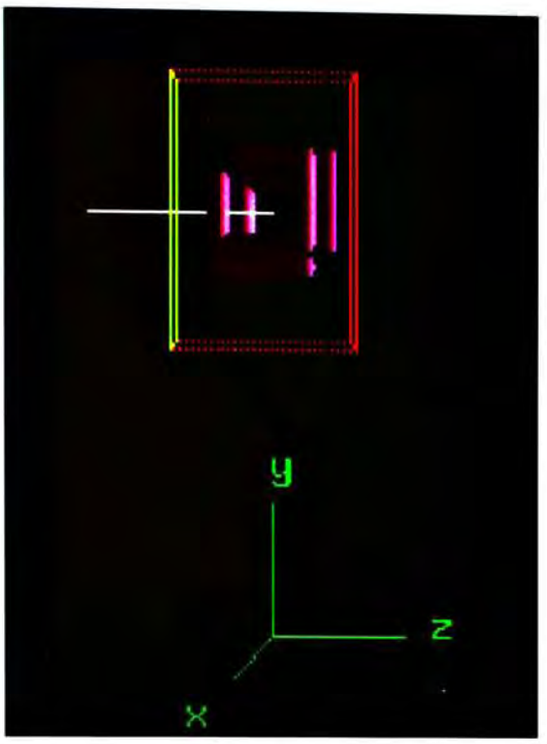


Figure 5.1(a)-Global Object, SGI Logo

Figure 5.1(b)-Viewport Image

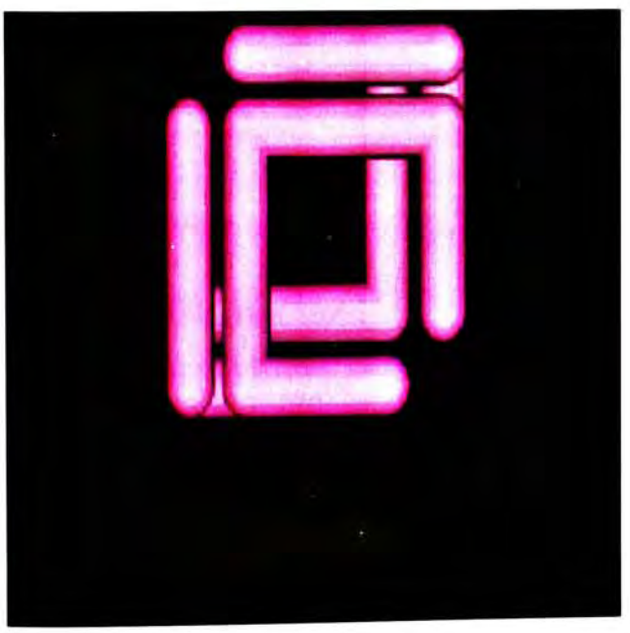


Figure 5.1(c)-Viewport Image of adjusting top plane.

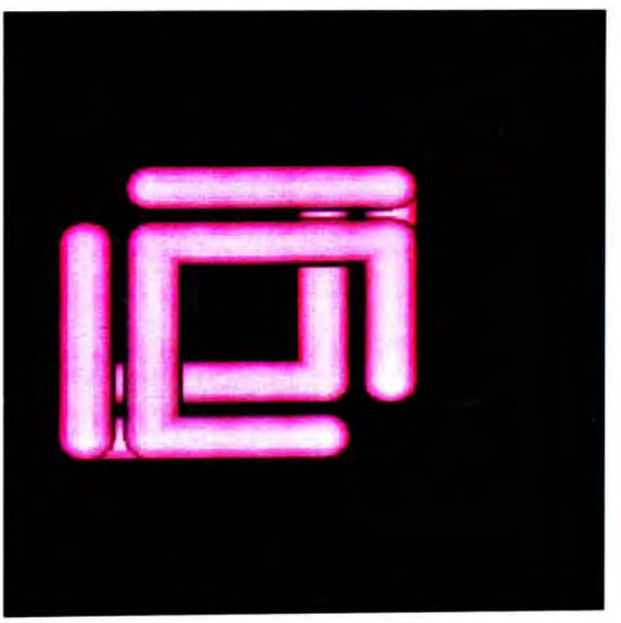


Figure 5.1(d)-Viewport Image of adjusting left Plane.

The next step of transformation is taking those viewer's coordinates into the viewport, which is referred as the viewport transformation. For different applications, the shape of viewport is different and hence different transformation matrix will be used. For example, the perspective transformation is similar to the actual scene of human eyes. The viewport shape is a frustum. However, as for our application, we would like the viewport as in the shape of rectangle. Thus, orthographic projection seems to be the most suitable transform. The homogenous matrix of orthographic projection is defined as equation 5.11.

$$\begin{bmatrix} \frac{2}{left-right} & 0 & 0 & 0 \\ 0 & \frac{2}{top-bottom} & 0 & 0 \\ 0 & 0 & \frac{2}{near-far} & 0 \\ \frac{right+left}{left-right} & \frac{top+bottom}{bottom-top} & \frac{far+near}{near-far} & 1 \end{bmatrix}$$

Equation 5.11

5.3 Z-buffering

While transforming the global object into the viewport, the z coordinates for the surface will also be mapped into somewhere on the viewport. Since the detector plane is a 2D array, the z coordinate may not be displayed on the viewer. However, it can be used for removing the hidden surface in or behind the front surface. The z-buffer is a 2D array that containing the z-coordinates on the scene. When a point is drawn, the global coordinate will be transformed in to the viewport coordinate. The obtained z coordinate would be used for comparing the value corresponding to that on z-buffer. When the value on z-buffer is greater than the z coordinate, the pixel is then allowed to overwrite the graphic display memory. This hidden surface removal

algorithm is called the z-buffering which is commonly used for the computer graphics. The size of z-buffer in Silicon Graphics Personal Iris [46] is 1280x1024x24-bit for such kind of operations.

As mentioned in above, the z coordinate is stored in the z-buffer while transforming the global coordinate into the graphics display. We might see that the z coordinate of the viewport is exactly the distance from viewer to the target surface. Hence, we can simply draw a scene for the object on the graphic display. The range map will be automatically generated in the z-buffer on the viewport. We can then read the content in the z-buffer and save it to the disks.

5.4 Generation of Range Map

By all means, z-buffering for hidden-surface removal can generate the range map for any scene. We can therefore read the 3D objects from file and draw the scenes on the graphic display. The viewer's coordinate can be adjusted in order to select the best view. Figure 5.2 shows the block diagram to set up for generating such range map.

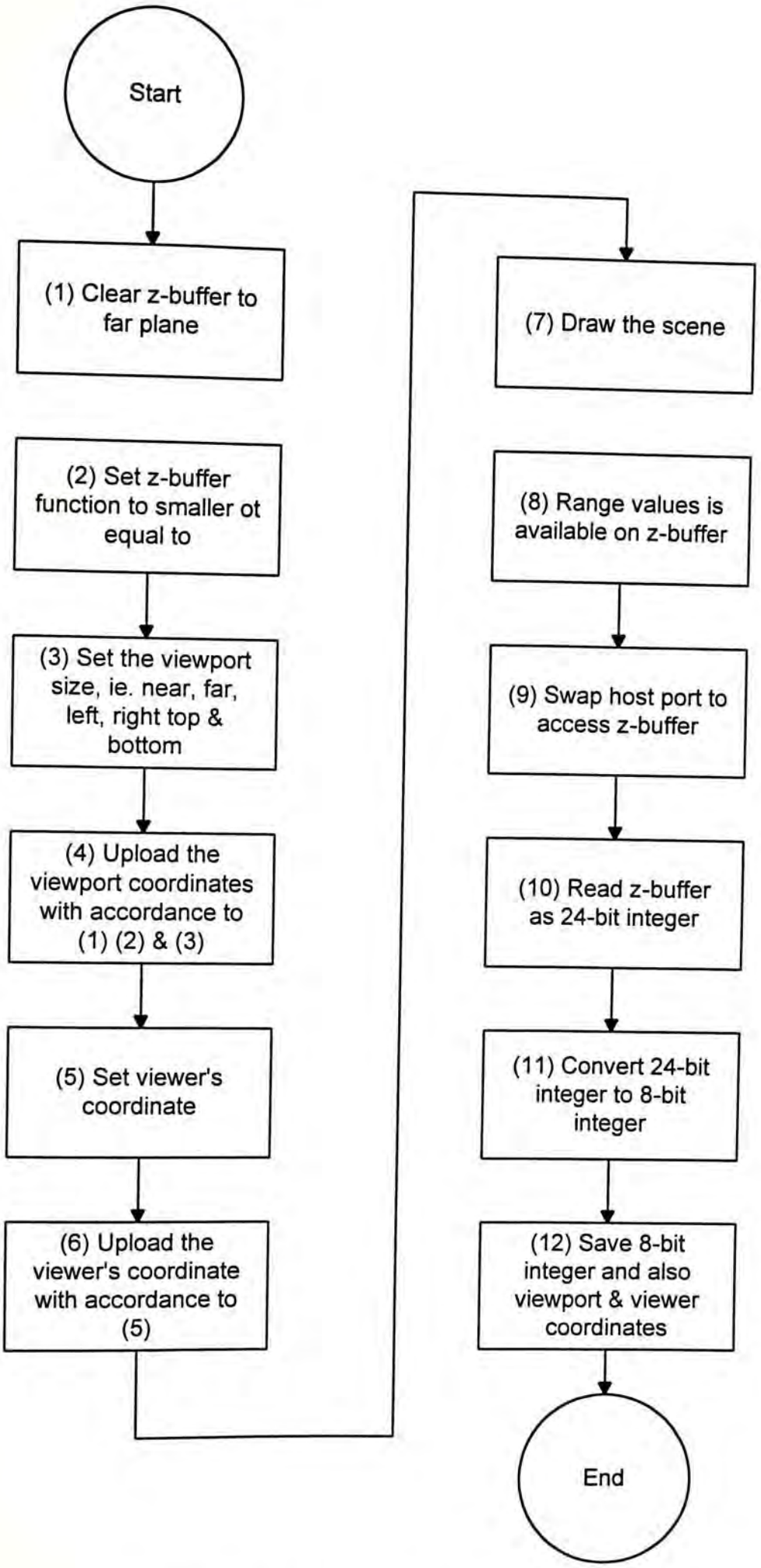


Figure 5.2 - Block diagram to generate range map

Blocks (1) and (2) are used to initialize z-buffer such that next pixels could be drawn on the viewport. Otherwise, the state is undetermined because the previous operations in z-buffer are unknown. Hence, nothing could be displayed. As mentioned, two transformations are used to project the 3D world space into 2D graphic space, which are the viewport and viewer transformation. Blocks (3) and (4) are used to define and upload the viewport transformation matrix to the vector processor of SGI. Likewisely, blocks (5) and (6) are used for viewer transformation as the purpose of (3) & (4). Two matrices are defined in different space for the SGI Graphics Library (GL). We could therefore change the individual matrix for different purpose. For example, we could change the viewer's coordinate and see the instantaneous result on graphic screen. After drawing the 3D scene (block (7)), the range image would be available on z-buffer (block (8)). Since the SGI would be accessible with image buffer as the normal operations, we should change the host port to access z-buffer as indicated on block (9). SGI would accept gray scale image with 8-bit maximum. Since the z-buffer is on the format of 24-bit integer, we must convert that into 8-bit format as shown on block (10) and (11). As shown on block (12), the range image, the transformation matrix and clipping plane coordinates would be finally saved for later analysis.

5.5 Experimental Results

Several 3D objects are used to simulate the process of range sensing. The intensity images are also saved for the references, which are including :-

(i) Volkswagen - figure 5.3 (a), (b) & (c);

(ii) Doughnut - figure 5.4 (a), (b) & (c);

(iii) X29 - figure 5.5 (a), (b) & (c);

(iv) Martini - figure 5.6 (a), (b) & (c);

(v) SGI logo - figure 5.7 (a), (b) & (c);

(vi) Chair - figure 5.8 (a), (b) & (c).

On the above figures, the global objects are shown on figure 5.3 - 5.8 (a). The figures 5.3 - 5.8 (b) show the range maps of one particular view. The figures 5.3 - 5.8 (c) show the corresponding intensity images.

The viewer's coordinates and clipping planes are recorded and tabulated on table 5.1 :-

Name	**v_x, v_y, v_z**	**p_x, p_y, p_z**	**twist**	**near/far**	**bottom/ top**	**left/ right**
Volkswa-gen	-0.16,0.24,0.38	0,0,0	0	0.5/0.89	-0.5/0.5	-0.5/0.5
Dough-nut	-0.32,0.32,0.51	0,0,0	0	0/0.19	-0.82/ 0.82	-0.74/ 0.74
X29	-0.32,0.24,0.38	0,0,0	0	0/0.19	-0.82/ 0.82	-0.74/ 0.74
Martini	-0.32,0.24,0.38	0,0,0	0	0/0.89	-0.5/0.69	-0.5/0.5
SGI logo	-0.16,0.24,0.38	0,0,0	0	0.05/ 0.89	-0.5/0.5	-0.5/0.5
Chair	-0.32,0.24,0.38	0,0,0	0	0/0.89	-0.71/0.7	-0.61/ 0.61

Table 5.1 Viewer's coordinates - v_x, v_y, v_z directions - p_x, p_y, p_z, *and Clipping planes.*

In a summary, we have reviewed how to use the computer graphics to generate the range map. We have also developed a program to obtain the range maps for any 3D object in file which are ideal for the further data processing experiments that requires noise-free images. The files are saved in standard Silicon Graphics Object Flip formats [46].

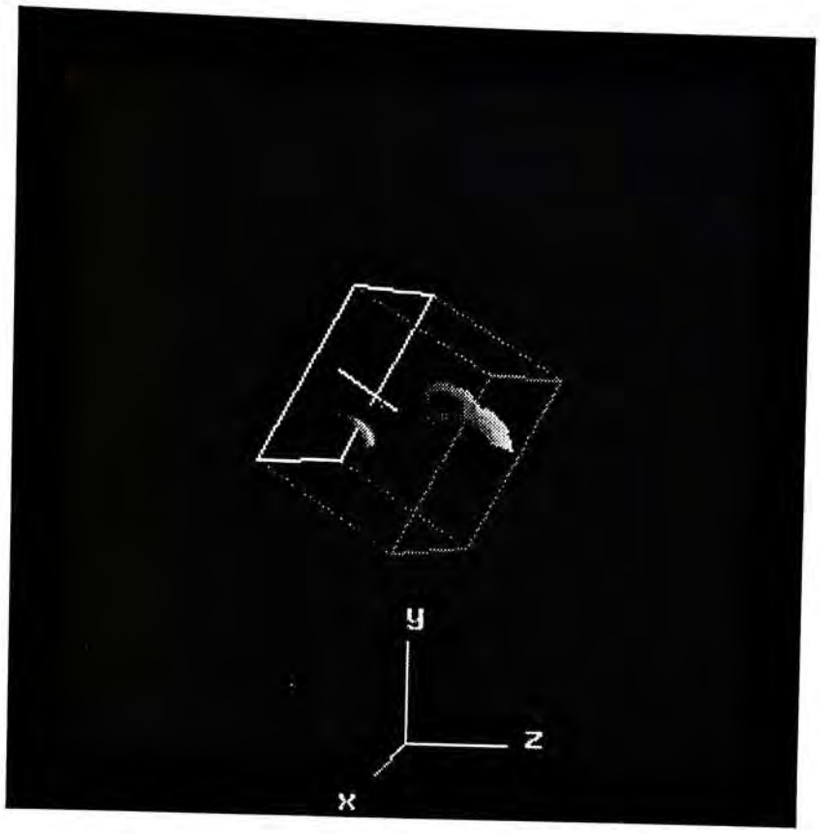


Figure 5.3 (a) Global Object - Volkswagen.

Figure 5.3 (b) Depth map - Volkswagen.

Figure 5.3 (c) Intensity image - Volkswagen.

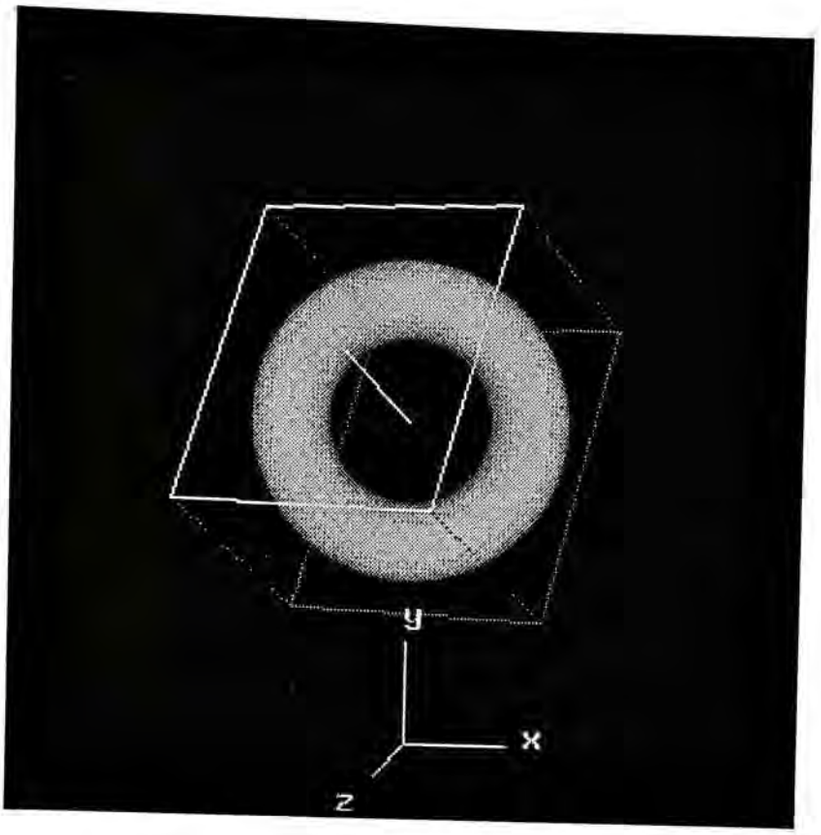


Figure 5.4 (a) - Global object - Doughnut.

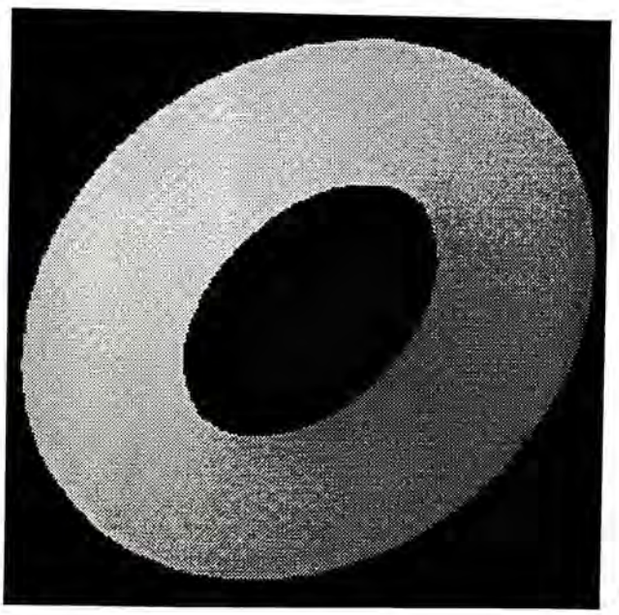


Figure 5.4 (b) Depth map - Doughnut.

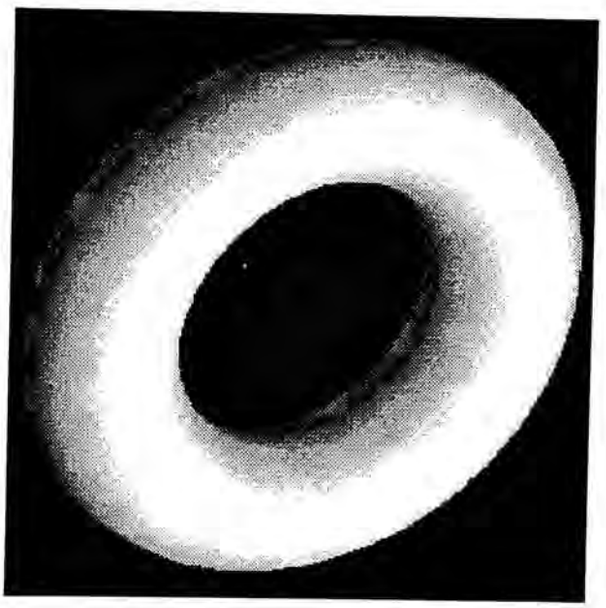


Figure 5.4 (c) - Intensity map - Doughnut.

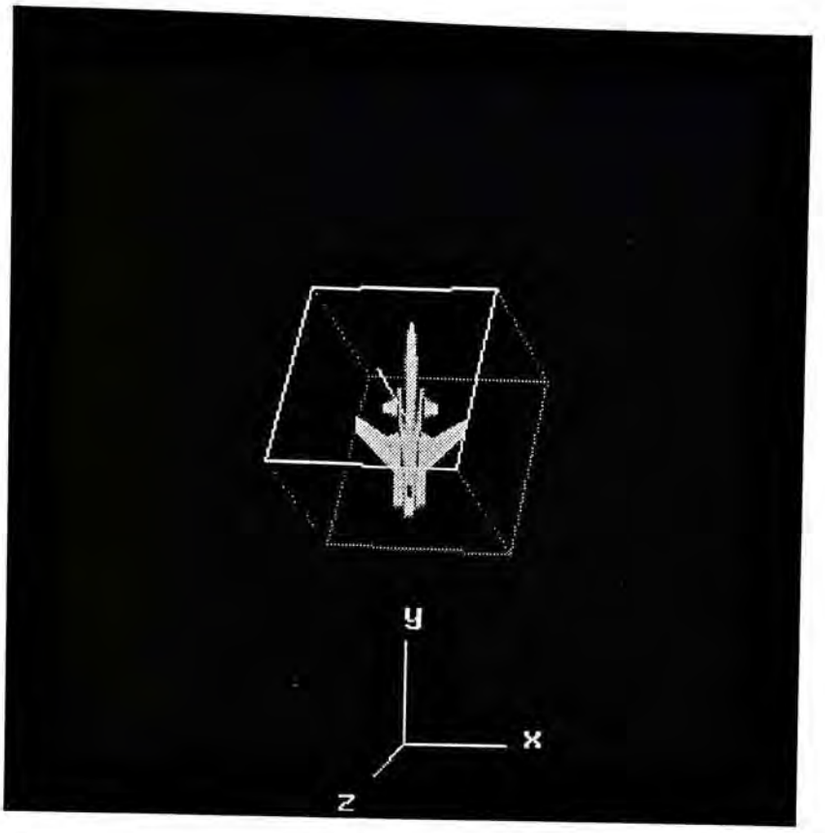


Figure 5.5 (a) Global object - X29.

Figure 5.5 (b) Depth map - X29.

Figure 5.5 (c) Intensity image - X29.

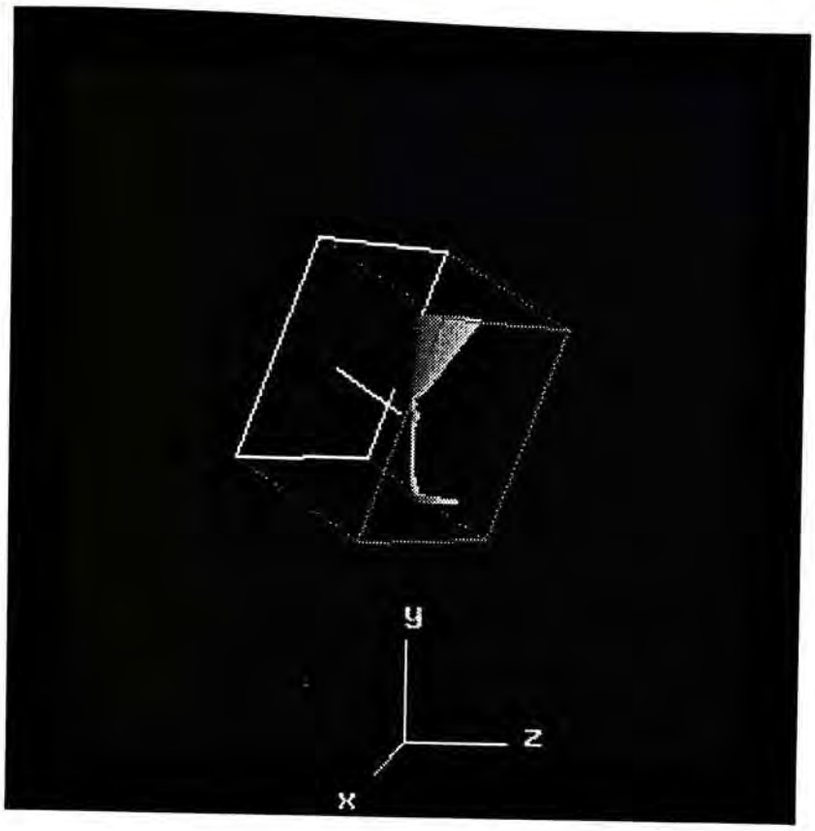


Figure 5.6 (a) Global object - Martini.

Figure 5.6 (b) Depth map - Martini.

Figure 5.7 (c) Intensity image - Martini.

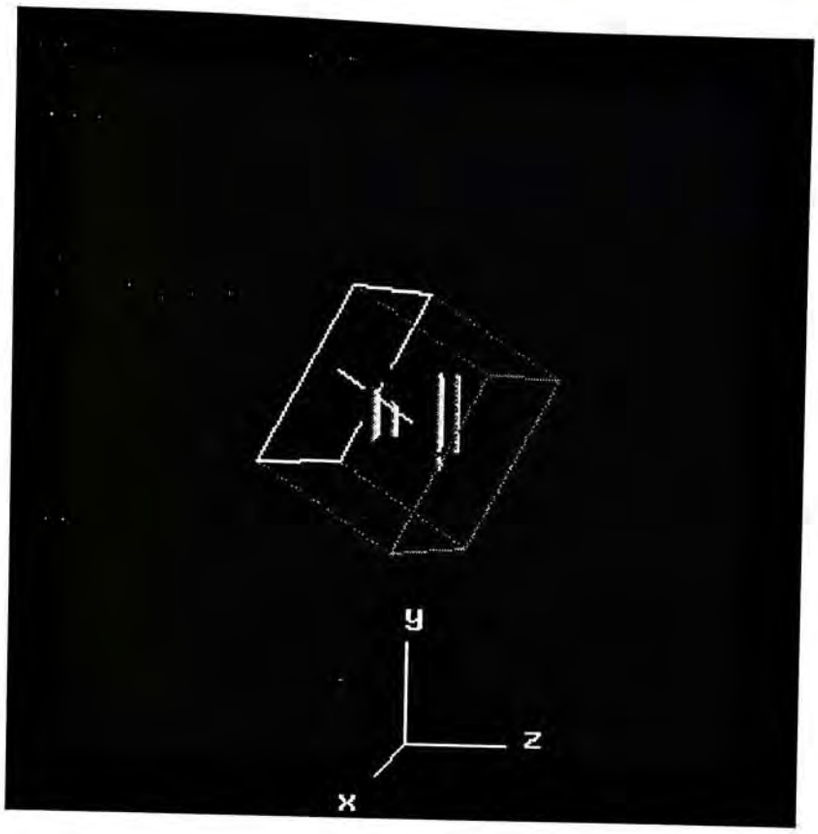


Figure 5.7 - (a) Global Object - SGI Logo.

Figure 5.7 (b) - Depth map - SGI Logo

Figure 5.7 (c) - Intensity imag - SGI logo.

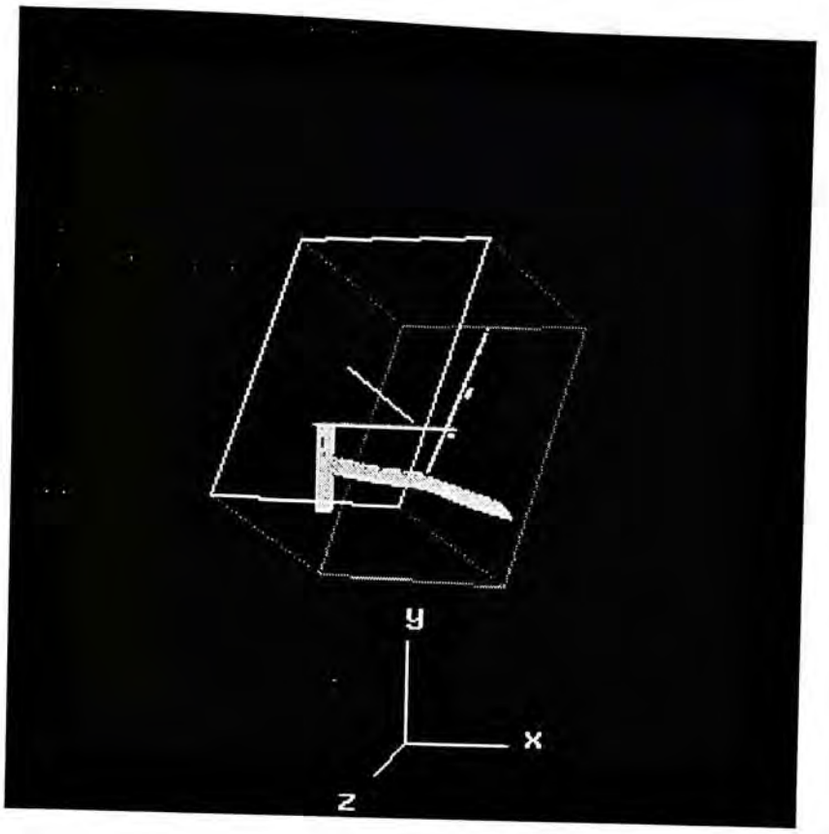


Figure 5.8 (a) - Global object - Chair.

Figure 5.8 (b) - Depth image - Chair.

Figure 5.8 (c) - Intensity image - Chair.

Chapter 6

Characterization of Range Map

It is our goal to reconstruct the complete 3D models from the partial characteristic views of range images. To achieve this goal, computing surface characteristics and selecting good primitives are important, because we propose to adopt the approach to match visible-invariant features in characteristic views. Thus, the immediate problem in recognizing the range image features is then addressed to handle arbitrary surface shapes at different viewing directions.

Since the characteristic views of range images are taken from different viewing directions, our chosen features must be independent in different characteristic views. Geometrical invariant transformation is one of the solutions for view independence. In past years, Besl and Jain [1] proposed that differential geometry could extract the visible-invariant features. In this chapter, we are going to explore those theories of differential geometry when applying to range imaging.

6.1 Mean and Gaussian Curvatures

Basically, differential geometry uses the curvature of surface to assign the feature types. Curvature of surface is the local property which does not change as changing the viewing directions. It does not like global properties, e.g. planar or quadratic surface equation, which would give a different parameters as changing the viewing direction or coordinate.

Generally, we would like to describe curvature as maximum and minimum principal curvatures, which are α_1 and α_2 respectively. Alternatively, curvature can be defined by Gaussian K and mean curvature H as :-

$$K = \alpha_1 \alpha_2$$

Equation 6.1

$$H = \frac{(\alpha_1 + \alpha_2)}{2}$$

Equation 6.2

where α_1 = maximum curvature; and α_2 = minimum curvature.

However, computation of derivatives is simpler than that of the maximum and minimum principal curvatures in discrete spaces. Thus, we define Gaussian and mean curvature as [1] :-

$$K = \frac{f_{uu} f_{vv} - f_{uv}^{\ 2}}{(1 + f_u^{\ 2} + f_v^{\ 2})^2}$$

Equation 6.3

$$H = \frac{f_{uu} + f_{vv} + f_{uu} f_v^{\ 2} + f_{vv} f_u^{\ 2} - 2 f_u f_v f_{uv}}{2(1 + f_u^{\ 2} + f_v^{\ 2})^{3/2}}$$

Equation 6.4

where f_u, f_v = the partial derivatives of first order;

f_{uu}, f_{vv}, f_{uv} = the partial derivatives of second order.

After obtaining the Gaussian and mean curvature, we can defined the feature types according to Besl [1], which are defined and tabulated in table 6.1.

	K<0	K=0	K>0
H<0	peak	ridge	saddle
H=0	(unknown)	flat	minimal
H>0	pit	valley	saddle

Table 6.1-Feature type definition.

6.2 Methods of Curvature Generation

There are two ways to calculate the Gaussian and mean curvature, viz. convolution and local surface patching. Convolution employs the gradient filters to obtain derivatives of f_u,f_v, f_{uu}, f_{vv}, and f_{uv} respectively. However, such kind of filters amplify noises in the range images. Therefore usually, a Gaussian filter is used to remove the noise components, but Gaussian filter reduces the resolution of the resultant features. Local surface patching is another method to generate the parameters of second order surface by least square fit.

6.2.1 Convolution

According to a Taylor's series expansion [38], we could approximate the result of function *f(x)* at x_1, and x_{-1} as :-

$$f(x_1) = f_1 = f_0 + hf_0' + \frac{h^2}{2} f_0'' + \frac{h^3}{6} f_0''' + \frac{h^4}{24} f_0^{iv} + \ldots$$

Equation 6.5

$$f(x_{-1}) = f_{-1} = f_0 - hf_0' + \frac{h^2}{2} f_0'' - \frac{h^3}{6} f_0''' + \frac{h^4}{24} f_0^{iv} + \ldots$$

Equation 6.6

where h is the quanitization space. By subtracting Equation 6.5 & 6.6, we could have

$$f_0' = \frac{f_1 - f_{-1}}{2h} + O(h^2)$$

Equation 6.7

$$f_0'' = \frac{f_1 - 2f_0 + f_{-1}}{h^2} + O(h^2)$$

Equation 6.8

where O(h) is the error function. Usually, the smaller the value of h, the more accurate the approximation is.

Similarly, we could take higher order approximation of Taylor's series expansion for $f(x_2)$ and $f(x_{-2})$, then f_0' and f_0'' would be

$$f_0' = \frac{-f_2 + 8f_1 - 8f_{-1} + f_{-2}}{12h} + O(h^4)$$

Equation 6.9

$$f_0'' = \frac{-f_2 + 16f_1 - 30f_0 + 16f_{-1} - f_{-2}}{12h} + O(h^4)$$

Equation 6.10

Then, we expand the derivatives from 3 points to 5 points. The error of approximation is small when the order of Taylor's series is high. However, the calculation would be so complicated for higher order approximation. Usually, 7-point and 5-point approximation of derivatives would be the rule of thumb. The 7-point derivative can be computed as :-

$$f_0' = \frac{-f_3 + 3f_2 + 3f_1 - 3f_{-1} - 3f_{-2} + f_3}{12h} + O(h^6)$$

Equation 6.11

$$f_0'' = \frac{-f_3 + 4f_2 + 17f_1 - 40f_0 + 17f_{-1} - 4f_{-2} - f_3}{24h^2} + O(h^6)$$

Equation 6.12

We may express the approximation in a filter form :-

a. Delta function $\vec{d}_0 = \frac{1}{7}[0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0]$

Equation 6.13

b. First order derivative $\vec{d}_1 = [-1 \quad 3 \quad 3 \quad 0 \quad -3 \quad -3 \quad 1]$

Equation 6.14

c. Second order derivative $\vec{d}_2 = [-1 \quad 4 \quad 17 \quad -40 \quad 17 \quad 4 \quad -1]$

Equation 6.15

The 2D filter is simply obtained by multiplying the 1D filter into the transpose itself :-

$$[D_u] = \vec{d}_0 \vec{d}_1^{\,T},\ [D_v] = \vec{d}_1 \vec{d}_0^{\,T},\ [D_{uu}] = \vec{d}_0 \vec{d}_2^{\,T},$$
$$[D_{vv}] = \vec{d}_2 \vec{d}_0^{\,T},\ [D_{uv}] = \vec{d}_1 \vec{d}_1^{\,T}$$

Equation 6.16

And, the complete estimation of derivatives can be expressed as :-

$$f_u] = [D_u] * [S] * [f],\ f_v] = [D_v] * [S] * [f],\ f_{uu}] = [D_{uu}] * [S] * [f],$$
$$f_{vv}] = [D_{vv}] * [S] * [f],\ f_{uv}] = [D_{uv}] * [S] * [f]$$

Equation 6.17

where [S] = lowpass filter; [f] = the image itself; [*] = convolution operator.

Since the convolutions is composed of row and column filter as defined in equation 6.16, we can reduce the computation by NxN convolutions into 2xN. This could be achieved by filtering the row components first using the row filter. Then, we can apply the column filter in the similar fashion. This is known as the separable convolution and can help us to reduce the computations in convolution.

As shown in figure 6.1, we chose the top view of an object to generate the range image. The range image was smoothed by a 7x7 Gaussian filter. Figure 6.2 to 6.6 show the partial derivatives for the range image. The Gaussian and mean curvature was generated by the convolution filter as the above and shown in figure 6.7 and 6.8 respectively. According to table 6.1, we can extract the features by the sign of Gaussian and mean curvatures. The corresponding feature map is plotted on figure 6.9.

6.2.2 Local Surface patching

Although separable filter can help us to minimize the computations from N^2 to 2N. However, the filter still requires tremendous computations. In addition, these kinds of filters have the property/characteristic to amplify noises in the original image. In order to minimize the noise, one may use the local surface. A surface can be expressed in quadratic form as equation 6.18 which could give the first and second derivatives to calculate the Gaussian and mean curvatures.

$$\vec{f} = a_{00} + a_{10}x + a_{01}y + a_{11}xy + a_{20}x^2 + a_{02}y^2$$

Equation 6.18

In order to fit the model of equation 6.18, we could use the standard technique of least mean square error approximation to minimize the error [38] :-

$$\begin{bmatrix} N & \sum_i x_i & \sum_i y_i^2 & \sum_i x_i y_i & \sum_i x_i^2 & \sum_i y_i^2 \\ \sum_i x_i & \sum_i x_i^2 & \sum_i x_i y_i & \sum_i x_i^2 y_i & \sum_i x_i^3 & \sum_i x_i y_i^2 \\ \sum_i y_i & \sum_i x_i y_i & \sum_i y_i^2 & \sum_i x_i y_i^2 & \sum_i x_i^2 y_i & \sum_i y_i^3 \\ \sum_i x_i^2 & \sum_i x_i^3 & \sum_i x_i^2 y_i & \sum_i x_i^3 y_i & \sum_i x_i^4 & \sum_i x_i^2 y_i^2 \\ \sum_i y_i^2 & \sum_i x_i y_i^2 & \sum_i y_i^3 & \sum_i x_i y_i^3 & \sum_i x_i y_i^2 & \sum_i y_i^4 \end{bmatrix} \begin{bmatrix} a_{00} \\ a_{10} \\ a_{01} \\ a_{11} \\ a_{20} \\ a_{02} \end{bmatrix} = \begin{bmatrix} \sum_i f_i \\ \sum_i x_i f_i \\ \sum_i y_i f_i \\ \sum_i x_i y_i f_i \\ \sum_i x_i^2 f_i \\ \sum_i y_i^2 f_i \end{bmatrix}$$

Equation 6.19

We can put equation 6.19 into matrix from :-

$$\bar{C}\vec{A} = \bar{Z}$$
$$\vec{A} = \bar{C}^{-1}\bar{Z}$$

Equation 6.20

According to equation 6.19, we could obtain the parameters in quadratic patching among each samples within the surface. The summations in $\bar{C}$ are summing the coordinates within a window.. Hence, by solving inverse of $\bar{C}$ and calculating $\bar{Z}$, we can obtain parameters of the quadratic surface equation. As $\bar{C}$ consists of the

global coordinates on the surface which changes at different locations. Thus, $\bar{C}$ would not be a constant matrix within the object surface. Consequently, operations on this matrix inversion involve a tremendous amount of computations. For example a 5x5 windows, the summations would take the ±2 samples vertically and horizontally. There would be totally 25x25 summing and each summing would involve 12 mutiplications. And if the surface consists of 128x128 points, there should be 128x128x25x25x12 = $1.22\text{x}10^8$ mutliplications and 128x128x25x25 = $1.02\text{x}10^7$ additions.

However, if we rearrange the equation 6.20 into another form

$$\bar{f}(x_0,y_0) = a_{00} + a_{10}(x-x_0) + a_{01}(y-y_0) + a_{11}(x-x_0)(y-y_0) + a_{20}(x-x_0)^2 + a_{02}(y-y_0)^2$$

Equation 6.21

That changes the origin from (0,0) to (x_0,y_0) which is the local surface patch. The whole surface contains a number of local surface patches, the sizes of which are the same. The local surface coordinates remain unchanged from one patch to another patch since the grid coordinates are the same with respect to their local origins. As a result, the coordinates for $\bar{C}$ are the same for different local patches and matrix inversion is only required to calculate once. The remaining parameters in $\bar{Z}$ changes from patch to patch since the matrix contains the range values which is different among patches. However, calculation of $\bar{Z}$ is simpler than $\bar{C}$ since it involves only accumulations of six variables, i.e. Σf_i, $\Sigma x_i fi$, $\Sigma y_i f_i$, $\Sigma x_i y_i f_i$, $\Sigma x_i^2 f_i$ and $\Sigma y_i^2 f_i$, among the local patches.

f_u,f_v, f_{uu}, f_{vv}, and f_{uv} at point (x_0,y_0) are calculated as substituting (x_0,y_0) for the partial derivatives. The results are expressed in equation 6.22.

$$\hat{f}_u = a_{10},\ \hat{f}_v = a_{01},\ \hat{f}_{uu} = 2a_{20},\ \hat{f}_{vv} = a_{02},\ \hat{f}_{uv} = a_{11}$$

Equation 6.22

Hence, the partial derivatives are simply obtained by the parameters of local surface patching, i.e. a_{10}, a_{01}, a_{20}, a_{02} and a_{11}. Figure 6.14 shows the range image of doughnut viewed at a side. The results of partial derivatives are plotted as figure 6.15-6.19 respectively.

We may draw the conclusion that both convolution and local surface patching could generate the partial derivatives for the purpose of curvature calculations. Convolution seems to be the most common method for such operation and which can be implemented by VLSI. However, it would require tremendous computations and amplify the noise component. Local surface patching seems to be suitable for generating the partial derivative since it involves only six accumulations for each patch among whole surface. The two methods were used to generate the curvatures of 128x128 images by a 80486-DX66 computer, we can see that local surface patching is faster than convolution filter in table 6.2.

Convolution filter	35 seconds
Local surface patching	22 seconds

Table 6.2-Elapsed times for calculating the partial derivatives.

6.3 Feature Extraction

After the Gaussian and mean curvatures are obtained, the features can be identified according to table 6.1. Kong [37] explained that only four types of features are essentially used for the range images, i.e. saddle-ridge, saddle-valley, peak and pit. Thus, we chose these kinds of features for characterization.

Figure 6.9 and 6.10 show that the extracted features from the range images which were taken by two viewing angles as indicated in figure 6.11. Accordingly, the common features for types of peak, flat, ridge, minimal and pit are indicated. Peak surface on figure 6.10 seems to be the easiest features for identification and it is formed in both characteristic views.

The features on figure 6.9 and 6.10 are extracted by convolution. As details of the features are subjected to the size of Gaussian filter. Figure 6.12 - 6.13 are the feature maps which were obtained by a 7x7 Gaussian filter. When comparing to those obtained by a 5x5 filter in figure 6.9 and 6.10, we could find that some of the detail components are missing as the filter size was increased.

6.4 Conclusion

The extracted features would be used to identify the common overlapping portion of each pair of characteristic views. And a higher level relational matching algorithm will be used to solve that common features and mapping function. However, such process assumes that a minimal set of features has been extracted.

Usually, typical features are needed to be represented as a minimal set of coordinates with character types rather than a map of feature. We should therefore reduce the representation of feature map into a feature set. Other surface characteristics may also be taken into an account for the best representatives, for example, boundary, area, surface directions. Although we have not worked out that feature representation, the basic toolsets for generating the visible-invariant features have been developed. Features can be generated by convolution and local surface patching. Convolution needs a tremendous computation power; however, it can be implemented on the VLSI devices. Local surface patching can reduce the operations in convolution filter but which is required a general computation unit.

Size of low-pass filter is important to feature extraction. Large scale of low-pass filter could suppress the noise component for the gradient filter. However, it reduce the resolution of resultant features. Alternatively, small scale of low-pass filter could give a good resolution for the features. However, the noise component may be amplified by the high-pass filter. It is necessary to detect the features by a multi-resolution approach. Therefore, we should extract the features by a large scale of low-pass filter first. Then, based on those extracted features, we search the details by decreasing the size of low-pass filter. Such process repeats until minimum low-pass filtering is applied.

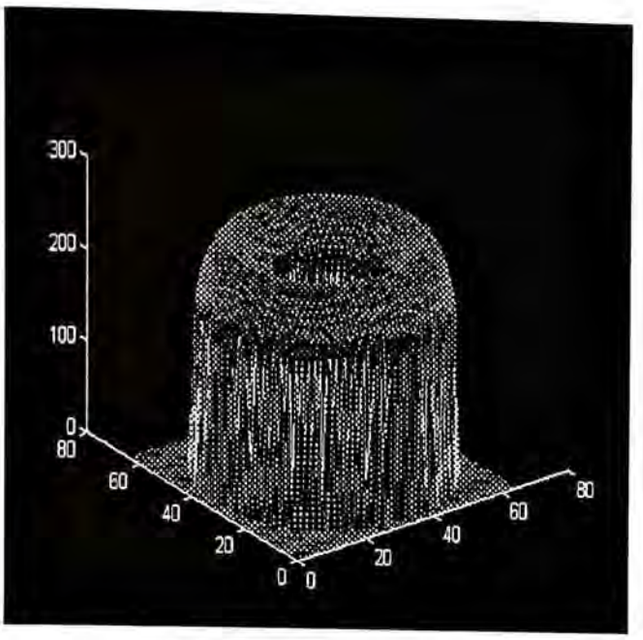


Figure 6.1 - Doughnut (mesh plot)

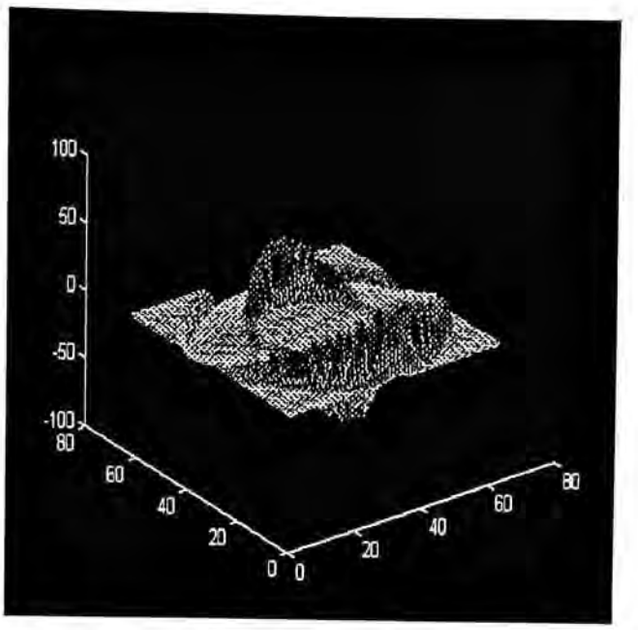


Figure 6.2 - Doughnut (f_u)

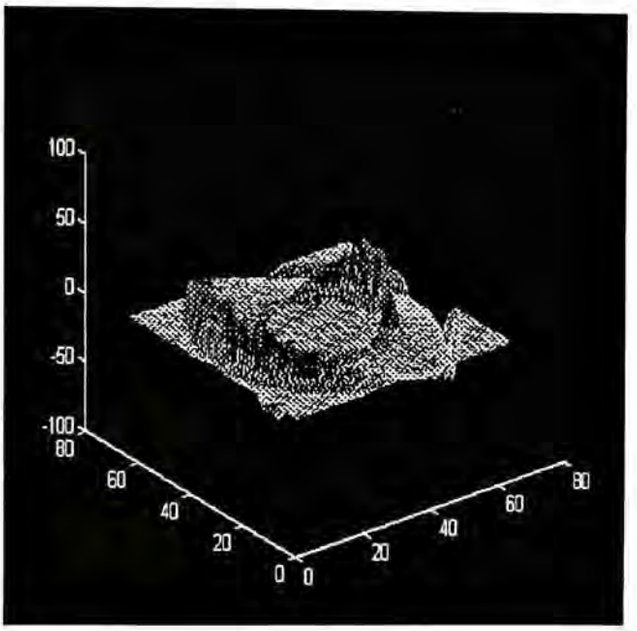


Figure 6.3 - Dougnut (f_v)

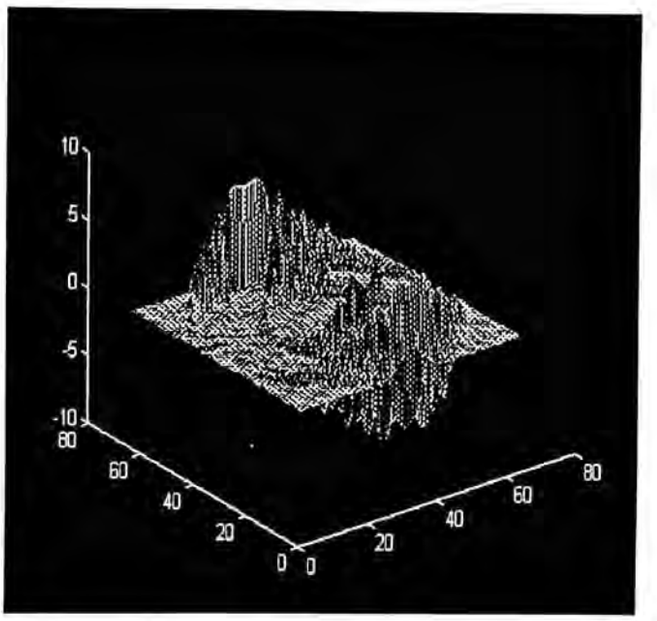


Figure 6.4 - Doughnut (f_{uu})

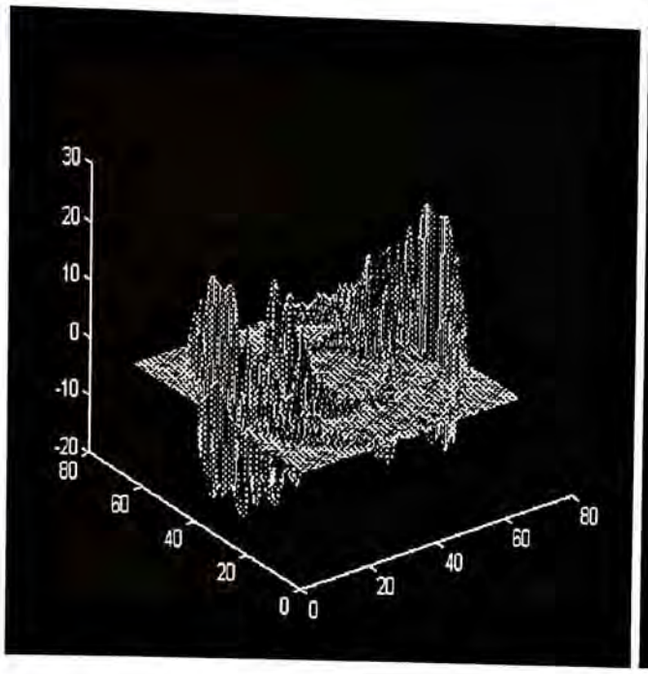


Figure 6.5 - Doughnut (f_{vv}).

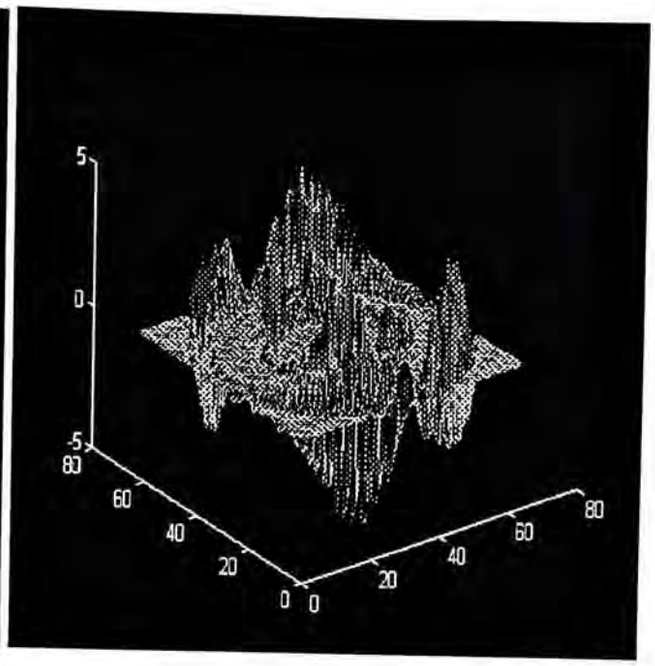


Figure 6.6 - Doughnut (f_{uv}).

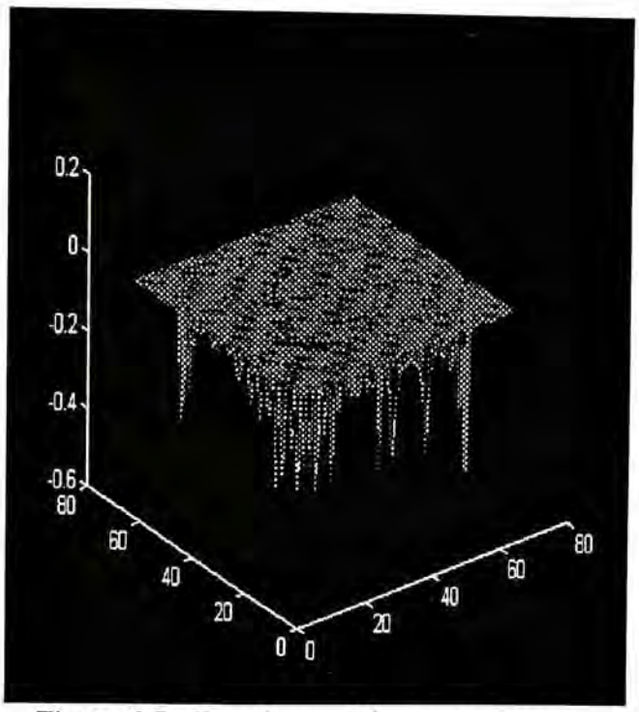


Figure 6.7 - Doughnut - Mean curvature.

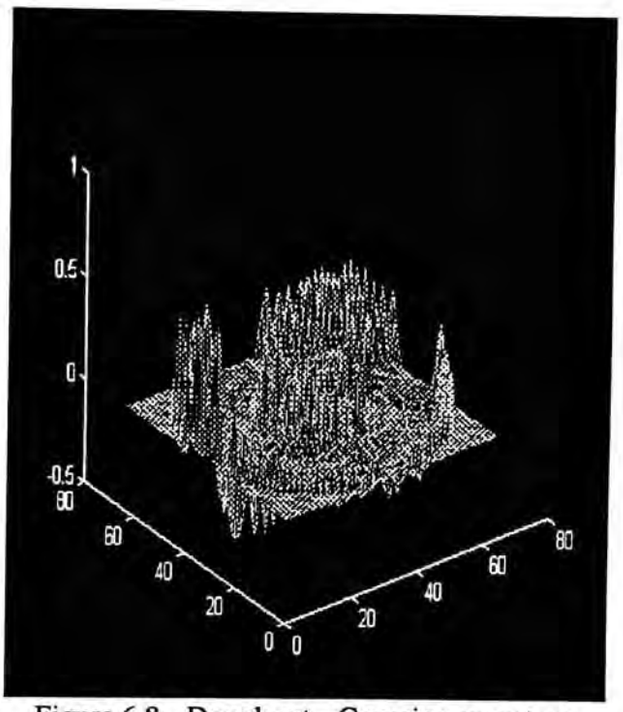


Figure 6.8 - Doughnut - Gaussian curvature.

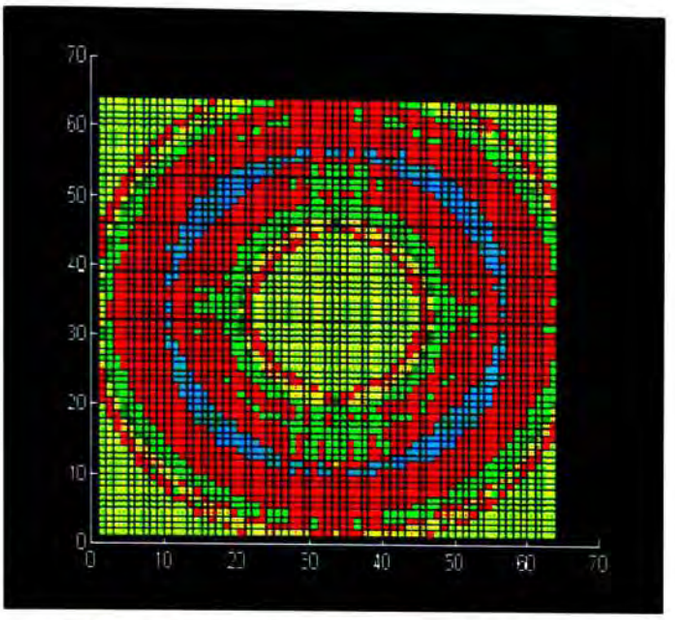


Figure 6.9 - Doughnut - Sign map.

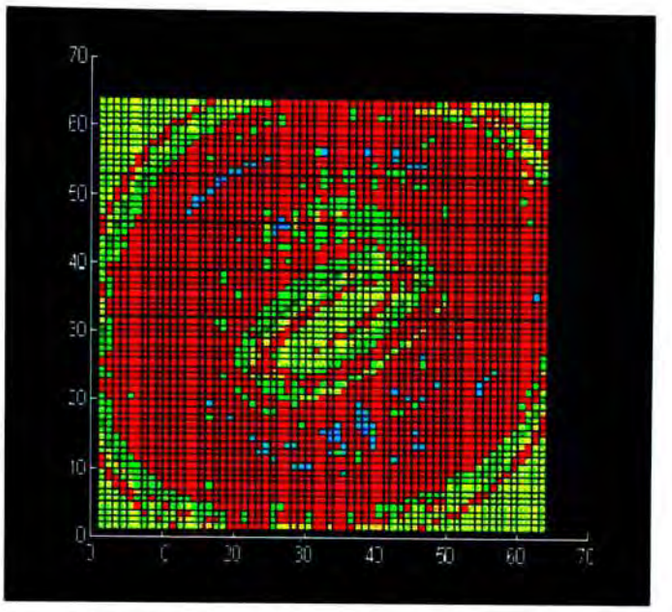


Figure 6.10 - Doughnut - Sign map.

Color Table : -
Yellow = Flat Surface
Blue = Peak Surface
Red = Ridge Surface
Purple = Minimal Surface
Green = Pit surface

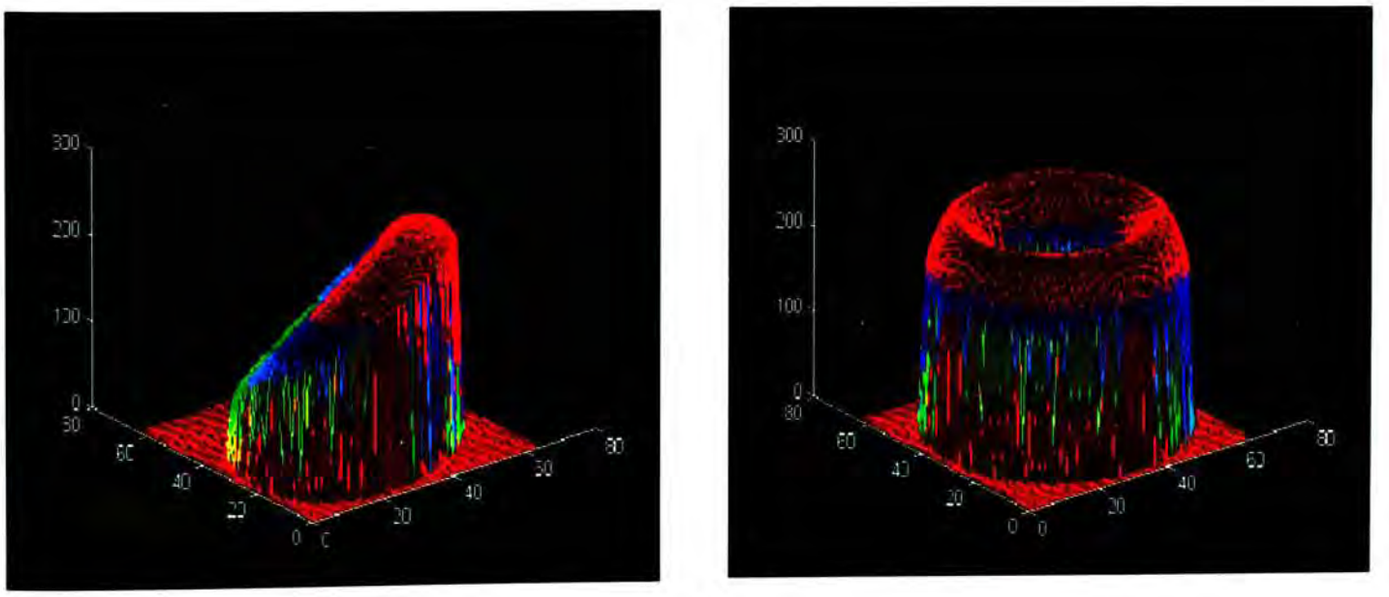


Figure 6.11 - Doughnut - Sign Map.

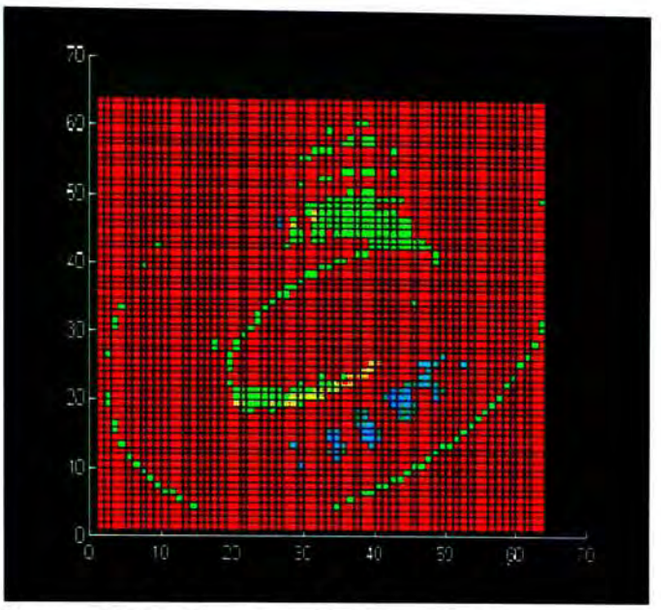


Figure 6.12 - Features for Guassian Filter = 7x7 (Side view).

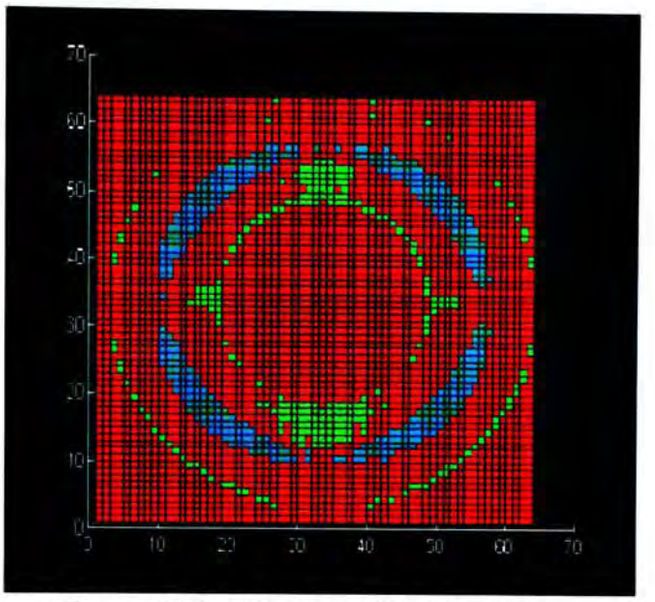


Figure 6.13 - Features for Guassian Filter = 7x7 (Top view).

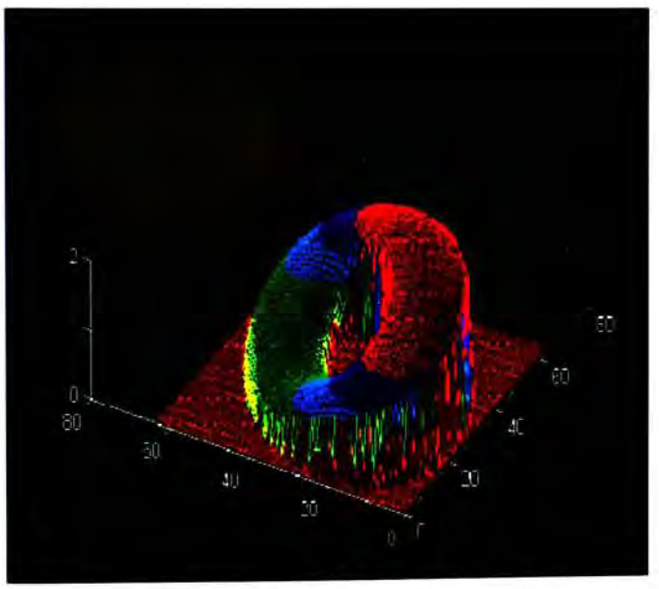


Figure 6.14 - Doughnut by Surface Patching.

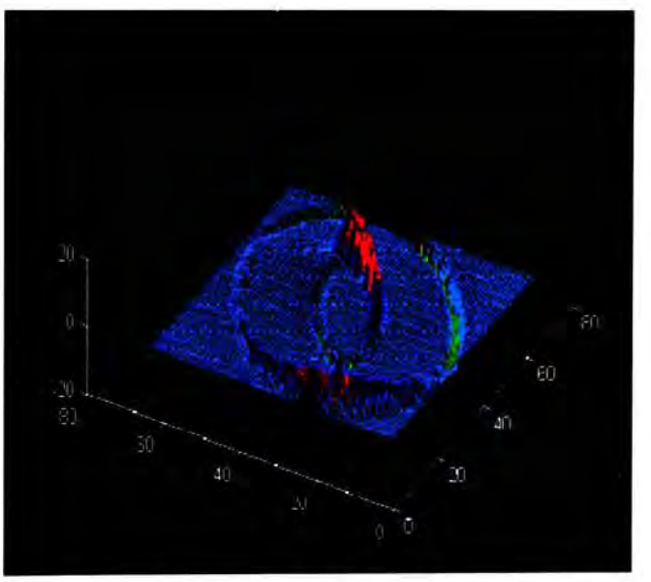


Figure 6.15 - Doughnut (a_{01}).

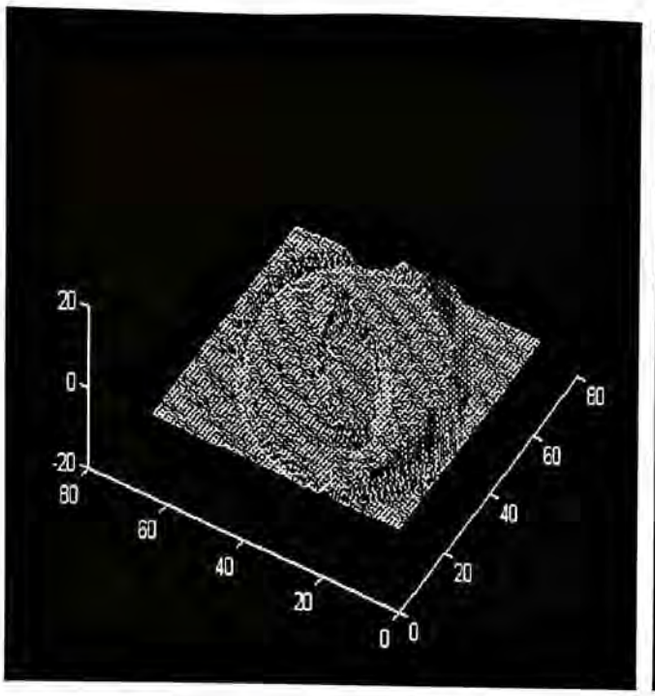


Figure 6.16 - Doughnut (a_{10}).

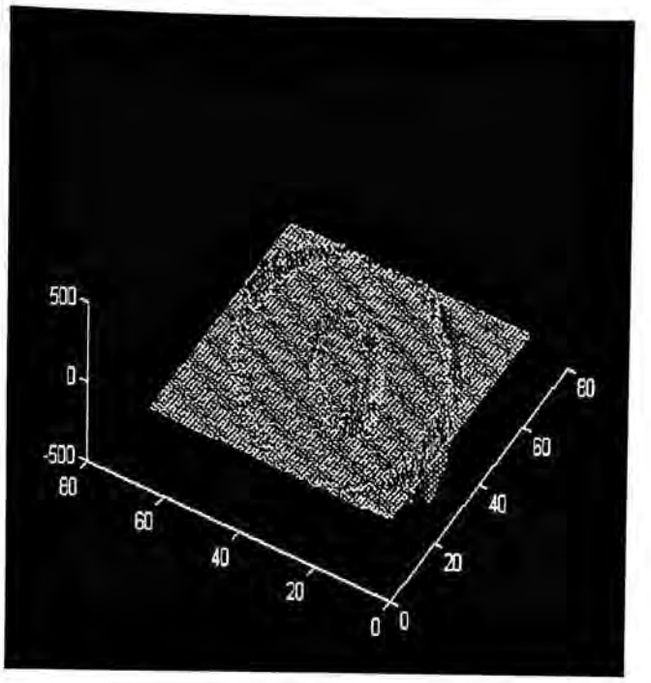


Figure 6.17 - Doughnut (a_{20}).

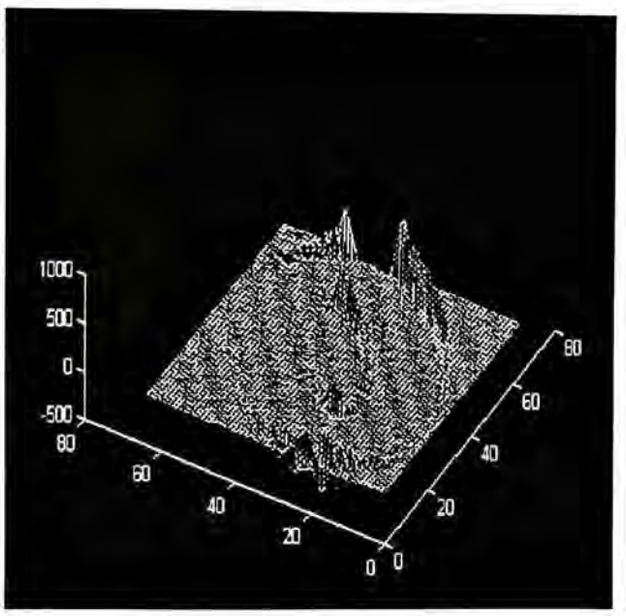


Figure 6.18 - Doughnut (a_{02}).

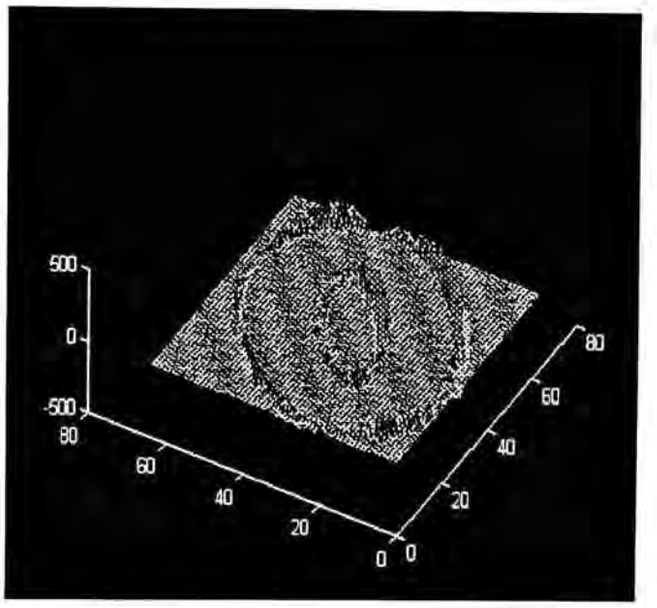


Figure 6.19 - Doughnut (a_{11}).

Chapter 7

Merging Multiple Characteristic Views

By utilizing the tools which have been developed for generating range images and extracting visible-invariant features of arbitrary 3D objects, we can then reconstruct the global 3D object. The reconstruction is archieved by transforming the characteristic views (CVs) at different view angles into a single view as the reference coordinate system which could be anyone of the captured views as the reference coordinate system. The other views are then transformed inversely back to the reference view by merging a pair of characteristic views, i.e. reference and other view. The pair of characteristic views must contain common features. Theoretically, at least four points common feature should exist in order to solve the homogeneous transformation matrix.

Recently, a rigid body model (RBM) has been proposed [39] for representing 3D objects and probabilistic relaxation matching has been employed for recognition of partially occluded objects. Merging to form a RBM can also be used by a match and transform process to a sequence of CVs [31]. In this chapter, we are going to discuss how we could merge the multiple characteristic views into a single coordinate system.

7.1 Rigid Body Model

As described in chapter 6, features in a characteristic view can be extracted using differential geometry, i.e. Gaussian and mean curvatures. Since Gaussian and

mean curvatures are both visible-invariant, the same features on the common overlapping portion could be obtained from different characteristic views. The feature list on a characteristic view is then grouped into a model for the purpose of matching. Since the feature lists are taken from different characteristic views, the coordinate systems in different characteristic views are also different. Hence, our model must be capable of handling such kind of transformations, e.g. rotation and translation.

The Rigid Body Model [39] (RBM) is based on those features extracted by the differential geometry. It is assumed that the object is a rigid body which is not deformable in any viewing transformation. Thus, the distances among the features on common overlapping portion should remain the same. RBM is a kind of model which is capable of handling arbitrary viewing transformation. This may come to our requirement of the property of visible-invariant.

Basically, RBM is the list of feature nodes on each characteristic views with their types and also viewport coordinates. The viewport coordinates are used to tabulate their distances among themselves. Figure 7.1 gives an illustration on the construction of RBM. An object is captured by two 2D views, namely "view-1" and "view-2" which are extracted with some features. The features a, b, c and d are the common features on the overlapping portion ***C*** and 1, 2, 3, 4, 5, and 6 are the other features not visible to both characteristic views. The common features are simply translated from view-2 to view-1 for this example. Two models can be created based on those features as :-

$S_1 = \{ f_a^1, f_b^1, f_c^1, f_d^1, f_1^1, f_2^1, f_3^1, d_{aa}^1, d_{ab}^1, d_{ac}^1, d_{ad}^1, d_{a1}^1, d_{a2}^1, d_{a3}^1, d_{ba}^1, .., d_{b3}^1, .., d_{3a}^1, .., d_{33}^1 \}$

$S_2 = \{ f_a^2, f_b^2, f_c^2, f_d^2, f_4^2, f_5^2, f_6^2, d_{aa}^2, d_{ab}^2, d_{ac}^2, d_{ad}^2, d_{a4}^2, d_{a5}^2, d_{a6}^2, d_{ba}^2, .., d_{b6}^2, .., d_{6a}^2, .., d_{66}^2 \}$

where S_i = the Rigid Body Model of view-i;

f_j^i = the types of features in node j for view-i;

d_{jk}^i = the distance between feature nodes j and k for view-i.

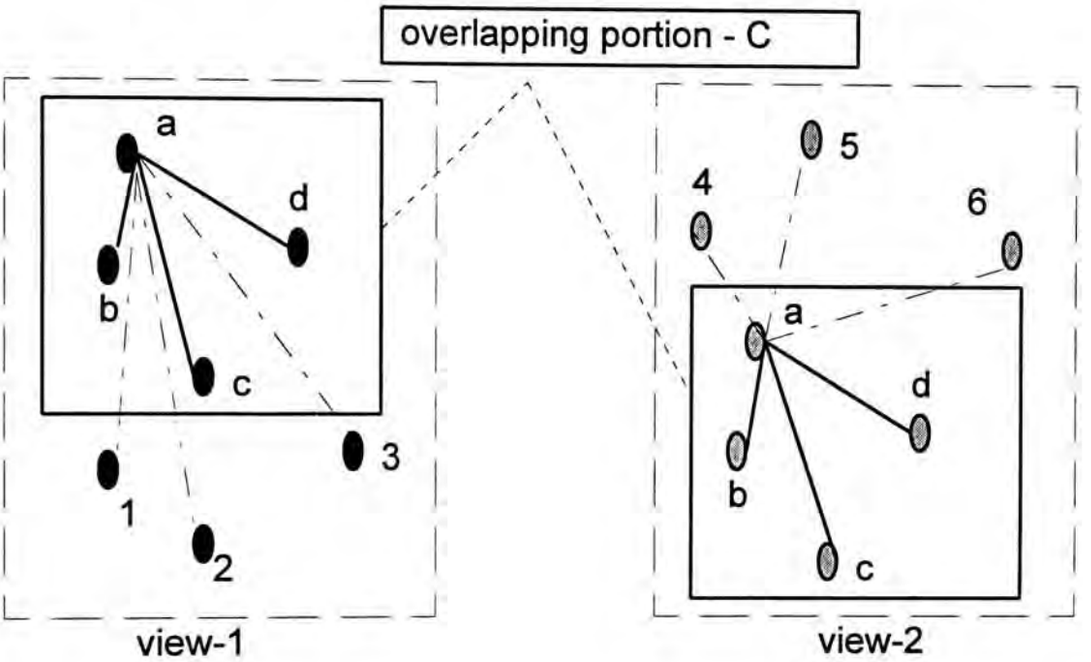


Figure 7.1 - Feature map of two characteristic views (CV).

The exact RBM is actually implemented with only the feature types and the viewport coordinates. The distances among nodes are then tabulated each time when the model is read. Calculation on tabulating distance values can be reduced from NxN to $\sum_{i=1}^{N} i$ times since d_{jk}^i is identical to d_{kj} and d_{jj} equals to zero. Thus, the tabulation only involves of d_{jk}^i where $j < k$ and $j \neq k$ which can much enhance the computational speed.

7.2 Sub-Rigid Body Model

The Sub-Rigid Body Model (SRBM) is defined as a subset of RBM. SRBM is the partial view of the original 3d object. In order to solve out the transformation of merging multiple characteristic views, the SRBM must be overlapped at least with one another. In the examples of figure 7.1, region ***C*** on two views is the overlapping area of view-1 and view-2. Again, with rigidity constraint (rigid body assumption), the distance among feature nodes should be visible-invariant to viewer's coordinate. Since SRBM is the sub-set of RBM, this property should also apply for transitional and rotational transformation. From figure 7.1, The overlapping features ***a, b, c,*** and ***d*** are simply translated from view-1 to view-2. We can easily find the common SRBM for region ***C*** in figure 7.1, which is :-

$$S_{12} = \{ f_a^{12}, f_b^{12}, f_c^{12}, f_d^{12}, d_{aa}^{12}, d_{ab}^{12}, d_{ac}^{12}, d_{ad}^{12}, d_{ba}^{12}, .., d_{bd}^{12}, .., d_{da}^{12}, .., d_{dd}^{12} \}$$

The matching is dependent on the distance constraints. We take an example in node ***a*** on both views. Node a^1 in view-1 is said to be matched node a^2 since we can find:-

$$d_{aa}^{1} = d_{aa}^{2}=0,\ d_{ab}^{1} = d_{ab}^{2},\ d_{ac}^{1} = d_{ac}^{2},\ d_{ad}^{1} = d_{ad}^{2};\ \text{and}$$

$$d_{a1}^{1} \neq d_{a4}^{2} \neq d_{a5}^{2} \neq d_{a6}^{2},\ d_{a2}^{1} \neq d_{a4}^{2} \neq d_{a5}^{2} \neq d_{a6}^{2},\ d_{a3}^{1} \neq d_{a4}^{2} \neq d_{a5}^{2} \neq d_{a6}^{2}.$$

The pair of features are said to be matched because there are three distances from nodes b, c, and d to node a on view-1 is identical to view-2. However, node a^1 in view-1 is not said to be matched with node b^2 in view-2 since we can only find the distance matches :-

$$d_{aa}^{1} = d_{bb}^{2}=0,\ d_{ab}^{1} = d_{ba}^{2}$$

Also, the distances among other nodes cannot be matched (i.e., a-c, a-d, a-1, a-2, a-3):-

$$d_{ac}^{\ 1} \neq d_{ba}^{\ 2} \neq d_{bb}^{\ 2} \neq d_{bc}^{\ 2} \neq d_{bd}^{\ 2} \neq d_{b4}^{\ 2} \neq d_{b5}^{\ 2} \neq d_{b6}^{\ 2}$$

$$d_{ad}^{\ 1} \neq d_{ba}^{\ 2} \neq d_{bb}^{\ 2} \neq d_{bc}^{\ 2} \neq d_{bd}^{\ 2} \neq d_{b4}^{\ 2} \neq d_{b5}^{\ 2} \neq d_{b6}^{\ 2}$$

$$d_{a1}^{\ 1} \neq d_{ba}^{\ 2} \neq d_{bb}^{\ 2} \neq d_{bc}^{\ 2} \neq d_{bd}^{\ 2} \neq d_{b4}^{\ 2} \neq d_{b5}^{\ 2} \neq d_{b6}^{\ 2}$$

$$d_{a2}^{\ 1} \neq d_{ba}^{\ 2} \neq d_{bb}^{\ 2} \neq d_{bc}^{\ 2} \neq d_{bd}^{\ 2} \neq d_{b4}^{\ 2} \neq d_{b5}^{\ 2} \neq d_{b6}^{\ 2}$$

$$d_{a3}^{\ 1} \neq d_{ba}^{\ 2} \neq d_{bb}^{\ 2} \neq d_{bc}^{\ 2} \neq d_{bd}^{\ 2} \neq d_{b4}^{\ 2} \neq d_{b5}^{\ 2} \neq d_{b6}^{\ 2}$$

However, the situations are not good enough when comparing to a^1 to a^2. Similarly, node a^1 is not matched to nodes c^2, d^2, 4^2, 5^2, 6^2 for the same reasons. In an other word, the feature is matched for a maximum score of closed distances among other nodes from one view to another. However, the above example is just simple illustration on matching the common features. A more general approach for RBM/SRBM matching will be discussed on the following section.

7.3 Probabilistic Relaxation Matching

The common features of two sub-rigid bodies can be matched by the probabilistic relaxation process which was established by Rosenfeld et al [41]. After identifying the overlapping regions, a mapping can be computed which will create a process for merging the multiple characteristic views into a single 3D object.

Suppose that there are two RBMs which are taken by views V_1 and V_2 respectively. The types of two RBMs are $\{ f_1^{\,i}, f_2^{\,i}, f_3^{\,i}, ..., f_{N1}^{\,i} \}$ and $\{ g_1^{\,i}, g_2^{\,i}, g_3^{\,i}, ..., g_{N2}^{\,i} \}$, where N_1 is the total number of features on view-1 and N_2 is the number on view-2 respectively. In order to identify the mapping function for those views, we should first create a probability vector

$$\bar{p}_i = [p(g_1 = f_i), p(g_2 = f_i), \ldots, p(g_{N2} = f_i)]^T$$

Equation 7.1

which is constructed for $N_1 \geq i \geq 1$. The component $p(g_m=f_i)$ indicates the probability that feature f_i on V_1 is matched to g_m on V_2. Rosenfeld at al. [41] suggested that we can utilize the distances among those features as spatial constraints to update this probability vector by a spatial vector, Q. Such that if d_{ij} is closed to d_{mn}, the probabilities $p(g_m=f_i)$ and $p(g_n=f_j)$ would be then enhanced. Meanwhile, the other vectors are suppressed by the corresponding Q vector. Hence, we can update the probability vector p_i from k^{th} to $k+1^{th}$ component according to the rule

$$\bar{p}_i^{k+1}(g_m = f_i) = \frac{\bar{p}_i^{k}(g_m = f_i)\bar{Q}_i^{k}(g_m = f_i)}{\sum_j \bar{p}_j^{k}(g_m = f_j)\bar{Q}_j^{k}(g_m = f_j)}$$

Equation 7.2

where $$\bar{Q}_i(g_m = f_i) = \sum_j c_{ij} \sum_n r_{ij}(m,n)\bar{p}_j^{k}(g_n = f_j);$$

Equation 7.3

C_{ij} = the weighting factor for the degree of influence from f_i to f_j;

$r_{ij}(n,m)$ = the compatibility coefficients.

$r_{ij}(n,m)$ is the key terms which determines the degree of enhancement or suppression of each component of probability vector according to the spatial constraints. Since we assume those features on each characteristic view has the equally weighted influence affecting the updating mechanism, C_{ij} is therefore chosen as a constant in both numerator and denominator and could be canceled out in equation 7.1.

Let the error between spatial constraints d_{ij} and d_{nm} be $\varepsilon = | d_{ij} - d_{nm} |$. The compatibility coefficients are some values in between 0 and 1, which should [39] :-

1) increases monotonically as ε decreases;

2) approach 0 as ε approaches infinity;

3) approach 1 as ε approaches 0;

4) the rate of change at ε near 0 is small, say 0.

The last condition is necessary to reduce the noise sensitivity of the compatibility coefficients because even if i=n and j=m, d_{ij} and d_{nm} would still have some difference for errors in extracting the features on different views. This unexpected noise is eliminated when ε is near zero. Many functions could be used to obtain the above objectives, e.g. Gaussian function. It has been shown in [39] that the exponential function:-

$$r_{ij}(n,m) = \exp\left(-\lambda \left\{ \frac{|d_{ij} - d_{nm}|}{|d_{ij} + d_{nm}} \right\}^2 \right)$$

Equation 7.4

could satisfy the above conditions. According to [39], λ is an arbitrary constant with range 3-15 which would have small amount of influence on the overall performance. We selected as λ=5 for our experiments.

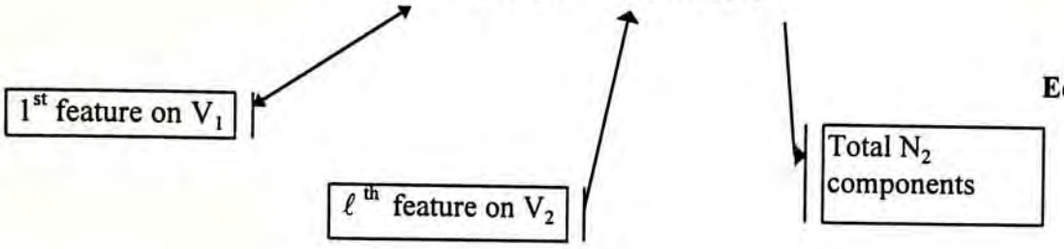
Therefore, we have defined the mechanism for updating the probability vectors from k^{th} to $k+1^{th}$ component. For the initial probability vector, we assumed that all the features on V_1 are equally probable to match the features on V_2. So, the probability vector for p_1^0 is :-

$$\bar{p}_1{}^0 = [\frac{1}{N_2}, \frac{1}{N_2}, .., \frac{1}{N_2}, \frac{1}{N_2}]^T$$

Equation 7.5

By applying the equation 7.2, we can update $p_1{}^0$ to $p_1{}^1$. The iteration would be carried on until one of the component in the vector approaches to unity. For example, if we stop the iteration on k^{th} probability vector, then $p_1{}^k$ would appear as:-

$$\bar{p}_1{}^k = [0,0,0,.....1,....,0,0]$$



Equation 7.6

We can say that feature *1* on V_1 is matched to the feature ℓ on V_2. Similarly, we could find that mapping function for all features on V_1 which are matched with the corresponding features on V_2, i.e. $G_1(i) = j$. For equation 7.6, the mapping function is $G_1(1) = \ell$.

Base upon the results of probabilistic relaxation, we could find the mapping function for two characteristic views. It is assumed that all the features on one view are matched to other view. However, some of the features are located on the non-overlapping portion. We visualized and discarded these non-overlapping features on the SGI workstation manually. All we need is that transformation for mapping of common overlapping portion only in order to merge those unexposed characteristic views.

7.4 Merging the Sub-rigid Body Model

We have developed the mechanism for determining mapping function of the common overlapping portion. Therefore, the two characteristic views could be merged into a single 3D object by the given mapping function. Since the one coordinate system of one view is related to another coordinate system :-

$$\begin{bmatrix} x^1 & y^1 & z^1 & 1 \end{bmatrix} = \begin{bmatrix} x^2 & y^2 & z^2 & 1 \end{bmatrix} \begin{bmatrix} T_{00} & T_{01} & T_{02} & 0 \\ T_{10} & T_{11} & T_{12} & 0 \\ T_{20} & T_{21} & T_{22} & 0 \\ T_{30} & T_{31} & T_{32} & 1 \end{bmatrix}$$

Equation 7.7

where $[x_n^{\,i}\ y_n^{\,i}\ z_n^{\,i}]$ is the coordinate of feature ***n*** of view ***i***;

[T] = the homogenous transformation, 4x4 matrix.

and, if we have four points of matching, the transformation could be written as :-

$$\begin{bmatrix} x_1^{\,1} & y_1^{\,1} & z_1^{\,1} & 1 \\ x_2^{\,1} & y_2^{\,1} & z_2^{\,1} & 1 \\ x_3^{\,1} & y_3^{\,1} & z_3^{\,1} & 1 \\ x_4^{\,1} & y_4^{\,1} & z_4^{\,1} & 1 \end{bmatrix} = \begin{bmatrix} x_1^{\,2} & y_1^{\,2} & z_1^{\,2} & 1 \\ x_2^{\,2} & y_2^{\,2} & z_2^{\,2} & 1 \\ x_3^{\,2} & y_3^{\,2} & z_3^{\,2} & 1 \\ x_4^{\,2} & y_4^{\,2} & z_4^{\,2} & 1 \end{bmatrix} \begin{bmatrix} T_{00} & T_{01} & T_{02} & 0 \\ T_{10} & T_{11} & T_{12} & 0 \\ T_{20} & T_{21} & T_{22} & 0 \\ T_{30} & T_{31} & T_{32} & 1 \end{bmatrix}$$

$$[\bar{v}_1] = [\bar{v}_2][\bar{T}]$$

Equation 7.8

It is quite straightforward to obtain the homogenous transformation as follows :-

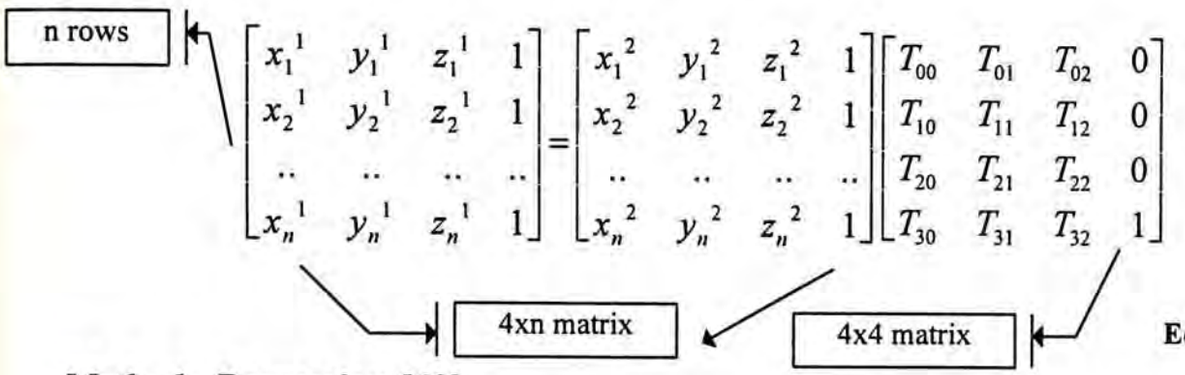
$$[\bar{T}] = [\bar{v}_2]^{-1}[\bar{v}_1]$$

Equation 7.9

Since the transformation is a 4x4 matrix, we must have four points of matching in order to solve matrix inversion. However, if the matching is more than four points, matrices $[\bar{v}_1]$ and $[\bar{v}_2]$ have more rows than columns and the equation is an overdetermined system. The solution is the matrix [T] which best fits to the

hyperplane in a least squares sense. In another word, we could use the extra points to minimize the error when merging those characteristic views.

Supposed that the mapping function could be solved as ***n*** pairs of matching. Then we could rewrite $[\bar{v}_1]$ and $[\bar{v}_2]$ in equation 7.8 as they are changed from 4x4 to 4xn matrix as shown in equation 7.10.



$$\begin{bmatrix} x_1^1 & y_1^1 & z_1^1 & 1 \\ x_2^1 & y_2^1 & z_2^1 & 1 \\ .. & .. & .. & .. \\ x_n^1 & y_n^1 & z_n^1 & 1 \end{bmatrix} = \begin{bmatrix} x_1^2 & y_1^2 & z_1^2 & 1 \\ x_2^2 & y_2^2 & z_2^2 & 1 \\ .. & .. & .. & .. \\ x_n^2 & y_n^2 & z_n^2 & 1 \end{bmatrix} \begin{bmatrix} T_{00} & T_{01} & T_{02} & 0 \\ T_{10} & T_{11} & T_{12} & 0 \\ T_{20} & T_{21} & T_{22} & 0 \\ T_{30} & T_{31} & T_{32} & 1 \end{bmatrix}$$

Equation 7.10

Method - Regression [40]

Since the matrix in equation 7.10 is not a square matrix, we cannot directly apply matrix inversion to solve the homogenous transformation. One of solution for that equation may come to the numerical method - regression. If we pre-multiply the two sides of equation 7.9 with the transpose of $[\bar{v}_2]^T$, we could have a square matrix.

$$[\bar{v}_2]^T[\bar{v}_1] = [\bar{v}_2]^T[\bar{v}_2][\bar{T}]$$

Equation 7.11

Let us define $[\bar{S}] = [\bar{v}_2]^T[\bar{v}_2]$ which is a square matrix of 4x4. We could rewrite equation 7.11 into :-

$$[\bar{v}_2]^T[\bar{v}_1] = [\bar{S}][\bar{T}]$$

Equation 7.12

Thus, the homogenous transformation could be solved out as follows :-

$$[\bar{T}] = [\bar{S}]^{-1}[\bar{v}_2]^T[\bar{v}_1]$$

$$[\bar{T}] = \left[[\bar{v}_2]^T [\bar{v}_2] \right]^{-1} [\bar{v}_2]^T [\bar{v}_1]$$

Equation 7.13

Since we assume that the features on each characteristic view is a rigid body, we could therefore transform all the non-overlapping portion into a single coordinate system. The transformation tries to minimize the spatial errors among those features for transforming from one characteristic view to another.

7.5 Illustrations

Let us demonstrate how effective when applying probabilistic relaxation to solve the mapping function in common overlapping portion of two characteristic views. Figure 7.2 (a) shows the global object of X29 plane which generated those range images as shown in figures 7.2 (b), 7.3 and 7.4. The characteristic views for the range images were selected such that some of the surfaces were commonly exposed to the viewer. So, we could merge those views into the single coordinate system. Accordingly, we could extract the features on those characteristic views as shown in figures 7.5, 7.6 and 7.7.

We take the examples of figures 7.6 and 7.7, namely CV_1 and CV_2 respectively, and try to merge them into a single coordinate system. The features of the two CVs are listed as tables 7.1 and 7.2 respectively. The coordinate systems of two characteristic views are defined in a unit cube viewport.

Feature node	x	y	z
1	0.55	0.99	0,27
2	0.65	0.36	0.36
3	0.51	0.73	0.32
4	0.51	0.43	0.30
5	0.44	0.31	0.39
6	0.74	0.45	0.45
7	0.62	0.26	0.32
8	0.64	0.40	0.33
9	0.54	0.00	0.28
10	0.61	0.07	0.32
11	0.58	0.63	0.32
12	0.52	0.22	0.29
13	0.52	0.05	0.30
14	0.60	0.51	0.31

Table 7.1 - Feature node listing of CV_1.

Feature node	x	y	z
1	0.35	0.49	0.54
2	0.48	0.42	0.43
3	0.48	0.27	0.43
4	0.45	0.56	0.43
5	0.55	0.99	0.32
6	0.49	0.66	0.33
7	0.55	0.01	0.34
8	0.45	0.05	0.20
9	0.50	0.47	0.36

Table 7.2- Feature node listing of CV_2.

In order to merge the characteristic views into a single coordinate system, we must first solve the mapping function of them. Thus, we applied the probabilistic relaxation matching those features. The result of matching CV_1 and CV_2 is tabulated as table 7.3.

View - CV_1	View - CV_2
5	1
1	6
3	7
2	8
6	11
4	2
8	5
7	9
9	14

Table 7.3 - Mapping function for V_1 to V_2.

Figures 7.8 and 7.9 show the mapping graph of the corresponding results. The features of CV_1 are indicated gray circles with green numbers, whereas CV_2 are indicated as red numbers. The individual mappings are indicated with yellow lines.

According to that mapping function after probabilistic relaxation, we could use the method of regression to obtain the homogenous transform matrix. The result for the corresponding mapping function is listed as equation 7.14 :-

$$[\bar{T}_{12}] = \begin{bmatrix} 0.13 & -0.04 & 0.86 & 0 \\ -0.08 & 1.02 & 0.10 & 0 \\ -1.17 & 0.06 & 0.19 & 0 \\ 0.84 & 0.00 & -0.26 & 1 \end{bmatrix}$$

Equation 7.14

Figures 7.10 and 7.11 show that the result for merging these two characteristic views of CV_1 and CV_2. The white portion corresponds to CV_1, whereas red portion to CV_2.

7.6 Merging Multiple Characteristic Views

In order to merge the multiple characteristic views into a single coordinate system, we could apply the probabilistic relaxation to match the other views into one of the merged views. As the example for the above, we could create another mapping function for converting characteristic view of figure 7.2 (b) (namely CV_0) to CV_1. The features of CV_0 are tabulated in table 7.4.

Feature node	x	y	z
1	0.51	0.98	0.37
2	0.76	0.46	0.34
3	0.24	0.46	0.34
4	0.38	0.56	0.34
5	0.63	0.56	0.34
6	0.51	0.09	0.54
7	0.44	0.08	0.33
8	0.56	0.08	0.33
9	0.50	0.49	0.43
10	0.39	0.42	0.32
11	0.62	0.41	0.33

Table 7.4 - Feature list of CV_0.

We could therefore find the mapping function for CV_0 to CV_1. The result of mapping is tabulated on table 7.5 and homogeneous transform on equation 7.15.

View - CV_1	View - CV_0
1	1
2	5
3	4
4	9
8	2
7	3
10	8
9	6
6	11

Table 7.5 - Mapping function for CV_1 to CV_0.

$$[T_{10}]=\begin{bmatrix} 0.70 & -0.24 & -0.88 & 0 \\ 0.00 & 1.01 & 0.01 & 0 \\ 0.67 & 0.26 & 0.87 & 0 \\ -0.06 & 0.06 & 0.59 & 1 \end{bmatrix}$$

Equation 7.15

We could merge those CVs into a single view. The merging sequence is listed on table 7.6 which has taken index 0 as the base view. The merged views of V_0, V_1, and V_2 are shown on figure 7.12 and 7.13.

Index	Range Image	Transformation	Matching pair of view	Direction
0	X29-0	-	-	-
1	X29-1	T_{10}	1->0	Inverse
2	X29-2	T_{12}	1->2	Forward

Table 7.6 - Merging sequence for CVs (1->0 & 2->1->0).

7.7 Mislocation of Feature Extraction

There are some mis-alignment for merging those views. The errors are caused by the slight spatial difference on extracting the features on each characteristic. For example, feature $f_0^1=(x_0^1\ y_0^1\ z_0^1)$ on view-1 is said to be matched to feature $f_n^2=(x_n^2+\delta_x,\ y_n^2+\delta_y,\ z_n^2+\delta_z)$ on view-2, where δ_x, δ_y, δ_z are the error due to feature extraction. Thus, if we solve the transform in between these two CVs according to this matching, the matrix would transform f_0^1 to f_n^2 with the error of δ_x, δ_y, δ_z respectively. This error is caused by the mislocation for feature extraction. In the following paragraph, we are going to explore how the mislocation affecting the merging and matching processing.

7.7.1 The Transform Matrix for Perfect Matching

In order to prove the error of the experiments are caused by the mislocation of feature extraction, we would like to calculate the transform matrix for the ideal case of the perfect matching. We would take this ideal transform for our reference to justify the results. The coordinates in each CV are confined with a unit cube which are defined by the viewport and viewer transform of corresponding views.

$$[\bar{v}] = [\bar{w}][\bar{M}][\bar{V}]$$

Equation 7.16

where $[\bar{v}]$ = coordinate in characteristic views;

$[\bar{w}]$ = global 3D coordinate;

$[\bar{M}]$ = the viewer's transform matrix;

$[\bar{V}]$ =viewport transform matrix.

To transform the CV_0 into their original 3D global coordinates, we could take the inverse transform as follows :-

$$[\bar{w}] = [\bar{v}_0][[\bar{M}_0][\bar{V}_0]]^{-1}$$

Equation 7.17

where $[\bar{M}_0]$ = the viewer's transform matrix in CV_0;

$[\bar{V}_0]$ =viewport transform matrix in CV_0.

Therefore, we could transform the CV_0 to CV_1 according to equation 7.16 :-

$$[\bar{v}_1] = [\bar{v}_0][[\bar{M}_0][\bar{V}_0]]^{-1}[[\bar{M}_1][\bar{V}_1]]$$

Equation 7.18

The matrix $[[\bar{M}_0][\bar{V}_0]]^{-1}[[\bar{M}_1][\bar{V}_1]]$ is the transformation from CV_0 to CV_1. Since the viewer and viewport transforms are saved with each CV, we could therefore

calculate the ideal matrix according to equation 7.18. For the above examples, we could obtain the matrices as follows :-

$$[\bar{T}_{10}]=\begin{bmatrix} 0.72 & 0 & -1.03 & 0 \\ 0 & 1 & 0 & 0 \\ 0.65 & 0 & 1.01 & 0 \\ -0.07 & 0 & 0.64 & 1 \end{bmatrix}$$

Equation 7.19

and,

$$[\bar{T}_{12}]=\begin{bmatrix} -0.08 & 0 & 1.13 & 0 \\ 0 & 1 & 0 & 0 \\ -0.93 & 0 & -0.09 & 0 \\ 0.84 & 0 & -0.26 & 1 \end{bmatrix}$$

Equation 7.20

Comparing with equations 7.14 & 15, there are significant errors for those two transformations which introduced the mis-alignments in merged views as shown in figures 7.15 and 7.16. Since we could use the original transformation matrices for each CV, the ideal merged object could be obtained by transforming all the CVs into the global 3D object according to equation 7.17. Figures 7.15 and 7.16 show the ideal merging of the X29 plane which merging sequence is defined as table 7.7.

Index	Range Image	Transformation	Matching pair of view	Direction
0	X29-0	M_0V_0	World->0	Inverse
1	X29-1	M_1V_1	World->1	Inverse
2	X29-2	M_2V_2	World->2	Inverse

Table 7.7 - Merging sequence of X29 CVs to 3D world.

where $[\bar{M}_0\bar{V}_0]=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -0.7 & 1 \end{bmatrix}$; and $[\bar{M}_1\bar{V}_1]=\begin{bmatrix} 0.72 & 0 & 0.69 & 0 \\ 0 & 1 & 0 & 0 \\ -0.69 & 0 & 0.72 & 0 \\ 0 & 0 & -0.97 & 1 \end{bmatrix}$;

$$[\bar{M}_2\bar{V}_2]=\begin{bmatrix} 0.13 & -0.04 & 0.86 & 0 \\ -0.08 & 1.02 & 0.10 & 0 \\ -1.17 & 0.06 & 0.19 & 0 \\ 0.84 & 0 & -0.26 & 1 \end{bmatrix}.$$

7.7.2 Introducing The Errors in Feature Set

We would like to carry on the experiments to prove that the mislocation in feature extraction could be led to significant errors in merging. First of all, the feature set of CV_1 would be converted to CV_0 by utilizing the saved transformation matrix according to equation 7.18. The transformed feature sets are tabulated on tables 7.8.

Feature node	x	y	z
1	0.50	0.99	0.35
2	0.64	0.55	0.33
3	0.50	0.73	0.44
4	0.50	0.43	0.42
5	0.50	0.03	0.59
6	0.76	0.45	0.33
7	0.58	0.26	0.32
8	0.61	0.40	0.32
9	0.51	0.01	0.36
10	0.58	0.07	0.33
11	0.56	0.63	0.36
12	0.50	0.22	0.40
13	0.50	0.05	0.41
14	0.57	0.51	0.33

Table 7.8 - Transformation of features in CV_0 to CV_1..

The mapping graph is shown in figure 7.17. The range map together with feature set are also shown in figure 7.18 and 7.19 for corresponding views respectively. The feature set of CV_0 transformed into CV_1 are also calculated as tabulated in table 7.9.

Feature node	x	y	z
1	0.77	0.49	0.33
2	0.61	0.42	0.32
3	0.61	0.27	0.32
4	0.63	0.56	0.34
5	0.46	0.99	0.36
6	0.54	0.66	0.40
7	0.51	0.01	0.34
8	0.50	0.05	0.57
9	0.55	0.47	0.36

Table 7.9 - Transformation of features in CV_2 to CV_0.

Then, we rearranged the order of the two feature sets and applied the rigid-body relaxation matching to these two feature sets. The mapping functions could be found and shown in tables 7.10 and 7.11 respectively.

View - CV_0	View - CV_1
1	1
3	3
6	6
11	11
7	7
12	12
4	4
5	5
8	8
10	10
14	14
9	9
13	13
2	2

Table 7.10 - Mapping Function of CV_0 to CV_1.

View - CV_0	View - CV_2
3	3
5	5
1	1
4	4
6	6
2	2
9	9
8	8
7	7

Table 7.11 - Mapping Function of CV_0 to CV_2.

Therefore, we could solve the transform matrices for these two mapping functions :-

$$[\bar{T}_{01}]=\begin{bmatrix} 0.72 & 0 & 0.74 & 0 \\ 0 & 1 & 0 & 0 \\ -0.46 & 0 & 0.52 & 0 \\ 0.34 & 0 & -0.28 & 1 \end{bmatrix}$$

Equation 7.21

and,

$$[\bar{T}_{02}]=\begin{bmatrix} -0.75 & 0 & 0.75 & 0 \\ 1 & 0 & 0 & 0 \\ -0.44 & 0 & -0.57 & 0 \\ 1.07 & 0 & 0.15 & 1 \end{bmatrix}$$

Equation 7.22

We would then compare these results with the perfect matching as the viewer and viewport are known. The matrices are given by equation 7.18 as follows :-

$$[\bar{T}_{01}]=\left[[\bar{M}_1\bar{V}_1]^{-1}[\bar{M}_0\bar{V}_0]\right]=\begin{bmatrix} 0.72 & 0 & 0.74 & 0 \\ 0 & 1 & 0 & 0 \\ -0.46 & 0 & 0.52 & 0 \\ 0.34 & 0 & -0.28 & 1 \end{bmatrix}$$

Equation 7.23

$$[\bar{T}_{02}] = [[\bar{M}_2\bar{V}_2]^{-1}[\bar{M}_0\bar{V}_0]] = \begin{bmatrix} -0.75 & 0 & 0.75 & 0 \\ 1 & 0 & 0 & 0 \\ -0.44 & 0 & -0.57 & 0 \\ 1.07 & 0 & 0.15 & 1 \end{bmatrix}$$

Equation 7.24

where $[\bar{M}_0\bar{V}_0]$, $[\bar{M}_1\bar{V}_1]$ and $[\bar{M}_2\bar{V}_2]$ are given in table 7.7.

The merging sequence for equations 7.21 and 7.22 is listed on table 7.12 :-

Index	Range Image	Transform Matrix	Matching Pair	Direction
0	X29-0	-	-	-
1	X29-1	T_{01} (eq 7.21)	0->1	Inverse
2	X29-2	T_{02} (eq 7.22)	0->2	Inverse

Table 7.12 - Merging sequence for ideal case (CV_0 + CV_1 + CV_2).

The merged views are shown in figures 7.21 and 7.22 for corresponding merging sequence. According to equation 7.23 and equation 7.24, the results are exactly the same as perfect matching. Thus, if the extraction could locate with the exact coordinates for identical features, our designed process of relaxation matching and homogenous merging could give the exact results as the same as perfect matching.

Then, we are going to see how the mislocation in feature extraction would alter the result in homogenous merging. The features 1, 2, 3 and 11 in table 7.8 were chosen and included some errors in the coordinates :-

Feature node	x	y	z
1	0.50	0.89 (-0.1)	0.47 (+0.12)
2	0.64	0.55	0.48 (+0.15)
3	0.50	0.73	0.39 (-0.05)
11	0.56	0.63	0.31 (-0.05)

Table 7.13 - Introducing errors in feature extraction.

The other nodes of the features set were unchanged. Figures 7.23 and 7.24 show the original feature maps and the one with error respectively. Then, we applied the relaxation matching to solve the mapping function :-

View - CV_0	View - CV_1
1	1
3	3
6	6
11	11
7	7
12	12
4	4
5	5
8	8
10	10
14	14
9	9
13	13
2	2

Table 7.14 - Mapping function for introducing errors.

The results of relaxation would be the same as table 7.10. According to table 7.14, we could find the transform matrix for corresponding mapping function :-

$$[\bar{T}^{e}{}_{01}] = \begin{bmatrix} 0.85 & -0.07 & 0.60 & 0 \\ 0.02 & 1.06 & -0.03 & 0 \\ -0.26 & 0.08 & 0.29 & 0 \\ 0.19 & -0.01 & -0.11 & 1 \end{bmatrix}$$

Equation 7.25

The merged view is shown in figure 7.25. We could see that there is significant error when comparing the result with equation 7.21. However, our matching would be the ideal one. Hence, we could conclude that the mislocation in feature extraction could lead to significant error in merging.

7.8 Summary

In this chapter, we have designed the system for reconstructing a single 3D object by the multiple range images of different characteristic views. We have also demonstrated how effective and efficient on matching and merging those views into a single coordinate system.

The system is based on the rigid-body relaxation matching to solve the mapping function of two characteristic views. After obtaining the mapping function, we could apply the generalized matrix inversion to solve the homogeneous transform. Therefore, each characteristic view could be transformed into another views. Accordingly, we could merge the multiple characteristic views as a global object which consists of only single coordinate system.

Figure 7.2 (a) - X29 object.

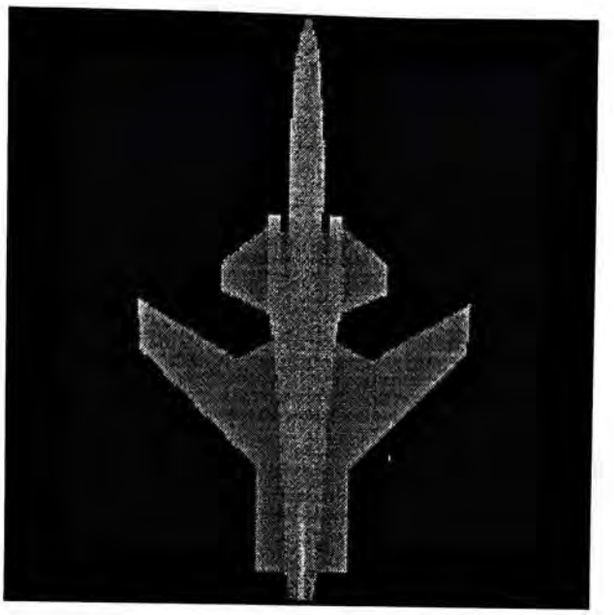


Figure 7.2 (b)- View V_0 of X29

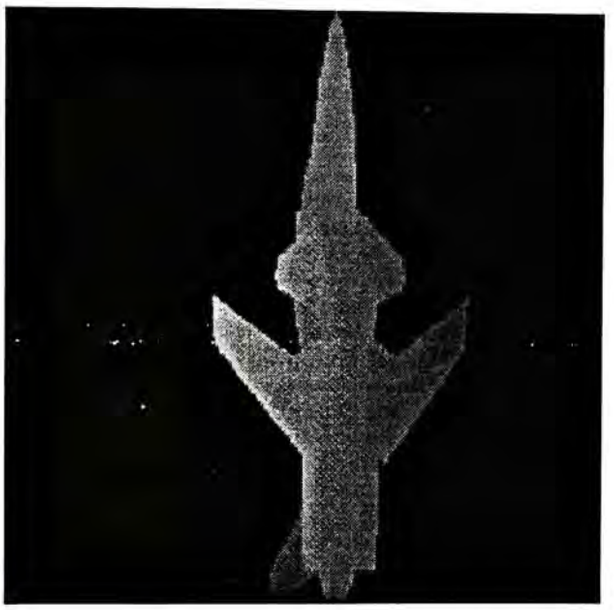


Figure 7.3 - View V_1 of X29.

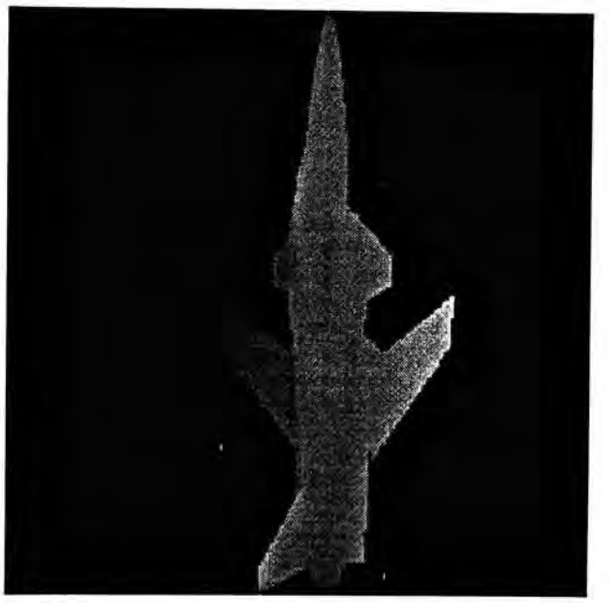


Figure 7.4 - View V_2 of X29

Figure 7.5 - CV + features (V_0).

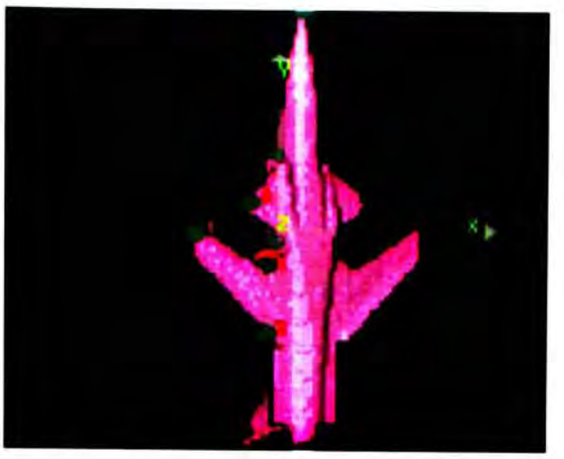


Figure 7.6 - CV + features (V_1).

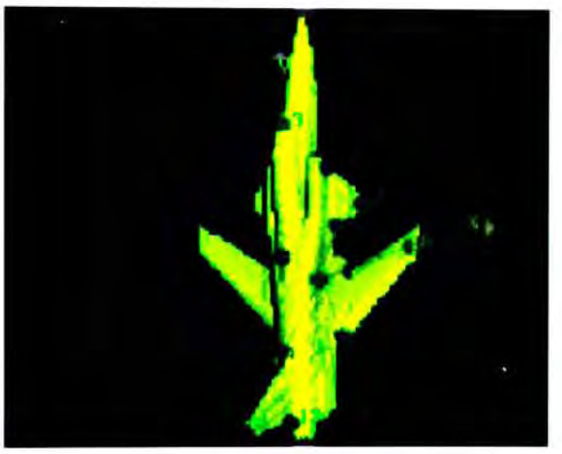


Figure 7.7 - CV + features (V_2).

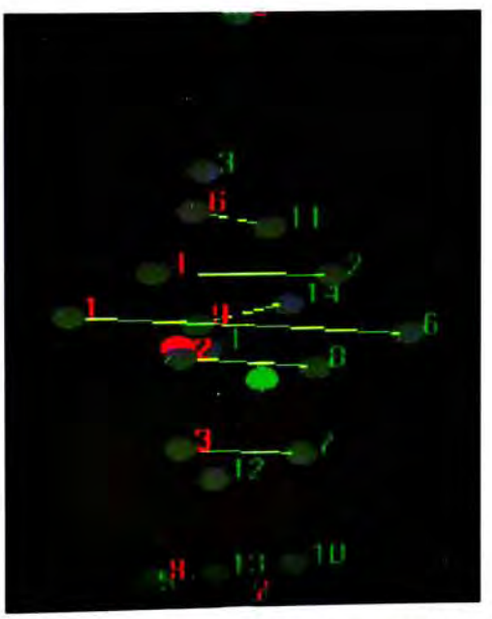


Figure 7.8 - Mapping graph V_1 to V_2.

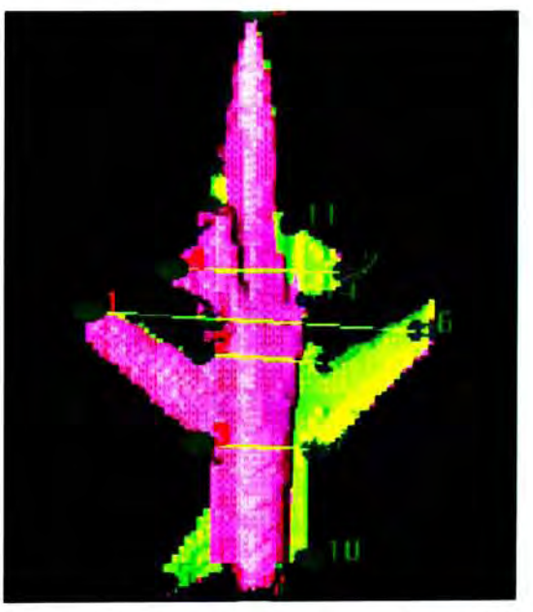


Figure 7.9 - Mapping graph + CV.

Figure 7.10 - Merged CV = $V_1 + V_2$.

White portion = V1.
Red portion = V2.

Figure 7.11 - Merged CV at other view.

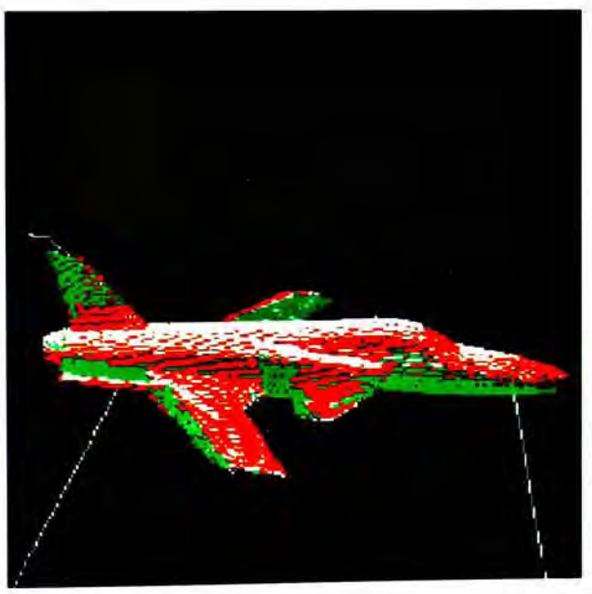


Figure 7.12 - Merged CV = $V_0 + V_1 + V_2$.

White portion = V0;
Green portion = V1;
Red portion = V2.

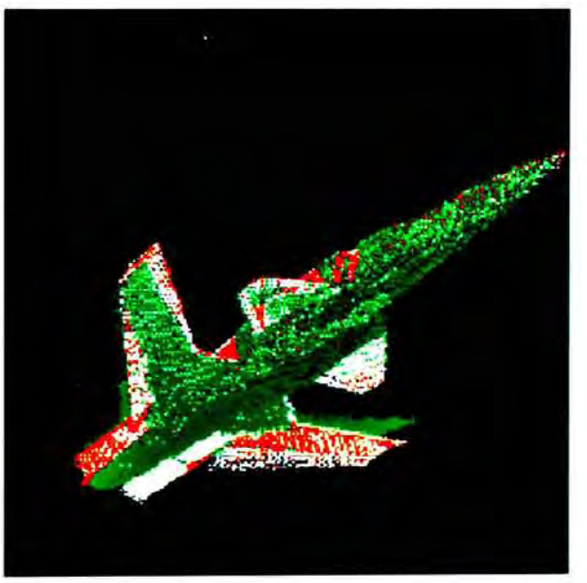


Figure 7.13 - Merged CV at other view.

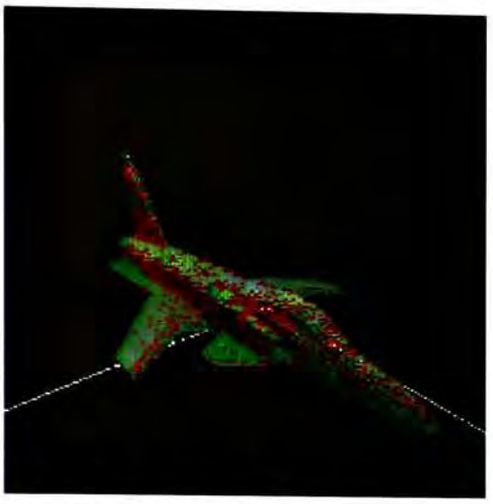


Figure 7.15 - Ideal Merging.

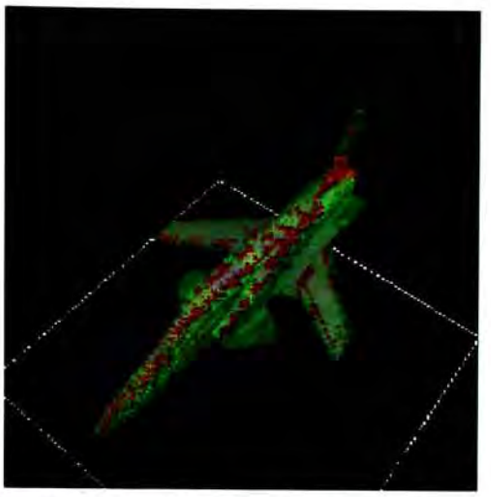


Figure 7.16 -Ideal Merging - another view.

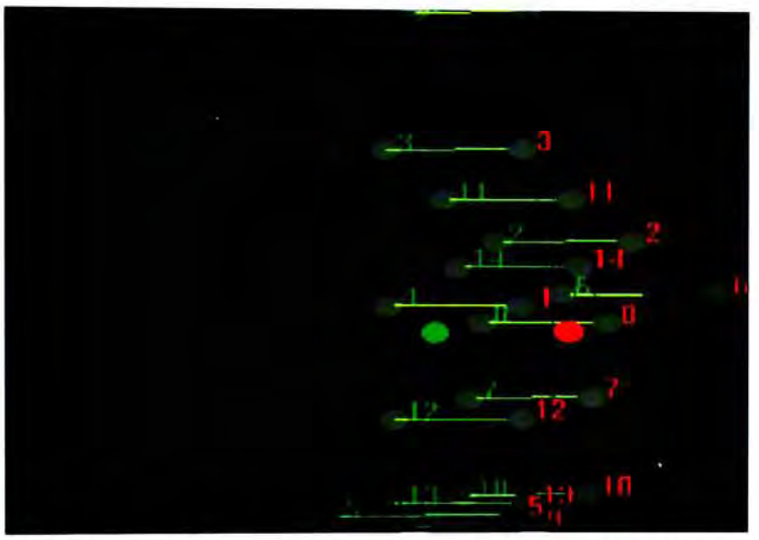


Figure 7.17 - Matching graph for ideal transform.

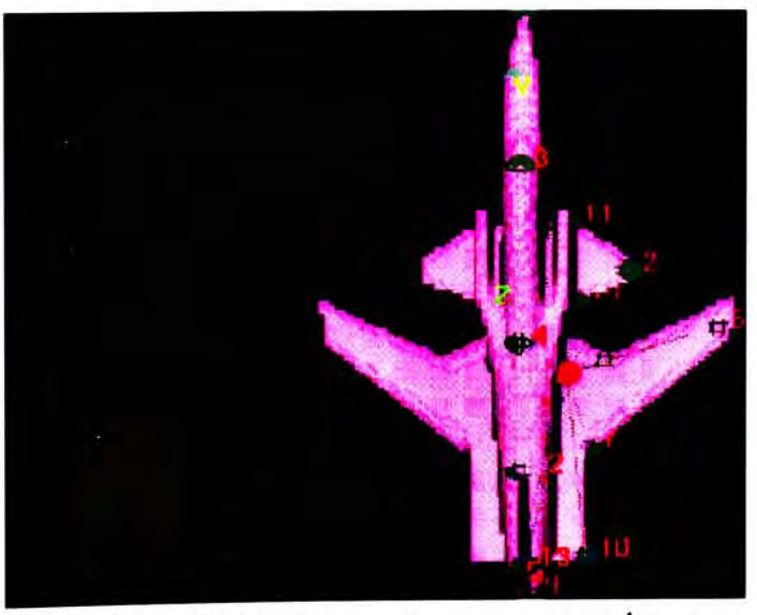


Figure 7.18 - Feature map + range image - view-0.

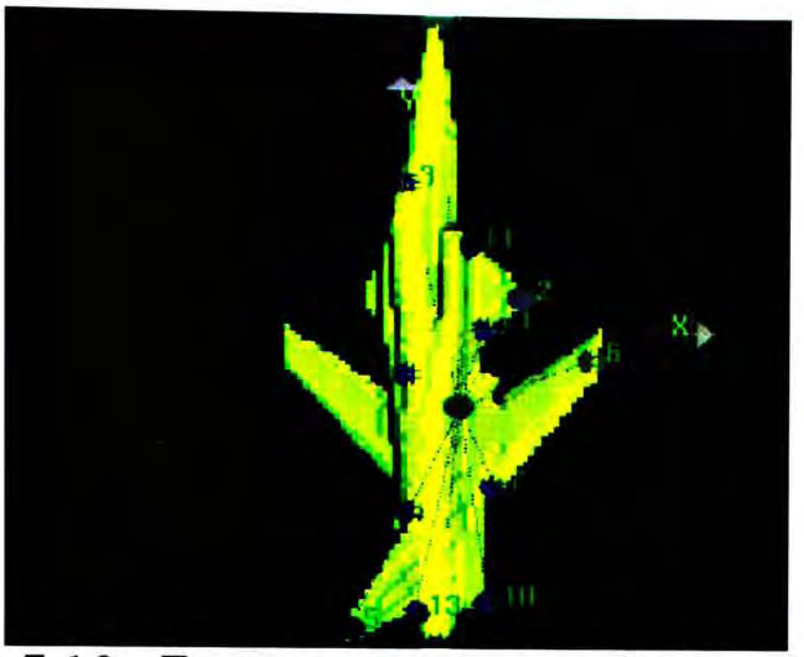


Figure 7.19 - Feature map + range image - view-1.

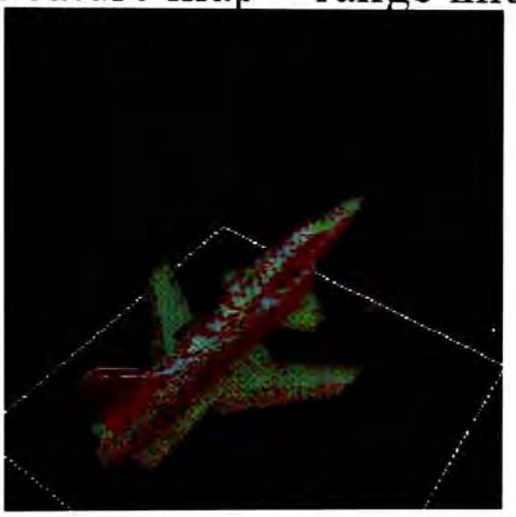


Figure 7.20 - Merged views = view 0 + 1 (ideal case).

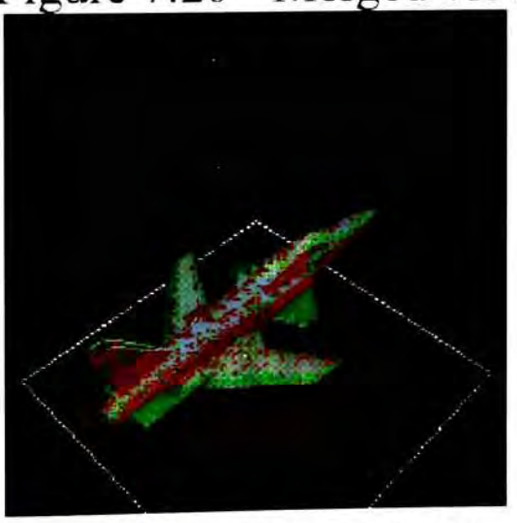


Figure 7.21 - Merged views = view 0 + 1 + 2 (ideal case)

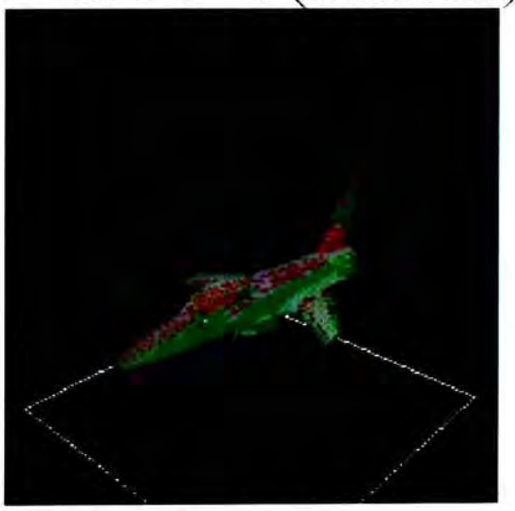


Figure 7.22 - Merged at another angle.

Figure 7.23 - Error on extraction - feature 1,2,3 and 11

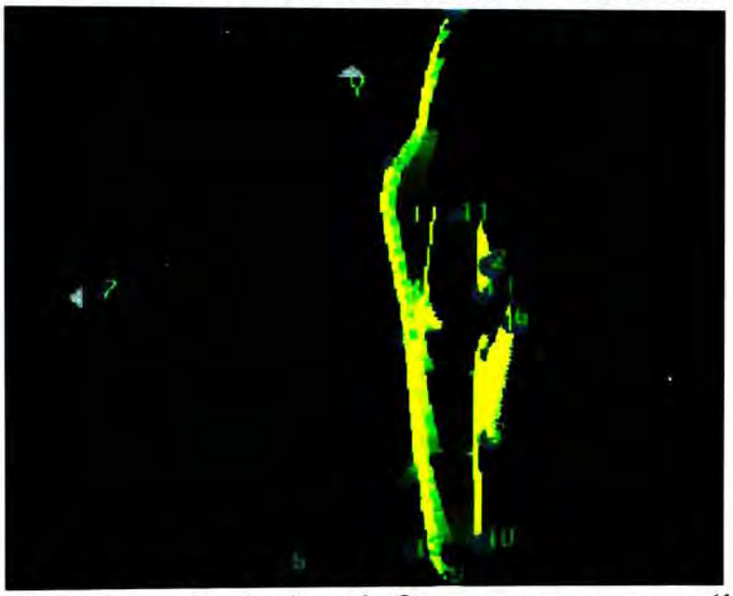


Figure 7.24 - Original features map (ideal case)

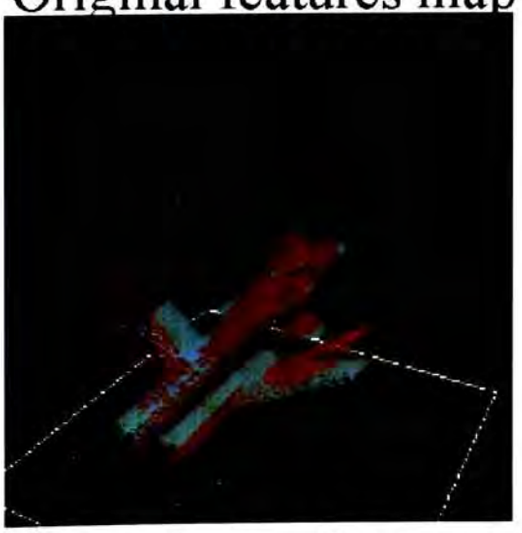


Figure 7.25 - Merged view for error in extraction.

Chapter 8

Conclusion

In this last chapter, I would like to make a summary of those previous chapters. The basic researches of acquisition, modeling and recognition of 3D objects has been addressed. Experimental results of active triangulation rangefinder have been presented. A 3D object reconstruction algorithm based on the data of rangefinder have been demonstrated.

The active rangefinder have been recognized as the direct and accurate method for acquiring 3D objects. A prototype based on the theory of triangulation has been designed in order to evaluate and verify the basic operations of that system. Although the prototype is still in early stage development and has not been proved as an accurate system, the basic concepts have been verified and proven.

Improvements

Since one of the objectives is to improve the performance of the old system, we would like to list those improvements for the new system. First of all, since the new system employs the movable optics rather than the x-y table as the scanning mechanism, the scanning speed can be archieved faster. Secondly, since the new system scans 3D objects by deflecting the laser beam, the object size can be variable as adjusting the scanning distance. Hence, the spatial resolution is dependent on the angular step and the scanning distance. Whereas, the old system employs the x-y table which has put restriction on resolution and object size according to the step size of

table. Thirdly, in the old system, we must change the position and orientation of object in order to capture the multiple views. Such operation could result in losing the global coordinates of the original objects. As scanning mechanism in the new one does not require moving the object, the existing state of object can be preserved as changing the viewing angle. Finally, since the movable optical accessories in the new system are manufactured in a small and compact scanning head, our rangefinder is potential to be embedded in robot arms or autonomous vehicle. On the contrast, since the old system must be fixed with scanning table, it is not appropriate for such kind of applications.

Drawbacks

Certain drawbacks exist for the new system when comparing with the old one. At first, since the movable optics are precise equipment with low inertia, they are very sensitive to vibration and impact. Secondly, the area of laser spot in the surface is changed as different incident angle. Since the range image is scanned by deflecting the laser beam, the excited area is also varied across the whole scanning region. This may cause the range/intensity cross-talk during acquisition. Finally, as scanning is adjusted by tiling the angle of scanning mirrors, a small angular error may cause a large positional error. Hence, the scanning head must provide angular feedback for the controller to ensure that the error is minimal and the positions are repeatable. Furthre details on the drawbacks of the developed system can be found in the performance analysis described in Appendix B.

Components

Components of the rangefinder are chosen to be commonly available ones. The components are listed as table 8.1.

Component	Purpose
CCD square pixel camera (CCIR-50Hz)	Image Capture
Image Frame Grabber (512x512)	Image Digitization
X-Y Scanning Head	Deflection of laser beam
Diode laser (0.5μW)	Light source
Scanning Mirror Controller	Scanning head control
Optical Table	Mounting of Optical Accessories
8086 Computer	Central Processing Unit

Table 8.1 - Rangefinder system breakdown.

The range value is determined by searching the maxima of returned laser beam. By all means, higher model of the central processing unit, e.g. 80386 or 80486 could lead to a higher speed of searching and hence quicker acquisition. However, the bottle-neck of the rangefinder was not so critically due to searching. We discovered that most of the time was wasted on settling the scanning head, especially for the point scanning operation. Thus, we inserted a delay when the laser beam traveling in between each step. Otherwise, the laser spot could be overshot when traveling for a large step and resulted in the errors. The problem was possibly due to improper setting of the gain of scanning mirror controller. However, we have not found out the proper value for the gain.

Reconstruction for 3D object from partial views

This is our next achievement of reconstruction for 3D object from the partial views of range images which are captured by the rangefinder. Figure 8.1 shows the system for reconstructing a 3D object from partial views of a rangefinder.

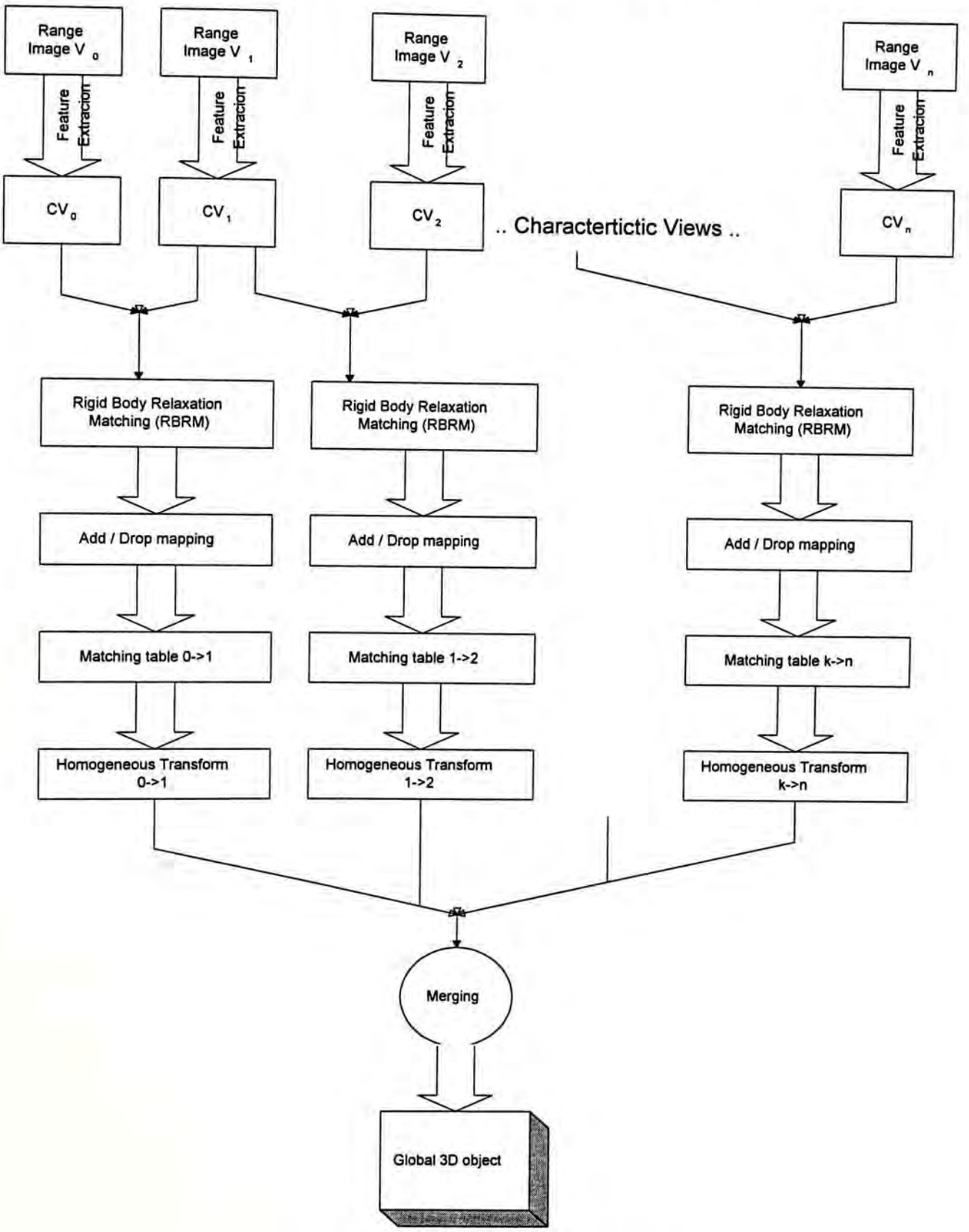


Figure 8.1 - 3D object reconstruction system.

In order to reconstruct the 3D object, several range images with some overlapping area are acquired. The range images are first extracted with their own features. Then, the features are tabulated with distances among others and also used as our rigid body model. Since the rigid body model is visible-invariant, it does not change with different viewing angle. Thus, we could establish a matching table from each pair of characteristic views. Accordingly, the homogeneous transformation could be obtained by solving that matching function. A such exercise for matching two partial views for a real object is described in Appendix C.

Irregular object matching

It should be explained that limitation of rigid-body probabilistic relaxation exists for determining the matching function of 3D object. Since the rigid-body uses the spatial distance as the matching constraint, the matching is selected when the score of matched distances among one view to another is maximal. However, if the distribution of features to be identified on the surface is a regular shape, e.g. sphere, the feature distances within themselves should remain the same. Thus, for this typical example, rigid-body relaxation matching will fail. However, the features are generally distributed randomly on the object surface. Hence, the rigid-body relaxation matching is generally applicable on 3D irregular object.

Further developments

Although we have developed the approach for reconstructing the global 3D object from the partial views of rangefinder, this is still necessary to solve some problems in order to improve the system.

1) The feature extraction for our development is used as the guideline for picking up and locating the exact coordinate manually. This process mainly depends on human being which does not have a general rule to govern. For example, the peaks would be sometimes the good features in highly irregular object. However, it would not be applied to some regular object, e.g. figure 5.4 (Doughnut). Therefore, it seems to be necessary to develop some algorithms of artificial intelligence, e.g. Neural Network, Genetic Algorithms, etc., to locate these features of different object generally.

2) It seems to be necessary to extract the features by means of differential geometry. However, our experiences were those features on range images may not be so rich to describe the characteristic views. As my opinion, feature extraction could be also based on the information which coming from the intensity images. The human vision system usually discriminates features by intensity. It would be a good trial for using features for both range and intensity images.

3) Rigid-body relaxation matching could be used to find the best matching for two different characteristic views. Since the 3D object reconstruction must be existing overlapping and non-overlapping portion, some surfaces may not be exposed to other characteristic views. Relaxation matching could not distinguish in between those features on overlapping and non-overlapping portion. Our approach is to discard the non-overlapping mapping function manually. However, it should be a more intelligent algorithm to select those non-overlapping features.

4) Finally, it could not be avoided for errors in extracting the feature locations, but which could be caused the error for solving the homogeneous transformation. Currently, our approach is to fine tune the homogenous transformation manually until the best matching is obtained. This error could be compensated by rotation and translation with a small quantity. It would be a good idea to find tune the transformation in an automatic process.

References

[1] P J Besl and R C Jain, "Invariant Surface Characteristics for 3D Object Recognition in Range Images," In *Proc. Computer Vision, Graphics, and Image Processing*, Vol. 33, 1986, pp 33-80.

[2] Ting-jun Fan, *Describing and Recognizing 3D Objects Using Surface Properties*, Springer-Verlag, 1990.

[3] P J Besl, *Surfaces in Range Image Understanding*, Springler-Verlag, 1988.

[4] P J Besl and R C Jain, "Intrinsic and Extrinsic Surface Characteristics", In *IEEE Transaction on Computers*, 1985, pp 226-233.

[5] D Zuk, F Pont, R Franklin and R Larrowe, *A System for Autonomous Land Navigation,* Technical Report IR-85-540, ERIM, Ann Arbor, MI, 1985.

[6] Perceptron, Product Information, Farminton Hills, Michigan. 1987.

[7] R Dunlay and D Morgenthaler, "Obstacle Detection and Avoidance from Range Data", In *Proc. SPIE Mobile Robots Conf.*, Cambridge, MA, 1986.

[8] E Krotkov and R Hoffman, "Results in Terrain Mapping for a Walking Planetary Rover", In *Proceedings of the Intl. Conf. on Advanced Robotics, Tokyo Japan*, 1993 pp 103-108.

[9] H R Everett, "Survey of Collision Avoidance and Ranging Sensors for Mobile Robots", In *Robotics and Autonumous Systems,* Vol. 5, 1989, pp. 5-67.

[10] P J Besl and R C Jain, "Range Image Understanding", In *IEEE Transaction on Computers*, 1985, pp 226-233.

[11] M Rioux and L Cournoyer, "The NRCC Three-dimensional Image Data Files", National Research Council, Canada, June 1988.

[12] M Rioux, F Blasis, J-A Beraldin and F Boulanger, "Range Imageing Sensors Development at NRC Laboratories", In *Proc. Workshop on Interpretation of Sences*, IEEE Computer Society Press 1989, pp. 154-160.

[13] P J Besl, J L C Sanz (ed.), "Active Optical Range Imaging Sensors", *Advances in Machine Vision,* Springler-Verlag in Perception Engineering, 1990, Chap. 1.

[14] E P Krotkov (ed.), Active Computer Vision by Cooperative Focus and Stereo, Springler-Verlag in Perception Engineering, 1989.

[15] R Watts, F Pont and D Zuk, "Characterization of the ERIM/ALV Sensor- Range and Reflectance", Technical Report, ERIM, Ann Arbor, MI, 1985.

[16] T Lee and S B Wong, "An Active Triangular Range Finder and Reflectance Sensing Using Scanning Mirror", In *Proc. of ICARCV'92,* Singapore, Vol. 1, CV5-6.p1-5, 1992.

[17] H T Tsui and D W Chan, "Lazer Range Finder with Intensity Imaging", In *Proc. ICARCV'90,* Singapore, pp. 1054-1058. 1990.

[18] M Hebert and E Krotkov, "3D Measurement from Imaging Laser Radars: How Good Are They?", *In Intl. Journal of Image and Vision Computing*, Vol. 10(3), pp. 170-178, April 1992.

[19] Cyberware Coorporation, 1993, *Production Information*, Monterey, CA.

[20] J Y Cartoux, J T Lapreste and M Richetin, "Face Autentication or Recognition by Profile Extraction from Range Images", In *Proc. Workshop on Interpretation of Sences*, IEEE Computer Society Press 1989, pp. 194-199.

[21] J J Rodríguez and J K Aggarwal, "Navigation Using Image Sequence Analysis and 3-D Terrain Matching", In *Proc. Workshop on*

Interpretation of Sences, IEEE Computer Society Press 1989, pp. 200-207.

[22] J A Jalkio, *The Analysis and Design of Triangulation Based Three Dimensional Vision Systems,* PhD Thesis, University of Minnesota, 1988.

[23] R K Miller, *3-D Computing, Modeling, Image Processing, and Visualization,* SEAI Technical Publications, 1992.

[24] J Ens, P Lawrence, "A Matrix Based Method for Determining Depth From Focus", in *Proc. of Computer Vision and Pattern Recognition,* IEEE Computer Society Press, 1991, pp. 600-606.

[25] Y C Kim, J K Aggarwal, "Finding Range from Stereo Images", in *Proc. of Computer Vision and Pattern Recognition,* IEEE Computer Society Press, 1985, pp.289-294.

[26] William Hoff, Narenda Ahuja, "Surfaces from Stereo : Integration Feature Matching, Disparity Esimation, and Contour Detection", in *IEEE Tran. on PAMI,* Vol. 11, No. 2, Feb., 1989, pp.121-136.

[27] Paul Dupuis, John Oliensis, "Direct Method for Reconstructing Shapes from Shading", In *Proc. of Computer Vision and Pattern Recognition,* IEEE Computer Society Press, 1992, pp 453-458.

[28] M Bicesel, A P Pentland, "A Simple Algorithm for Shape from Shading", In Proc. of Computer Vision and Pattern Recognition, IEEE Computer Society Press, 1992, pp.459-471.

[29] Marco Jamini, *A Rangefinder Based on The Rotation of a Virtual Image For Application in Three Dimensional Robotic Vision,* PhD Thesis, Case Western Reserve University, 1987.

[30] Jean-Yves Hervé, Yiannis Aloimonos, "Exploratary Active Vision : Theory", In *Proc. of Computer Vision and Pattern Recognition,* IEEE Computer Society Press, 1992, pp. 10-15.

[31] John S M Chiang, T Lee and S B Wong, "Rigid Body Relaxation Matching For 3D Object Recognition", In *Proc. of Robotics, Vision and Parallel Processing for Industrial Automation*, Universiti Sains Malaysia, 1994.

[32] James J Clark, "Active Photometric Stereo", In *Proc. of Computer Vision & Pattern Recognition*, IEEE Computer Society Press, June 1992, pp. 29-34.

[33] Steven K Case, Jeffrey A Jalkio, and Richard C Kim, Takeo Kanade (ed.), "3-D Vision System Analysis and Design", *Three Dimensional Machine Vision*, Kluwer Academic Publishers, 1987.

[34] J A Jalkio, R C Kim and S K Case, "Triangulation Based Range Sensor Design", in *Proc. of SPIE*, IEEE SPIE press, 1987, pp. 243-248.

[35] General Scanning Inc., *XY Scan Head Series: X-Y Scan Head*, Watertown, Massachusetts.

[36] A Yariv, *Introduction to Opical Electronics*, New York : Holt, Rinehart and Winston, 2nd Ed.

[37] Kong Shao-hua, *3-D Irregular-Object Recognition System*, MPhil Thesis, The Chinese Universtity of Hong Kong, Department of Electronic Engineering, August 1992.

[38] Curtis F Gerald and Patrick O Wheatley, *Applied Numerical Analysis*, Addison-Welsey Publishing Company, 3rd Edition, June 1985.

[39] T Lee, R Li and S K Kong, , *Recognising Partially Occluded Objects by Rigid Body Relaxation Matching*, Electronics Letters, 21st June 1990 Vol. 26 No. 13, pp. 856-857.

[40] William H Press, *Numerical Reciepes In C*, 2nd Edition, Cambridge University Press, 1992.

[41] A Rosenfeld, "Scen Labeling by Relaxation Operations", in *IEEE Trans. Systems, Man and Cybernetics*, vol. SMC-6, 1976, pp. 420-433.

[42] ACM, *Collected Algorithms from ACM*, Association for Computing Machinery, Inc., New York, 1980.

[43] Yeon C Kim and J K Aggarwal, "Finding Range From Stereo Images", In *Proc. of Computer Vision and Pattern Recognition*, IEEE Computer Society Press, 1985, pp 289-294.

[44] R C Bolles and Patrice Horaud, Tekeo Kanade (ed.), "3DPO : A Three-Dimensional Part Orientation Systems", In *Three-Dimensional Machine Vision*, Kluwer Academic Publishers 1987, pp.399-450.

[45] O Monga, S Benayoun and O D Faugeras, "From Partial Derivatives of 3D Density Images to Ridge Lines", In *Proc. of Computer Vision and Pattern Recognition*, IEEE Computer Society Press, June 1992, pp. 354-359.

[46] Silicon Graphics Computer System, *Graphics Library Programming Guide*, Document No 007-1210-040, Silicon, Graphics, Inc., Mountain View, California.

[47] Yang Chen and Gerad Medioni, "Object Modelling by Registration of Multiple Range Images", *In Intl. Journal of Image and Vision Computing*, Vol. 10(3), pp. 145-155, April 1992, Butterworth-Heinmann Ltd..

Appendix A Projection of Object

In chapter 4, we have discussed that the object height ***h*** could be found by measuring the displacement of the line ***d*** in figure A.1. Next, we are going to find the relation between the object height, ***h*** in terms of with ***d***, the displacement in the image plane of the camera.

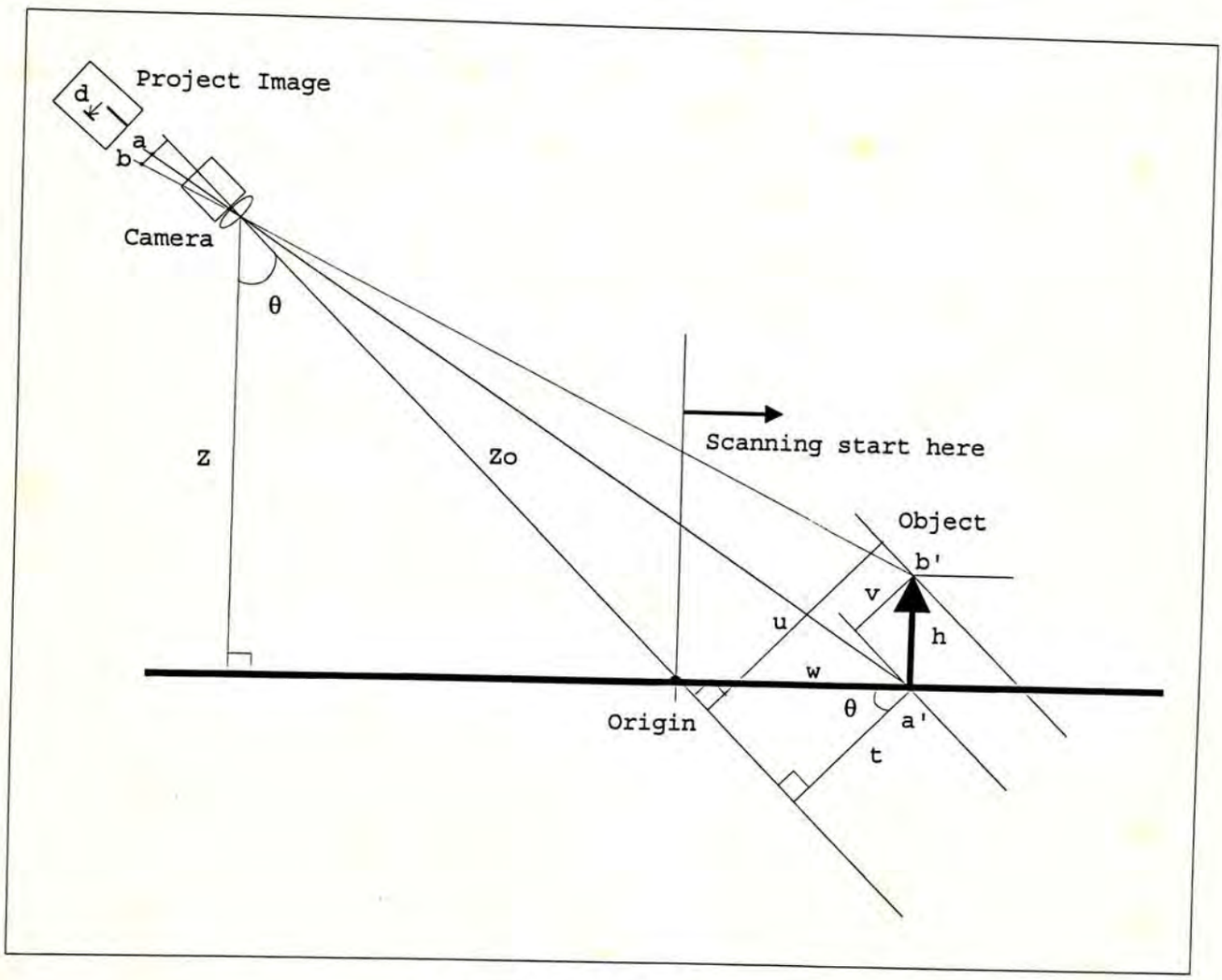


Figure A.1 Analytical Diagram

As shown in figure A.1, the camera is placed at a height ***Z*** with an angle θ. An object of height ***h*** is located at ***w*** apart from origin. The project image will be formed by ***a*** & ***b***. Thus, we can express :-

$$d = b - a$$

Equation A.1

We can find

$$a = \frac{f}{Z_0 + w\sin\theta} w\cos\theta$$

Equation A.2

where Z_0 = distance from origin to lens; f = focal length;

w = distance from origin to a'; θ = Angle of incidence.

And,

$$b = \frac{f}{Z_0 + w\sin\theta - h\cos\theta}(h\sin\theta + w\cos\theta)$$

Equation A.3

Finally, by substituting equation A.2 & A.3 to A.1, we can obtain

$$d = \frac{f(h\sin\theta + w\cos\theta)}{Z_0 + w\sin\theta - h\cos\theta} - \frac{fw\cos\theta}{Z_0 + w\sin\theta}$$

Equation A.4

And, Z_o can be calculated

$$Z_0 = \frac{Z}{\cos\theta}$$

Equation A.5

Thus, we can obtain the analytical relationship between ***d*** and object height ***h***.

Appendix B Performance Analysis on Rangefinder System

In order to evaluate actual performance of the rangefinder system, several experiments were carried including photometric distortion, range tolerance, mirror controller error.

B.1 Photometric Distortion

As discussed previously, the resolution of captured image is not homogeneously distributed on the screen. Since camera is viewing with an angle to the object, the image at closer distance is enlarged whereas the image at farther object is shrunk.

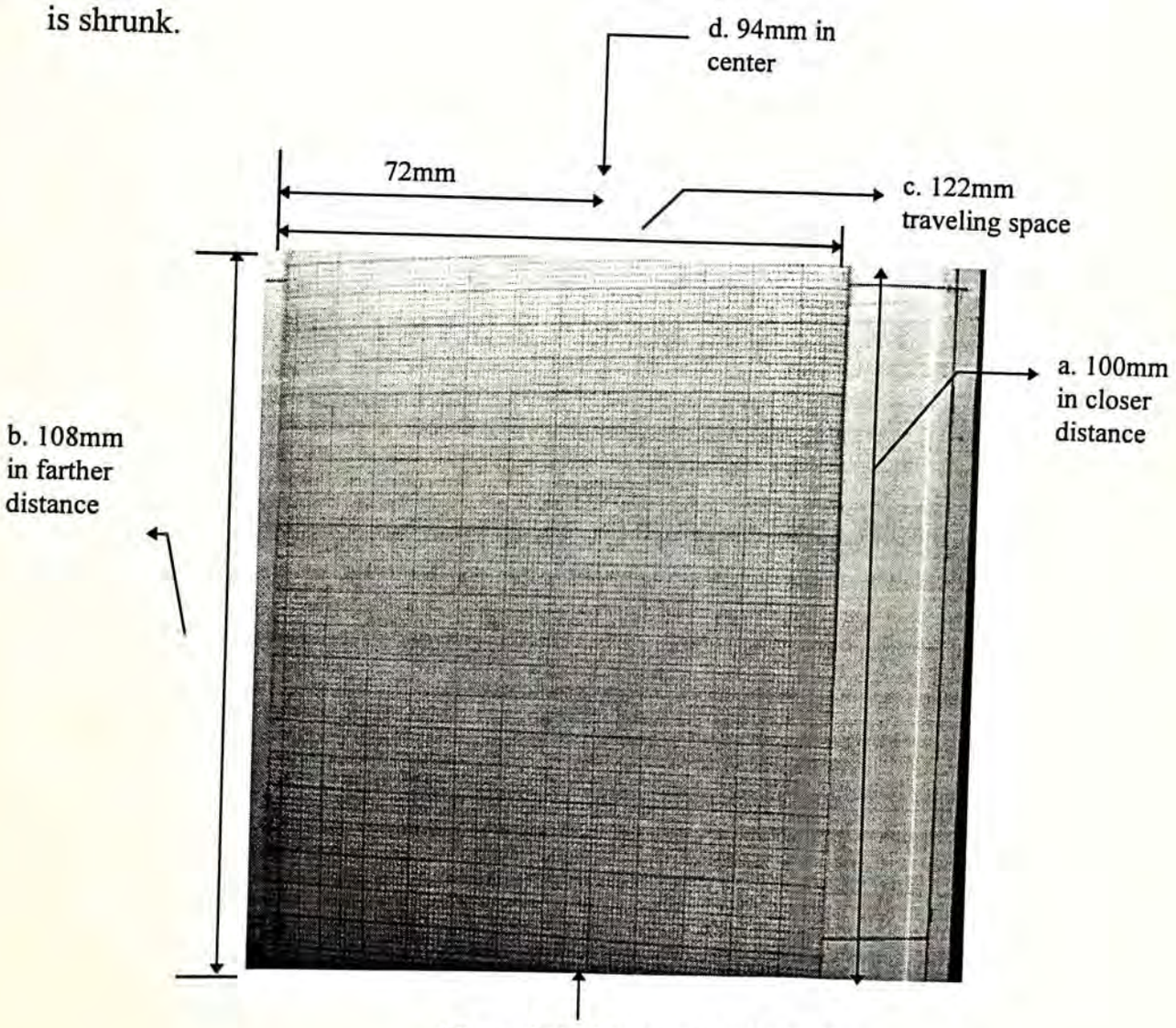


Figure B.1 - Photometric Distortion

Figure B.1 shows a snapshot from the camera of the rangefinder. The camera looks from the right hand side to the left hand side with an angle. We could measure width on the closer side (a) is 100mm while the farther side (b) is 108mm with the a traveling distance 112mm (c).

Taking with reference (d) as the center, we would like to calculate the magnification factor in the transversal axis.

Distance (mm)	Width (mm)
0	99
10	99.5
20	100
30	101
40	101.5
50	102
60	103
70	104
80	104.5
90	105
100	106
110	107
120	108

Table B.1 - Photometric Magnification

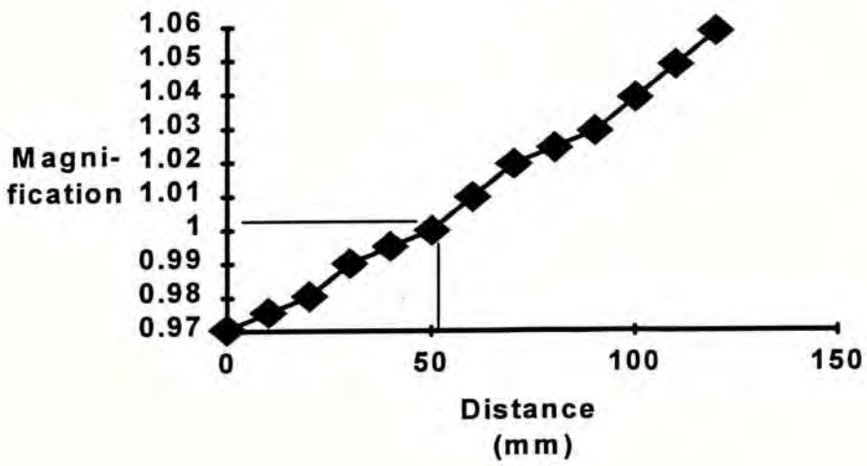


Figure B.2 - Magnification factor against distance

According to figure B.2, the magnification factor could be range from 0.98 to 1.06. Therefore, if the object is scanned over the surface, the portion which close to camera would be magnified whereas farther portion would be shrunk.

B.2 Range resolution

The next experiments were used to illustrate the accuracy of the rangefinder. At first, we would like to use several reference samples with known heights to give a overall transfer function. The experiment setup is shown on figure B.3.

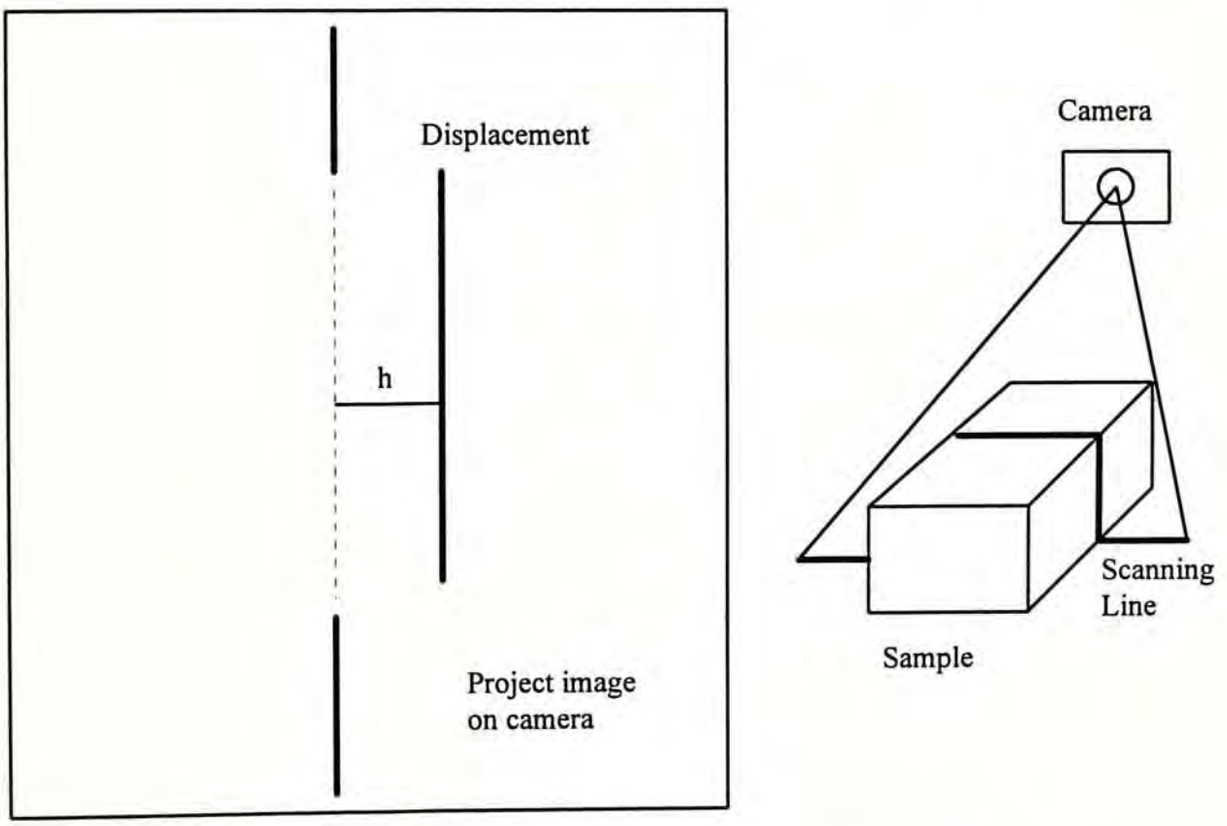


Figure B.3 - Experimental setup for range resolution.

The reference samples were simple rectangular metal pieces. They were made by the NC milling machine and hardened by heat treatment which could ensure the flatness of the surface within 2 microns. The sizes of sample is 4cm X 8cm with different heights, which were 2cm, 3cm, 5cm, 6cm, 7cm respectively.

The experiments were done to measure profiles ***h*** in figure B.3 along the projected line against different heights of samples. The results are plotted on figure B.4-9 as follows.

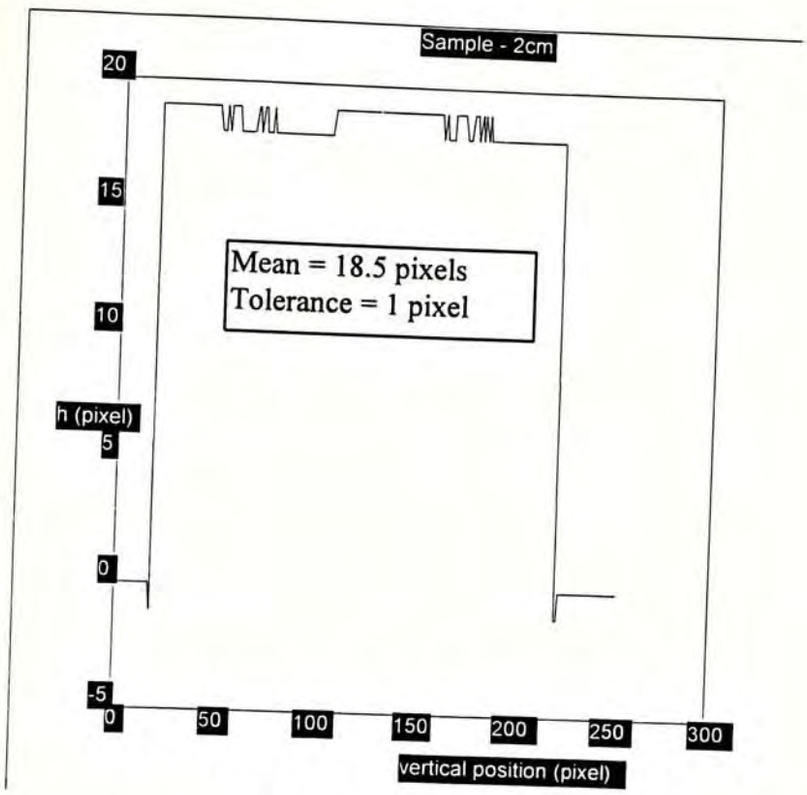


Figure B.4 - Surface Profile (2cm)

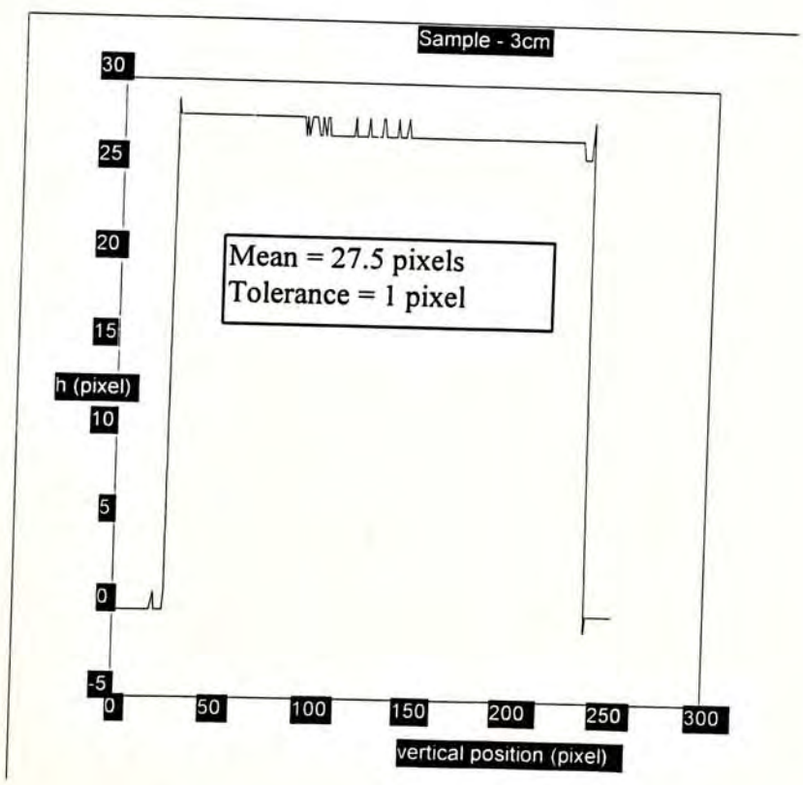


Figure B.5 - Surface Profile (3cm)

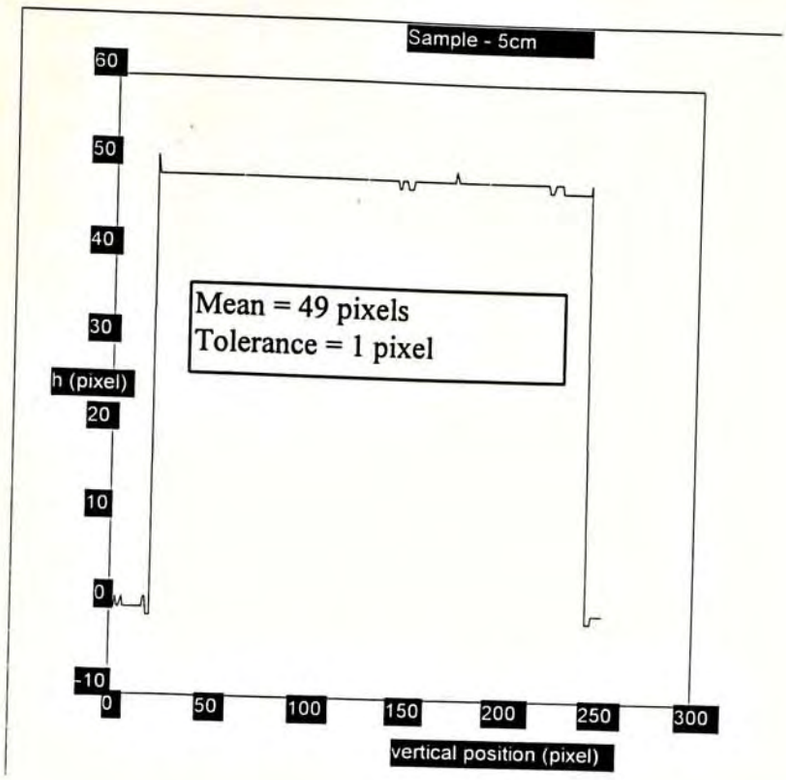


Figure B.6 - Surface Profile (5cm)

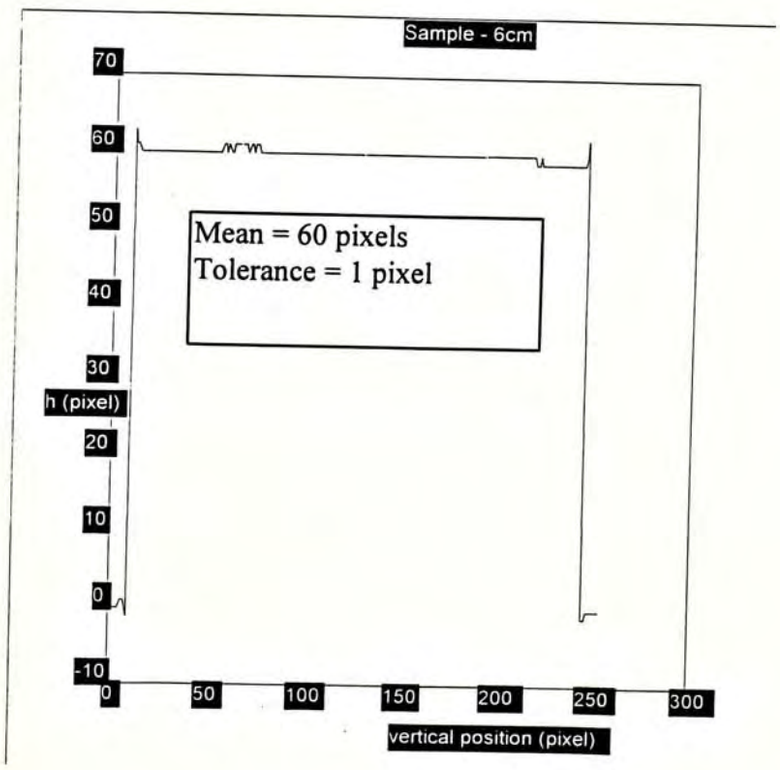


Figure B.7 - Surface Profile (6cm)

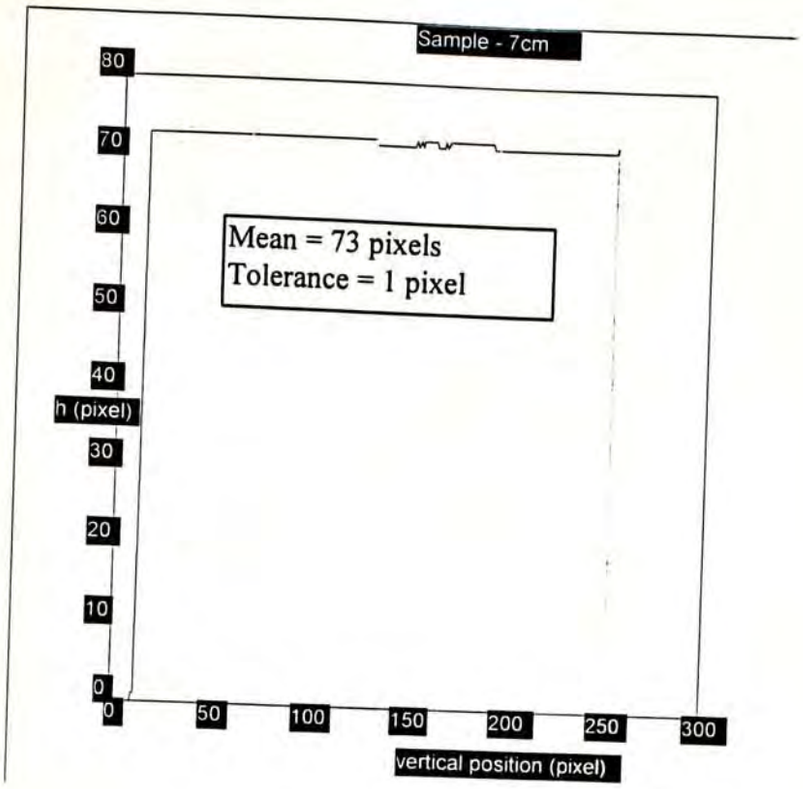


Figure B.8 - Surface Profile (7cm)

In order to analyze the range accuracy for our rangefinder system, we must first convert the pixel unit on the camera to physical dimension (mm). Therefore, we tabulate the average values of ***h*** against the sample height as in table B.2.

h (pixel)	Physical Height (mm)
18.5	20
27.5	30
49	50
60	60
73	70

Table B.2 - h in pixels vs. Physical Height (mm)

Then, we plot the graph of physical height against vertical position :-

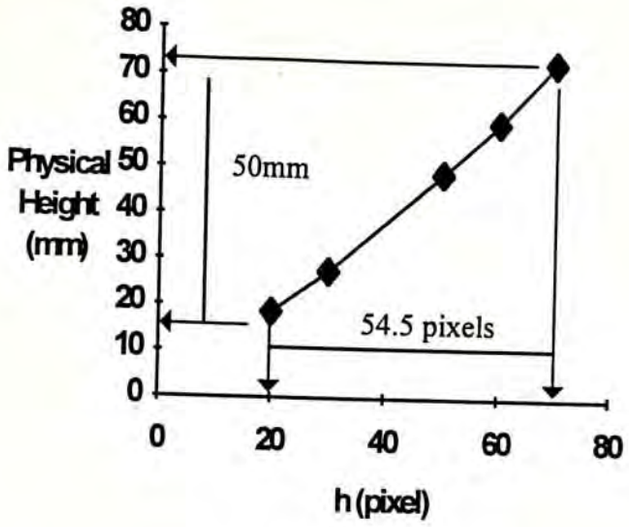


Figure B.9 - Plot *h* against Physical Height

We would like to use a linear relation to represent the above graph :-

$$M = \frac{\text{Physical Height}}{\text{Pixels on Camera}} = \frac{50\text{ mm}}{54.5\text{ pixels}} = 0.917\text{ mm / pixels}$$

According to figure B.4-9, the measurement accuracy of the camera is about ±1 pixel. Hence, range accuracy of the rangefinder is :-

$$\boxed{\text{Tolerance} = \text{M}x(1\text{ pixel}) = \pm\ 0.917\text{mm}}$$

B.3 Mirror Controller Linearity

Next, we would like to examine the performance of the mirror controller. An experiment in figure B.10 was used to detect the laser beam position against the variation of mirror deflection angle. The deflection angle is adjustable by setting the 8-bit D/A converter in mirror controller. Therefore, we measured the beam position while varying the D/A setting. The results were recorded and plotted in figure B.11.

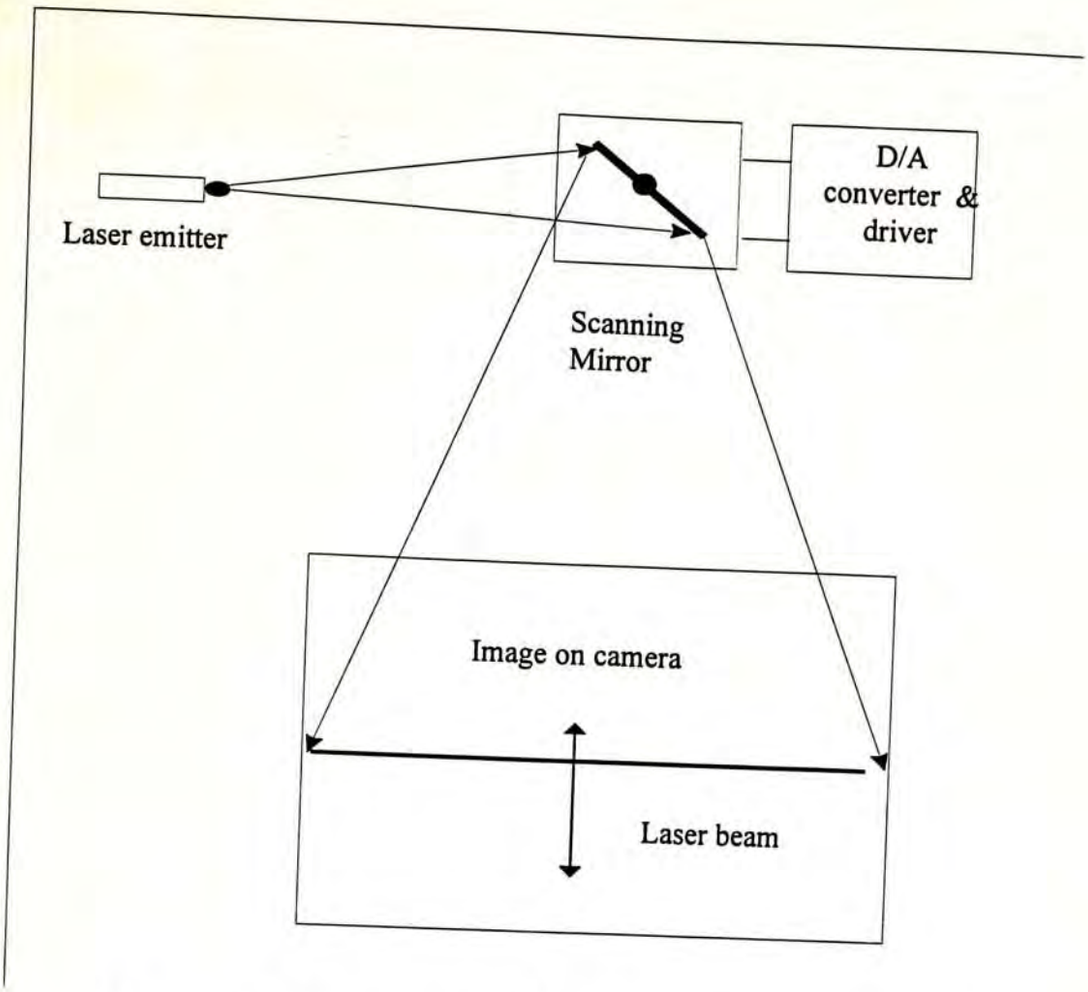


Figure B.10 - Experimental Setup for Mirror Controller Linearity

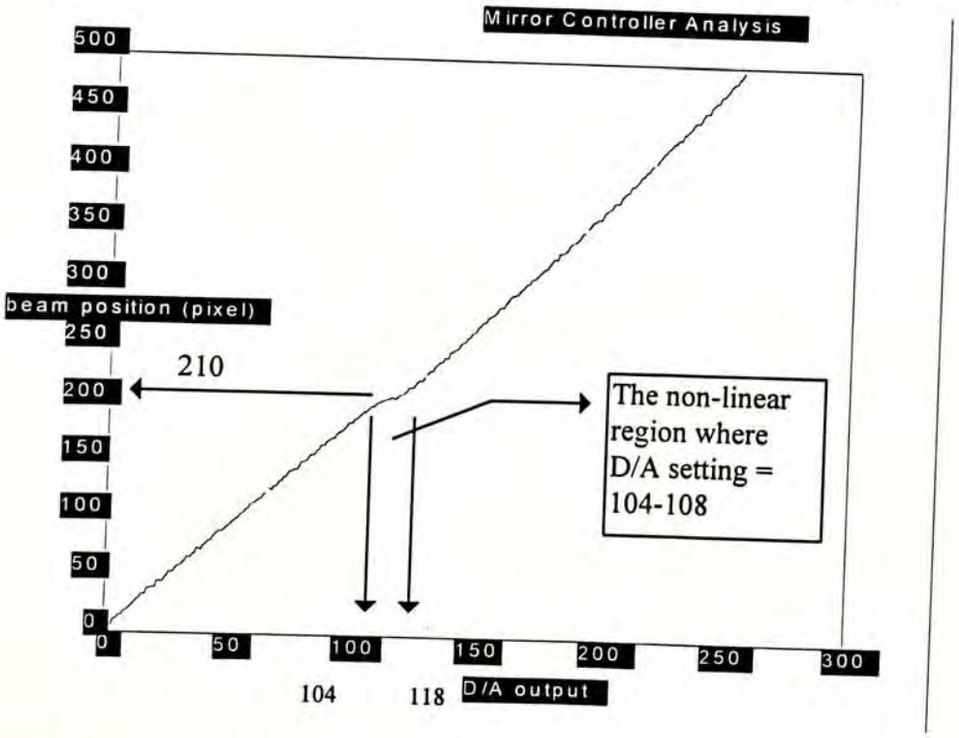


Figure B.11 - Beam Position vs. D/A Setting

According to figure B.11, we could see the laser beam position could be controlled by setting the value of D/A converter with a linear relationship except the

region of (104-118). The beam position in that region could only be adjustable with a small amount when comparing with others.

In order to explain the non-linear respond in the region (104-118), we must investigate the output driving circuit of scanning mirror. Our design chose the high current darlington transistor pairs as class B DC amplifier. It would suffer the problem of non-linear output on the zero-crossing region. Although we putted a DC bias to prevent this problem, the zero-crossing region still exists.

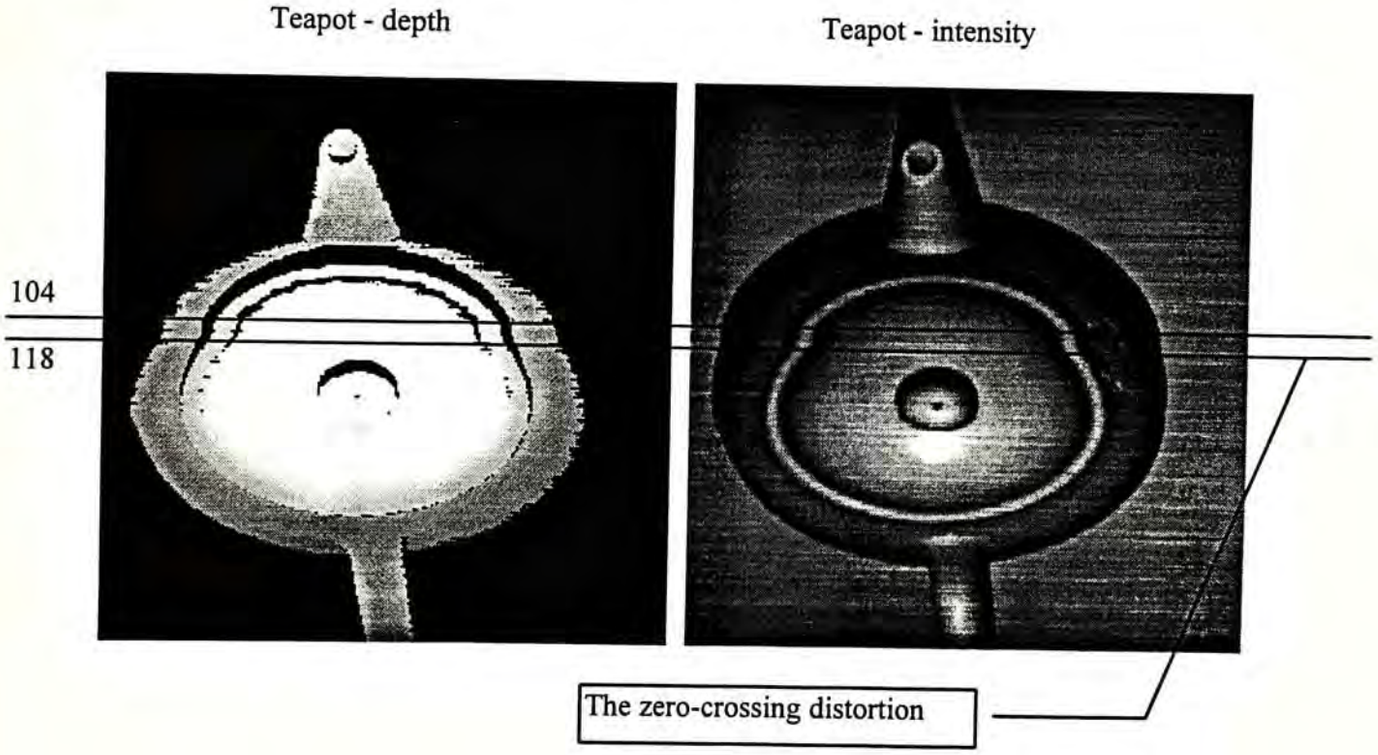


Figure B.12 - The zero-crossing distortion

The zero-crossing region would make a very small increment in the scanning. As indicated in figure B.12, the zero-crossing effect on the scanning controller could create a distortion on the output image.

Next, we would like to examine the transfer function for laser beam versus the D/A setting. Table B.3 tabulates the physical position versus the D/A values.

D/A Values	Physical Position (mm)
0	0
31	20
48	31
64	40
80	49
96	58
112	63
128	69
144	79
160	88
176	97
192	106
208	116

Table B.3 - Physical Position for different D/A values

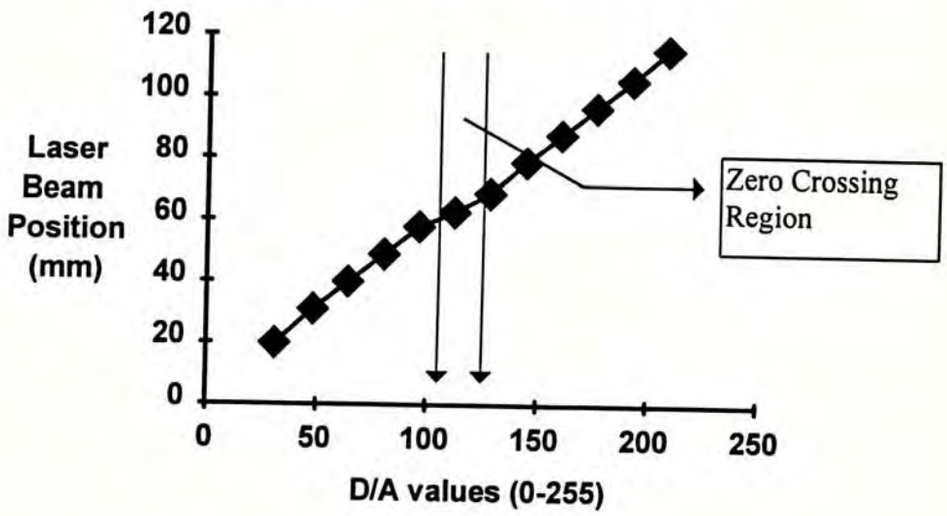


Figure B.13 - Transfer Function of Laser Position

B.4 Repeatability

Several experiments were carried on to test the repeatability of the rangefinder. A sample, as shows in figure B.14, was used to scan with 5 sets of data. and remained the same position.

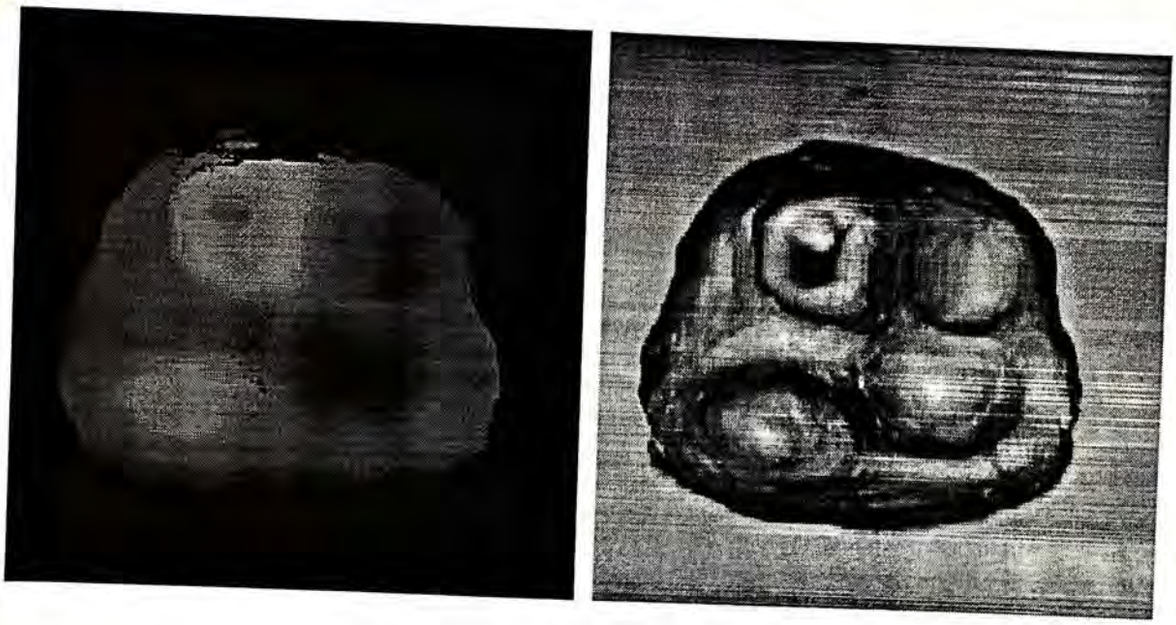


Figure B.14 - Sample Object

The repeatability test would examine the number of pixels in difference, max. difference and average difference.

ND = Total number of pixels in different value between two images

$$MD = \text{maximum difference} = Max\left[Abs(I_1 - I_2)\right]$$

$$AD = \text{average difference} = \frac{\sum Abs(I_1 - I_2)}{ND}$$

where I_1, I_2 = two different images under test. By taking one reference image (Obj1-1), we would compare the other four sets of range data :-

Image File	ND (pixel)	MD (pixel)	AD (pixel)
Obj1-2	35824	36	0.789
Obj1-3	35823	36	0.703
Obj1-4	35825	35	0.770
Obj1-5	35834	35	0.696

Table B.4 - Repeatability Test

There would be about **54.7%** along the images are different to each other. By converting the values to physical unit, the maximum difference is **35mm**. Also, the average difference is **0.72mm**.

B.5 Effect of Surface Flatness

Finally, different samples of surface flatness were used to examine the effect of range measurement. The samples were polished by different sizes of metallic sand, i.e. no. 60, 80, 120 (greater sand no. means smaller particles).

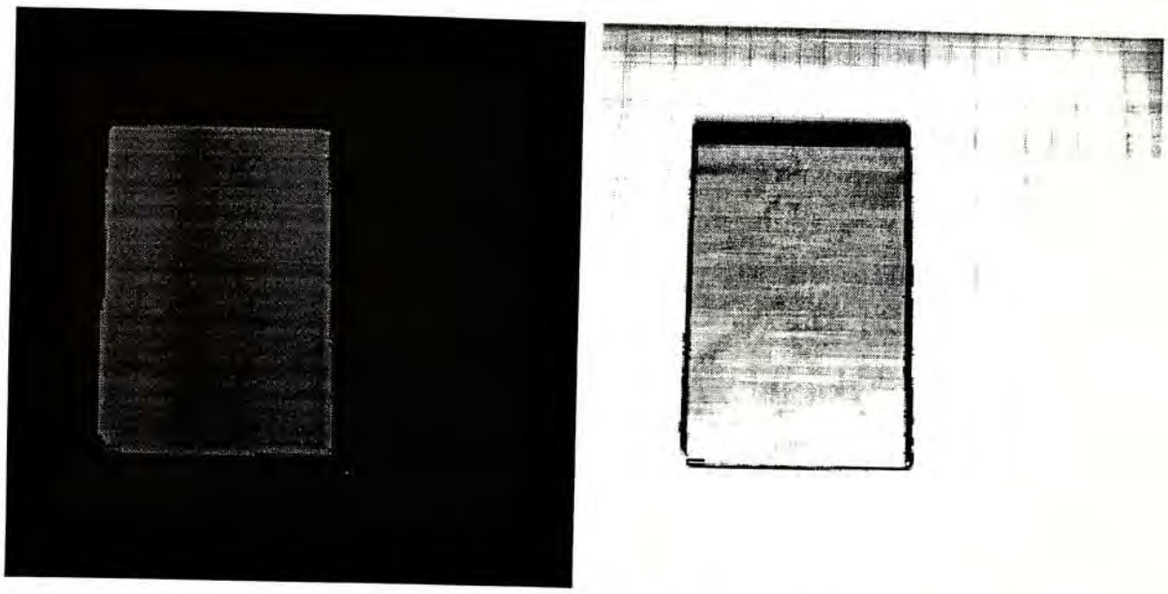


Figure B.15 - Sample 2cm (without polishing)

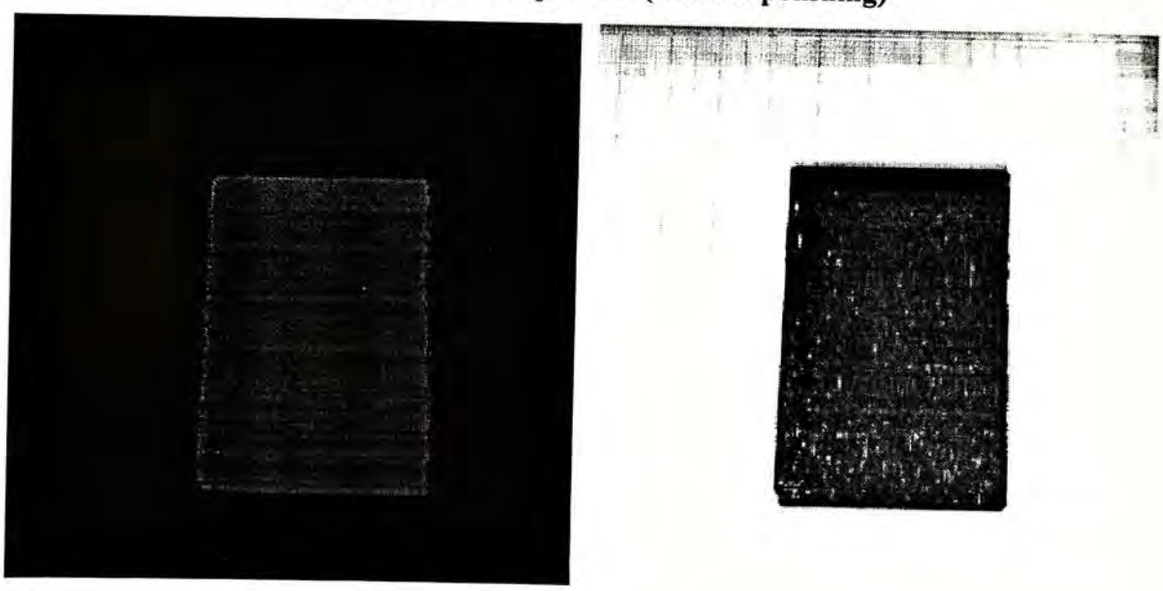


Figure B.16 - Sample 2cm (with no. 120 polishing)

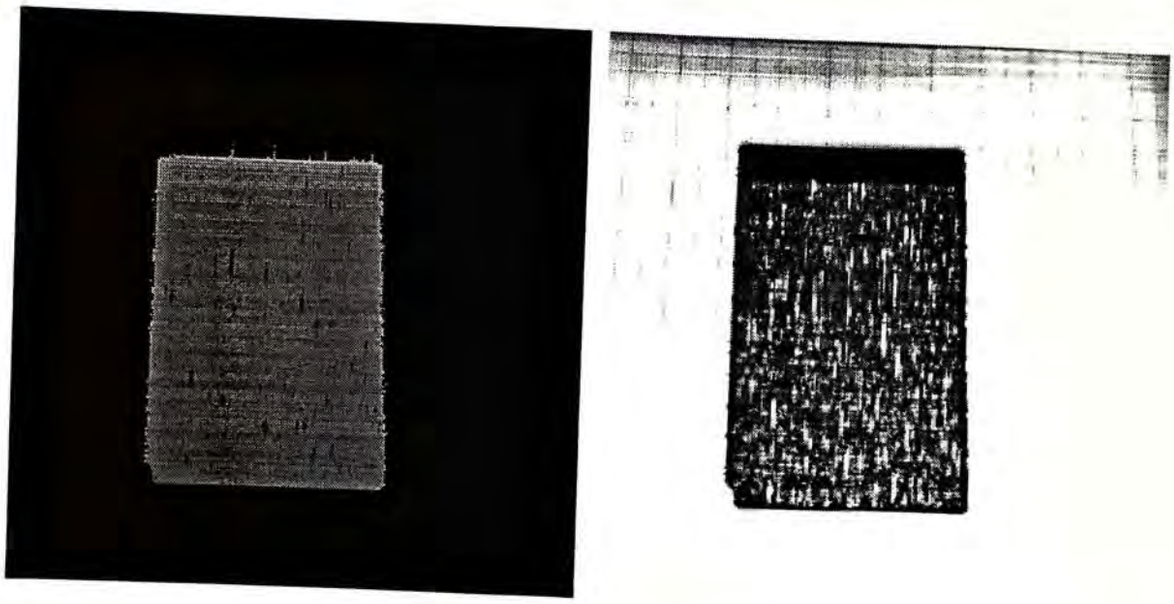


Figure B.17 - Sample 3cm (with no. 80 polishing)

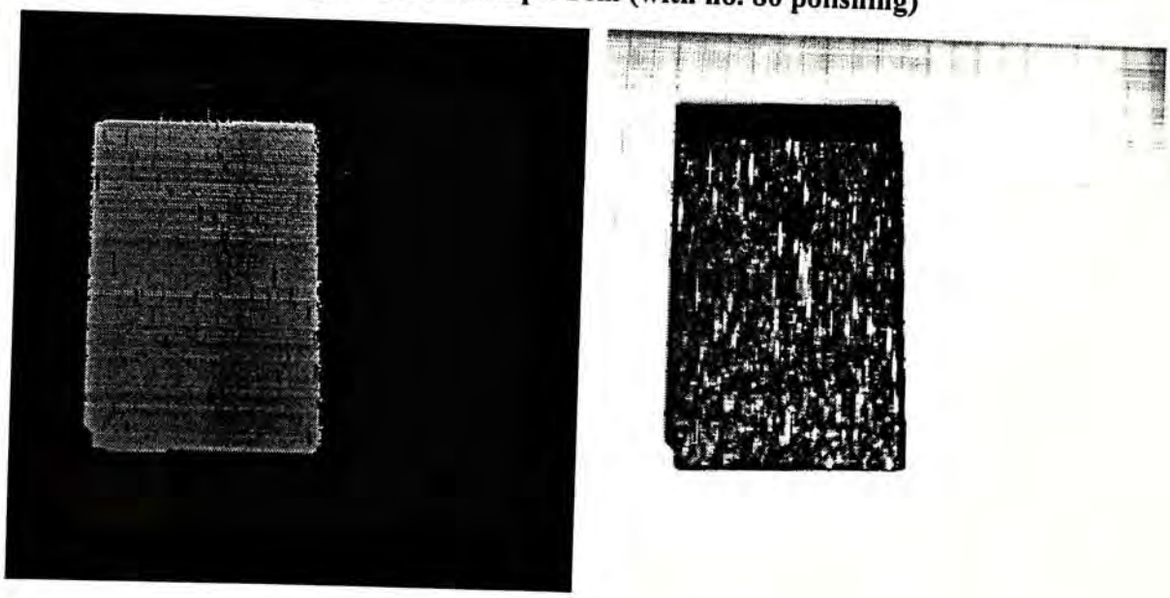


Figure B.18 - Sample 3cm (with no. 60 polishing)

According to figure B.15-18, we could see roughly the noise pattern is proportional to the surface flatness. Therefore, there would be much irregularities if the object surface is rough. In order to examine the phenomenon more details, we would like to extract a profile for the different surface flatness which are indicated in figure B.19-21 respectively.

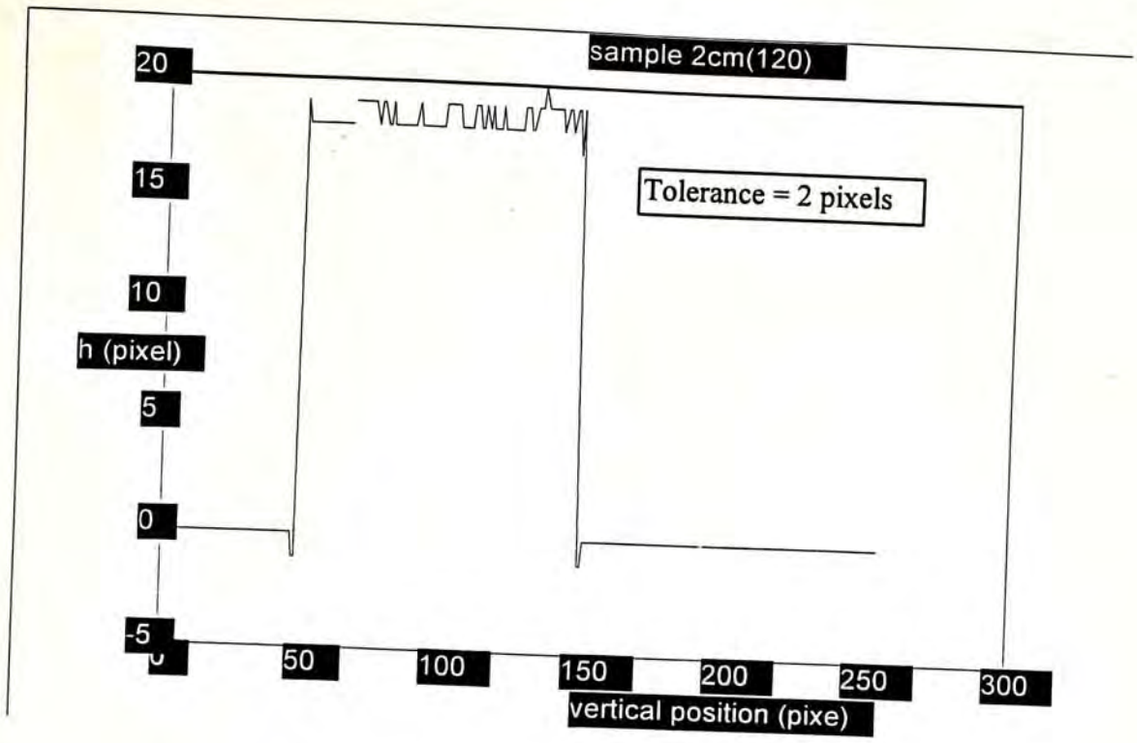


Figure B.19 - Surface profile 2cm (with no.120 polishing)

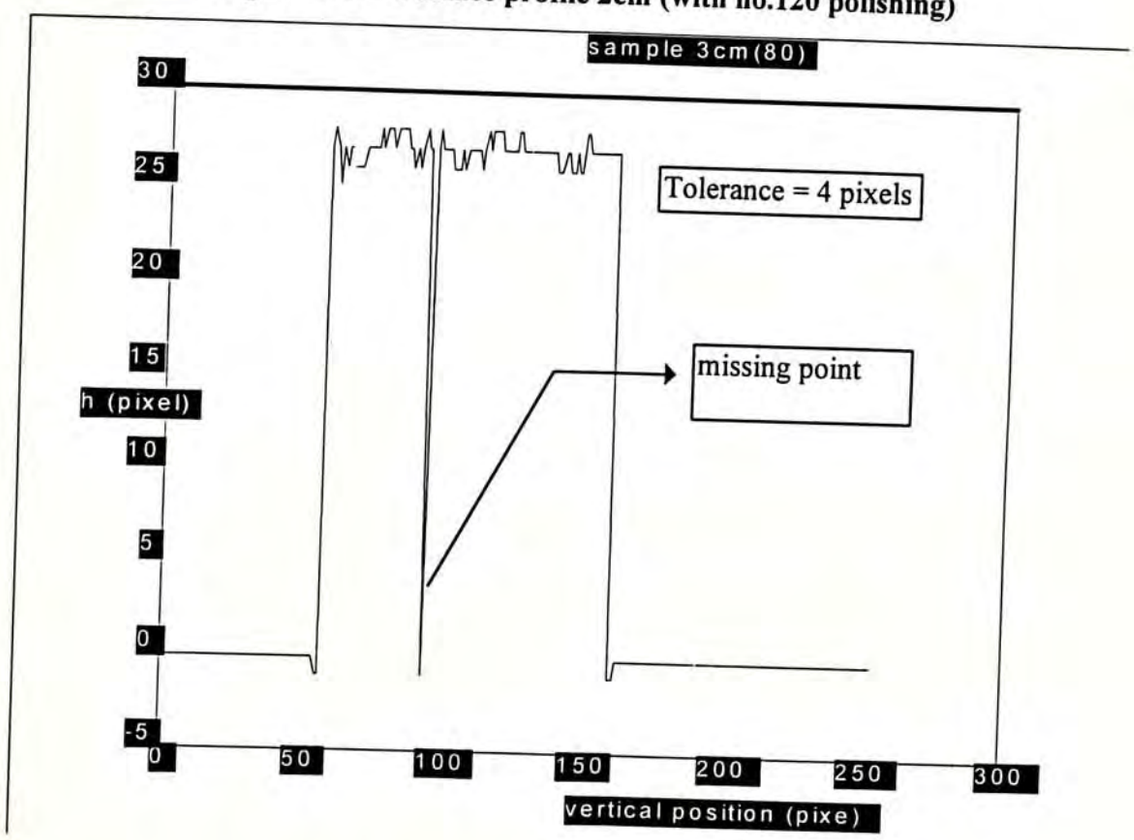


Figure B.20 - Surface profile 3cm (with no. 80 polishing)

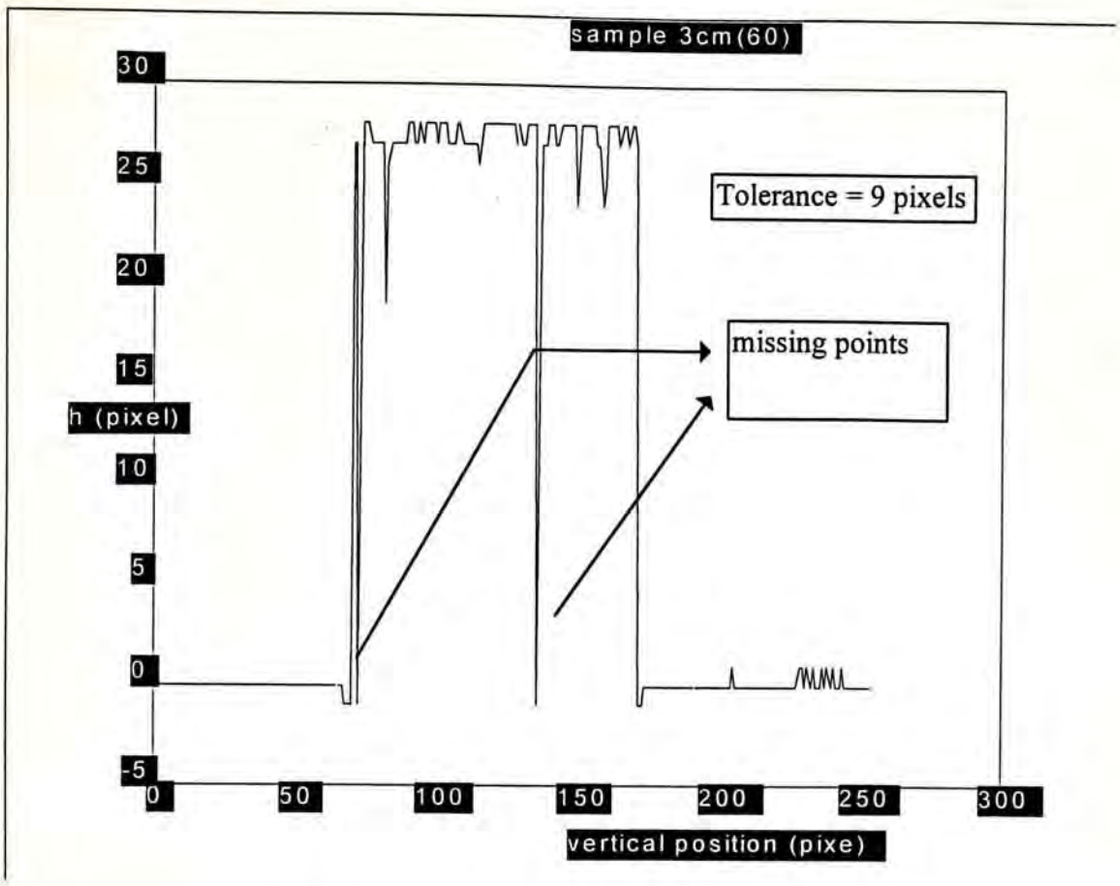


Figure B.21 - Surface Profile 3cm (with no. 60 polishing)

According to figures B.19-21, the range accuracy would be about 1.834mm for surface with metallic sand no. 120 polishing, 3.668mm with sand no. 80 polishing, and 8.253mm with sand no. 60. Conclusively, the range accuracy is low when the surface is rough. Also, a large amount of "missing points" is happened when the surface is rough enough.

B.6 Summary

According to the above experiments, we would like to make a summary on them. At first, the rangefinder would enlarge the object in closer distance whereas shrink the object in farther distance. The factor is range from 0.97 to 1.06 for -50mm and +70mm displacement respectively. Secondly, the range resolution is dependent on the camera's resolution which is within 1 pixel. Based on the experiments on the

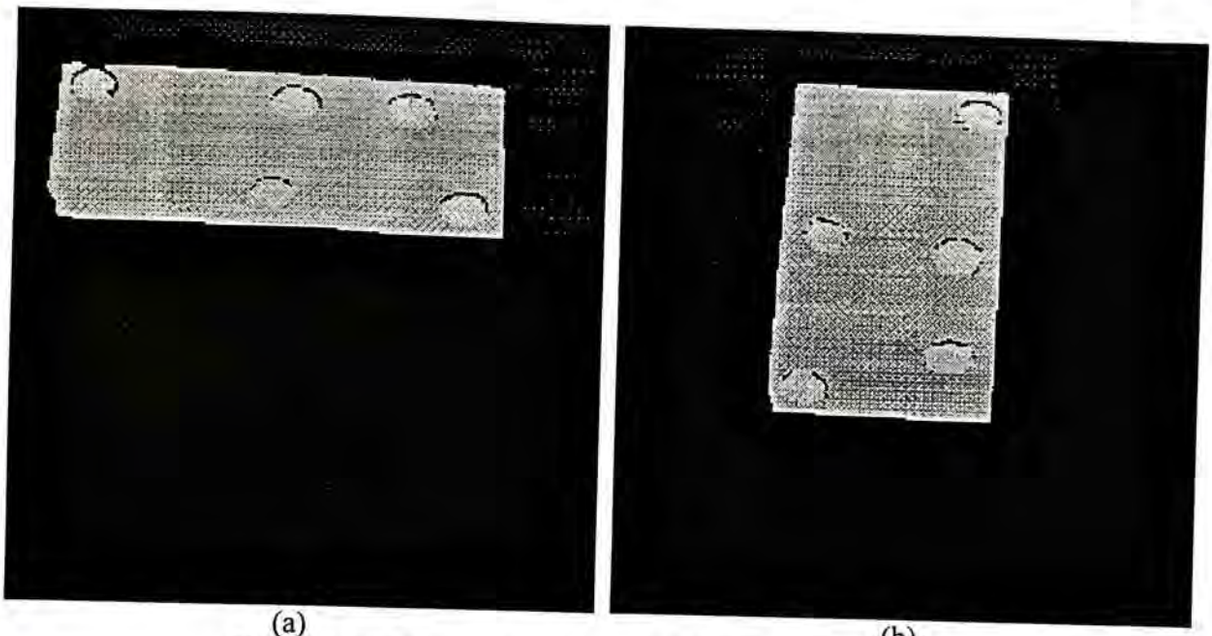
session 2, 1 pixel tolerance is representing 0.917mm range. Thus, accuracy of the rangefinder is 0.917mm. Thirdly, the mirror controller response is not linear which suffers a zero-crossing distortion on longitudinal position 210mm. Finally, the repeatability of the rangefinder 0.72mm (mean).

We would like to use a table to summarize on the above parameters :-

Parameter	Value
Photometric Magnification	0.97@0mm; 1@50mm; 1.06@120mm (see figure B.2 for more details)
Range Accuracy	0.917mm
Camera Resolution	0.917mm / pixel
Mirror Controller Linearity	1.706 mm / D/A unit; @D/A 0-103 0.815 mm / D/A unit; @D/A 104-108 1.736 mm / D/A unit; @D/A 109-255 (see figure 13 for more details)
Mean Repeatability	0.72mm

Appendix C Matching of Two Characteristic Views

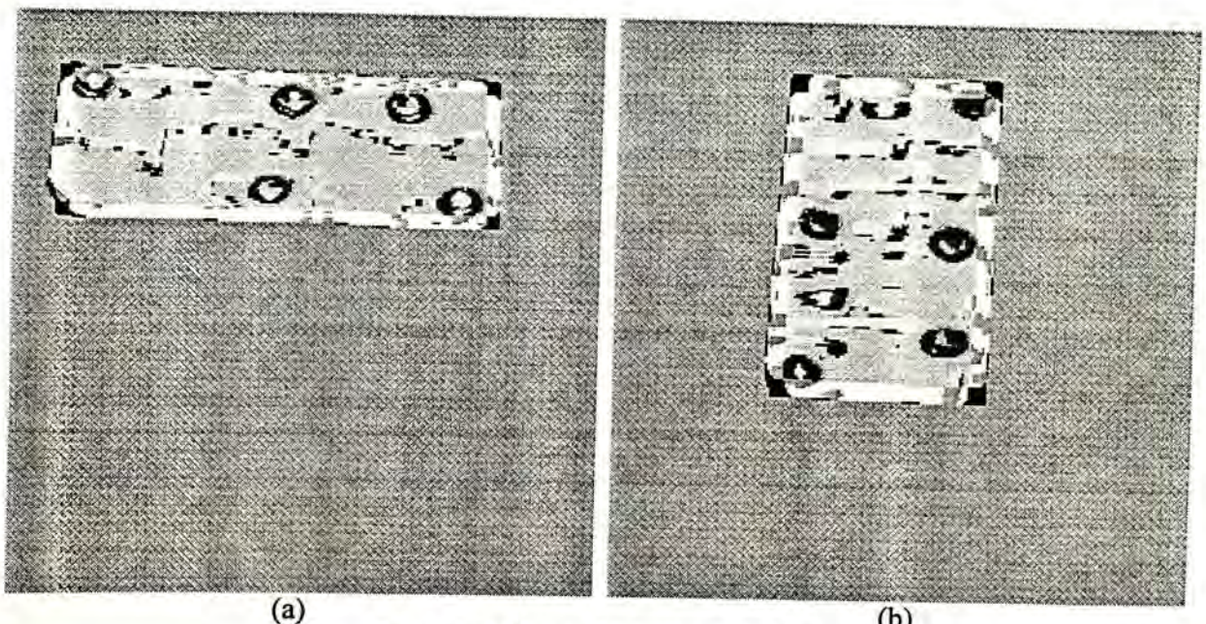
Two characteristic views of an object has been obtained and depicted in Fig. C.1. The views were corrected for geometric distortion caused by the scanning mechanism of the developed system. The correction ratio was determined from Table B.1 and Fig. B.2.



(a) (b)

Fig. C.1 (a) Characteristic View 1 (b) Characteristic View 2

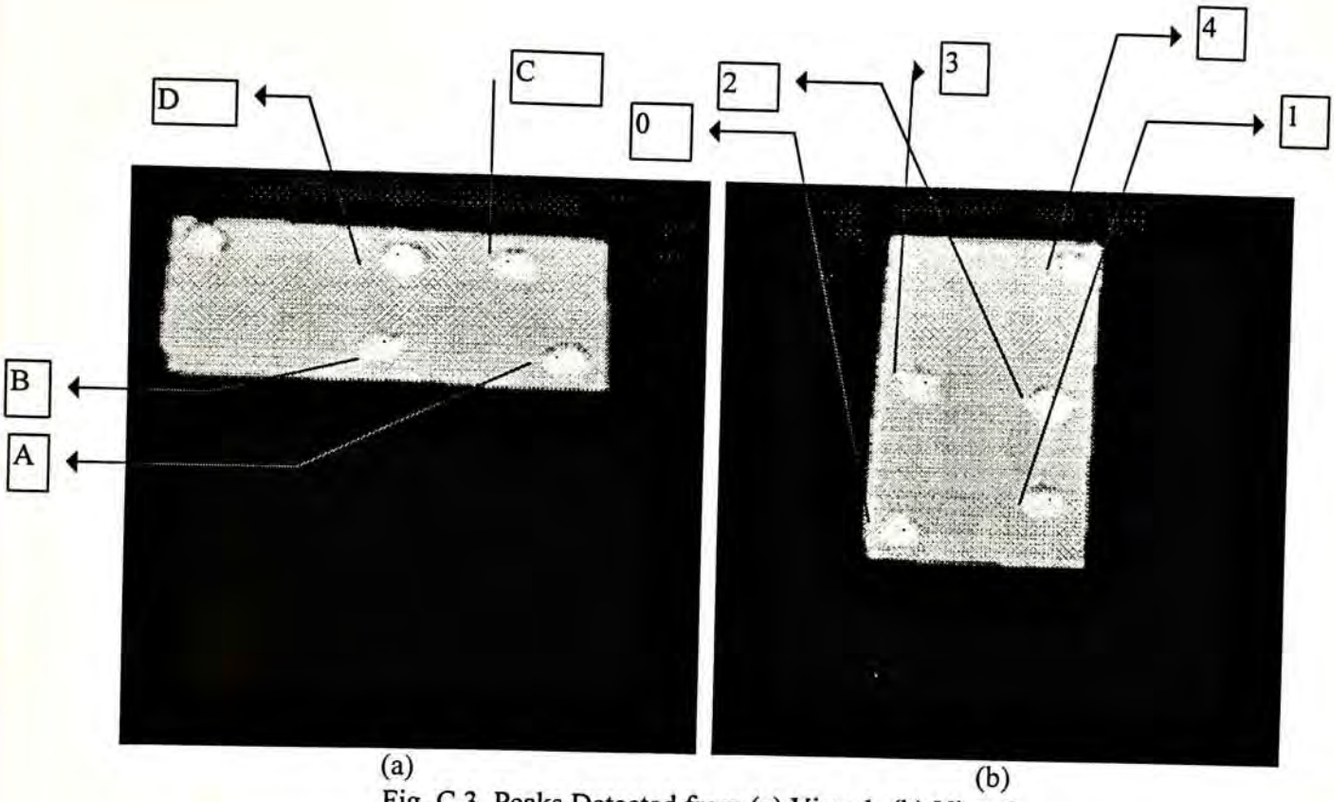
From the range maps, Guassian and mean curvatures have been computed using Equations 6.3 and 6.4. The maps were further labelled according to the signs of the curvatures as depicted in Table 6.1. The results were depicted in Fig. C.2.



(a) (b)

Fig. C.2 Label Maps of (a) View 1 (b) View 2

With the label maps from Fig. C.2, "peak" feature points were extracted using the peak-extraction algorithm depicted in Fig. C.3. Coordinates of the feature points so obtained were tabulated in Table C.1 and C.2. Hence a sub-rigid body model was formed from each CV. Next, the relaxation process was applied to find the mapping of features between two views. The mapping obtained was depicted in Table C.3.



(a) (b)

Fig. C.3 Peaks Detected from (a) View 1 (b) View 2

The Peak Detection Algorithm

Step 1: Let N be the template window size = 31

Step 2: IF the centre pixel of the window is of PEAK label THEN

The centre pixel is a PEAK if it has the maximum range value in the window.

Step 3: Move the window to next pixel and go to Step 2.

Step 4: Stop when the whole image has been checked

Fig. C.3 The Peak Extraction Algorithm

Feature	x	y	z
A	0.74	0.68	0.56
B	0.44	0.70	0.53
C	0.65	0.85	0.59
D	0.46	0.86	0.60

Table C.1 - Features Extracted from View 1.

Feature	x	y	z
0	0.33	0.40	0.48
1	0.58	0.46	0.48
2	0.58	0.64	0.56
3	0.36	0.66	0.51
4	0.59	0.86	0.56

Table C.2 - Features Extracted from View 2.

Table C.3 Feature Mapping obtained by the Relaxation Process

View 1	View 2
A	0
B	3
C	1
D	2

By inspection, the matching successfully found the feature correspondence. However, this exercise merely illustrated the processes proposed in Chapter 7 for matching real data. The whole system has to be further evaluated using objects with more features, preferably of different feature types and with multiple views. In particular, the feature extraction process has to be carefully evaluated before the system can become a practical one.