uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Christopher S. Yang
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.A.Sc. (Electrical Engineering)
GRADE / DEGREE

School of Information Technology and Engineering
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Design and Calibration of a Multi-modal Range Sensor using Passive Stereo, Structured Lighting, and Active Triangulation Laser Range Finder

TITRE DE LA THÈSE / TITLE OF THESIS

P. Payeur
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

R. Laganière

M. Frize

Gary W. Slater
LE DOYEN DE LA FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTDOCTORALES /
DEAN OF THE FACULTY OF GRADUATE AND POSTDOCORAL STUDIES

DESIGN AND CALIBRATION OF A MULTI-MODAL RANGE SENSOR USING PASSIVE STEREO, STRUCTURED LIGHTING, AND ACTIVE TRIANGULATION LASER RANGE FINDER

by

Christopher S. Yang

A thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements of the degree of
Master of Applied Sciences
in Electrical and Computer Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering
School of Information Technology and Engineering
Faculty of Engineering
University of Ottawa

Your file Votre référence
ISBN: 0-494-14975-2
Our file Notre référence
ISBN: 0-494-14975-2

Abstract

Collecting dense range measurements in uncontrolled environments is a challenging problem as lighting and surface textures significantly influence the quality of the measurements. This dependency affects the registration and data fusion processes and consequently degrades the accuracy of surface or occupancy models that are computed from the range measurements.

Typical approaches to address this issue have concentrated on improving a specific type of range sensor. On the other hand, the overall quality of the sensing can also be enhanced through the development of a mechanism that combines various range sensing technologies to form a multi-modal range sensor. The resulting problem of merging datasets collected with the respective modalities can then be solved in two ways: system calibration of the multi-modal sensor or data fitting of all datasets into a single model, the latter being more widely used. The lack of multi-modal system calibration approaches is due to their complicated and lengthy nature, where individual calibration approaches must be applied to each subsystem and then applied between subsystems of the multi-modal range sensor.

This thesis proposes a technique to alleviate the problems encountered in multi-modal systems calibration. Straightforward and generic guidelines for calibration are defined and applied to an in-house integrated multi-modal system built from a laser range finder, two structured lighting systems, and a stereovision system. The system's intra- and inter-calibration processes are detailed. Reconstructed renderings of datasets collected with the calibrated multi-modal range sensor, without the use of data fitting, are also presented. From these results, the potential benefits of multi-modal calibration that reduces the need for data fitting and the advantages of merging subsystem's strengths to complement other subsystem's weaknesses are put into evidence.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to acknowledge Dr. Pierre Payeur for his unwavering support over the past two years. His guidance has proven to be extremely invaluable during the journey of the research and production of this work. Besides his diligence in providing a lending hand when I am faced with turmoil and worries, his optimistic personality has been uplifting and made my graduate studies a gratifying experience.

I would like to thank the professors at the University of Ottawa School of Information Technology and Engineering (SITE) Vision, Imaging, Video and Autonomous Systems (VIVA) Research Laboratory, especially Dr. Martin Bouchard, Dr. Eric Dubois and Dr. Andy Adler, for their instructional support and opening my mind to the various possibilities in machine vision.

Finally, I would like thank my family for their everlasting pillar of support that I have consistently used to not only lean upon, but use as a driving force towards the completion of this thesis. Their love has made me a better, a wiser and a stronger person.

GLOSSARY

AM-CW: Amplitude Modulated Continuous Wave

CCD: Charged Coupled Device

FM-CW: Frequency Modulated Continuous Wave

HVS: Human Vision System

LUT: Lookup Table

MAE: Mean Absolute Error

MSE: Mean Square Error

RGB: Red Green Blue colour space

TOF: Time of Flight

YCrCb: Luma-Chroma-Chroma colour space

Chapter 1

INTRODUCTION

The human vision system has the unique ability to provide sensory and perceptive information that can be used to determine the depth and three-dimensional modeling of an object. Although each human eye provides redundant data, we are capable of discerning ambiguous objects. In the field of computer vision, it is the goal to help computers to "see" as humans do. By the use of visual receptors and geometric properties, it is possible to construct three-dimensional models of objects by the use of various range finding technologies. Such visual capabilities could facilitate automated processes, such as robot path planning, vehicle navigation, surveillance and security, medical imagery, and virtual reality three-dimensional modeling.

The popularity in the development of range finding technology can be seen from the recent launch of the National Aeronautics and Space Administration (NASA) Jet Propulsion Laboratory (JPL) Mars Pathfinder, which débuted its first scan of the Martian landscape on July 4, 1997 [1]. Among the various equipment used to explore Mars a stereo imaging system built from two charged coupled device (CCD) colour cameras and external housing unit for protection was introduced to scan the diverse landscapes as well as track the navigation of the Pathfinder's rover. The feedback from the imaging system helped operators back on Earth to evaluate the surroundings of the Pathfinder rover as well as perform topographical surveys of the planet. More prevalent use of range finding technology can be found in military and marine biology applications such as sound navigation and ranging or more commonly known as sonar for the use of detection of submerged objects. Whether range finding technology is used for deep space exploration or found in our daily use, its usefulness goes beyond our imagination.

The introduction of smaller and portable range sensing technologies has opened the door for researchers to explore new dimensions of the world in applications like the Mars Pathfinder. However, with scalability of technology and the reduced manufacturing costs of range sensing devices, there remains speculation of their efficiency and reliability. Even though today's consumers can purchase newly improved high-resolution cameras for a fraction of the cost of their predecessors, software implementations that handle disparity techniques have stood at a standstill with little improvement.

The introduction of active sensing techniques, such as laser range scanning [2, 3] and active triangulation [4] through structured lighting has provided a different approach to range sensing which resolves conditions where classical stereovision cannot perceive depth. For example a light pattern projected on a scene can be used to detect the depth of objects in obscure and darkened environments, whereas stereovision is dependent upon the illumination of the environment and the texturing of the scene [5, 6]. Laser range sensors, which lately have been reducing power consumption and size, have been at the forefront in depth sensing providing far more accurate depth estimation than previous techniques. However, the greatest feature drawback of most current active sensors is their ability to detect depth over only a single plane or a sparse detection grid of a non-reflective object within close proximity of the scanner. This implies the repetitive process of moving the active scanner to various and strategic poses to complete a full scene scan, unlike stereovision where one sampling is sufficient. In addition, the cost of laser range finders has remained stagnantly high recently, which forces developers to search for inexpensive methods.

With each technology providing advantages and drawbacks in their respective domains, a promising solution consists of the combination of the efforts of various range sensing techniques to create a multi-modal or a joint range sensing technique that would provide different depth perspectives of a scene from a common viewpoint. Although, this approach in providing optimal depth sensing data is very promising, the question of registration

between range sensors remains a critical issue to ensure consistency between the measurements.

The iterative and tedious manual processes, inherent in using multiple original equipment manufacturer (OEM) systems, lead to errors in registration and a lack of repeatability that must be avoided. This implies the necessity of a fully integrated sensing system, which merges various OEM solutions to ensure that each subsystem operates in tandem with each other [7]. Such an integrated solution can only be efficient if robust intra- and inter-calibration procedures are defined.

In this thesis, an original approach that achieves automated calibration within multi-modal scanning systems is proposed. The robotic integration of a prototype of a multi-modal range sensor is also discussed. Finally, the quality of the datasets and the validation of the calibration procedure are analyzed with experimental range measurements collected using the prototype.

1.1 Three-Dimensional Reconstruction

Of the many methods for three-dimensional reconstruction such as stereovision mentioned above, range sensing can be classified into two major categories: active and passive range sensing [8]. Active range sensing methods are defined when a vision system's parameters are modified purposively for reconstruction as such in the reconstruction from focusing and defocusing where the imaging system controls the camera lens to produce focused and defocused images. These two images are then used to estimate the local shape. In the method of reconstruction from motion, it is considered active if the camera is moved to various positions to generate a relative motion between sensor and scene, however it is passive if the objects in the field of view of the camera is not deliberately displaced [9]. Table 1.1 lists a few common range sensing techniques and their classifications of active and passive range sensing.

Method	Number of Images	Classification
Stereo	2 or More	Passive
Motion	Sequence	Active/Passive
Focus/Defocus	2 or More	Active
Zoom	2 or More	Active
Contours	Single	Passive
Texture	Single	Passive
Shading	Single	Passive
Structured Light	Single or Sequence	Active
Laser Range Finder (Active Triangulation, AM/FM-CW, TOF)	Not Applicable	Active
SONAR/RADAR	Not Applicable	Active

Table 1.1 – Reconstruction Methods and Their Classification

The following sections will briefly discuss the physical attributes of a few popular range sensing modalities and their reconstruction and calibration abilities. We present this data as an introduction to familiarize the reader with the technology that will be used in the in-house proposed multi-modal range sensor and its calibration and reconstruction processes.

1.2 Stereovision Reconstruction and Calibration

Passive stereo is a sensing technique where the range is computed by triangulation between the locations of matching pixels in images taken simultaneously from a pair of cameras of the same object or scene. The correspondences between pixels are established with the use of a search algorithm and by using correlation such as the sum of square difference measurements to compare local neighbourhoods [10]. This method of reconstruction resembles the human vision system (HVS) where redundant sets of images are made available and then correlated by the brain or a neural network. The raw output of the stereo system is an image that

provides the offset coordinates between the two images to distinguish depth. This offset is known as disparity and represents the inverse range between the images at each pixel [8].

The reconstruction from stereovision is dependent on the position of the two cameras to each other. The relative physical location from one camera to the other is represented by a mathematical relationship called extrinsic properties. These properties relate the translation and the rotation of the optical centre to each other. Traditionally, the relation is from the right camera optical centre to the left camera optical centre. In addition to the extrinsic properties, another mathematical relationship known as intrinsic properties must be determined between the optical centre of the camera to the image plane of the camera in pixel coordinates.

Although there are different approaches in stereovision calibration by using knowledge of planar surfaces [11], one-dimensional objects [12], or unique patterns [13], the majority of these approaches follow the traditional stereovision calibration that uses *a priori* characteristics of the calibration target. This thesis is not different as stereovision calibration is fundamentally based on Tsai camera calibration model [14], which is the foundation of camera and stereovision system calibration and effectively demonstrates how a calibration target with *a priori* patterns on planar surfaces can be used to determine both intrinsic and extrinsic camera properties.

1.3 Structured Lighting Reconstruction and Calibration

Historically, structured lighting was one of the first range imaging approaches used in robotics. The basic principle uses a spot, stripe, or grid of light that is projected onto the scene by a light projector or a laser projector. A sensor, typically a CCD camera, views the scene and extracts the projected structured light on the scene. Using a translational matrix or table, it is possible to construct a mapping between the camera image points to world points and vice versa without determining intrinsic and extrinsic properties of the camera and the

laser projector. Instead the translational matrix acts as the calibration matrix, which defines the relationship between camera image pixel locations to a desired reference frame. Unlike stereo triangulation where the difficult task of correspondence between images is used to establish object reconstruction, structured lighting avoids this problem and is practical in shape measurement tasks as long as the system is calibrated through the translational matrix.

To determine this translational matrix/table, *a priori* features that intersect the structured lighting pattern and detected by the camera are mapped together. Ideally each feature at a unique distance correlates to a pixel coordinate within the image. The coordinates of the feature in the image and its known position in the world reference frame is either kept in a table for interpolation or used in a system of equations to produce the calibration matrix.

1.4 Laser Range Finder Reconstruction and Calibration

Three major types of laser range finders are used in active laser ranging: amplitude modulated continuous wave (AM-CW) lasers, frequency modulated continuous wave (FM-CW) lasers, and time of flight (TOF) lasers. AM-CW lasers which operate by calculating the amplitude phase shift between the transmitted and received beams are commonly used in indoor environments. The accuracy and efficiency of the AM-CW lasers is increased when scanning objects at close to medium range with the use of two lasers offset in modulation frequency. This results in a much higher sampling rate at the cost of its maximum range. The method of detection of AM-CW lasers is greatly compromised by sudden changes in the range or in the surface reflectance. In these situations the reflectance change is the splitting of the light beam between surfaces of differing ranges. An example would be light being reflected across a medium with a higher refractive index (i.e. glass). Thus due to these constraints the AM-CW laser is only effective to a range of 50 meters in an ambient indoor lighting environment [8, 15, 16, 17].

FW-CW lasers, like its inverse method, depend upon the frequency shift between the transmitted and received beams of light. Yet unlike its counterpart, FW-CW is highly accurate and more efficient in outdoor settings. However, it is not widely used due to its complexity and the fragility [8, 18, 19, 20].

In replacement of the FW-CW laser, the TOF laser is just as effective, but affordable and robust. The range of an object is calculated by determining the time needed for the laser light to reach the target and return. Each laser light is emitted in a pulse format as opposed to the continuous wave, which can decay with range. Thus the pulsing laser can work to either detect singular or multiple pulse signals and can scan in either a one or two axis system. The TOF is then capable of achieving maximum ranges of one to several hundred meters to the accuracy of ± 5 centimetres. Thus the applications of TOF are mostly used in long-range mapping and navigation [17, 21, 22].

One key breakthrough in TOF is its ability to provide accurate range where surface reflectance is an issue. The method of "last pulse measuring" techniques is the concept where the train of echoes due to reflectance is detected and where the last echo (last pulse) is used for the range computation (somewhat similar to ultrasound applications). This concept ensures that in dynamic environments where fog, dust, or smoke may appear, the reflective signal from the target will always be received instead of being scattered from the medium. Although TOF laser scanners seem to be extremely suitable for long range sensing and undesirable conditions, it is still limited by its speed (few thousand samples per second) and by its acquired spatial density per scan [8].

To ensure that CCD receiver sensors are optimized to capture the reflective signal from the laser transmitter, an innovative approach called active triangulation has been shown to reduce noise and transient power. Two popular models that utilize active triangulation but with unique approaches are the *double aperture mask principle* and *synchronized spot-scanning principle* [23].

The double aperture mask principle is a novel approach by triangulating reflective signals from the scene through a camera lens shown in Figure 1-1. The camera lens focusing adjustment sets a reference plane such that if the camera lens is configured to be in focus at a given distance, the reference plane is then the same distance away from the lens. At this distance the double aperture mask has no effect and the projective point is a single point detected on the CCD image sensor device as in point A and its projective point A' on the sensor. However, any other point not on the reference plane will have two points detected on the CCD image sensor for which the separation is proportional to the distance between the distance of the point and the reference plane [24].

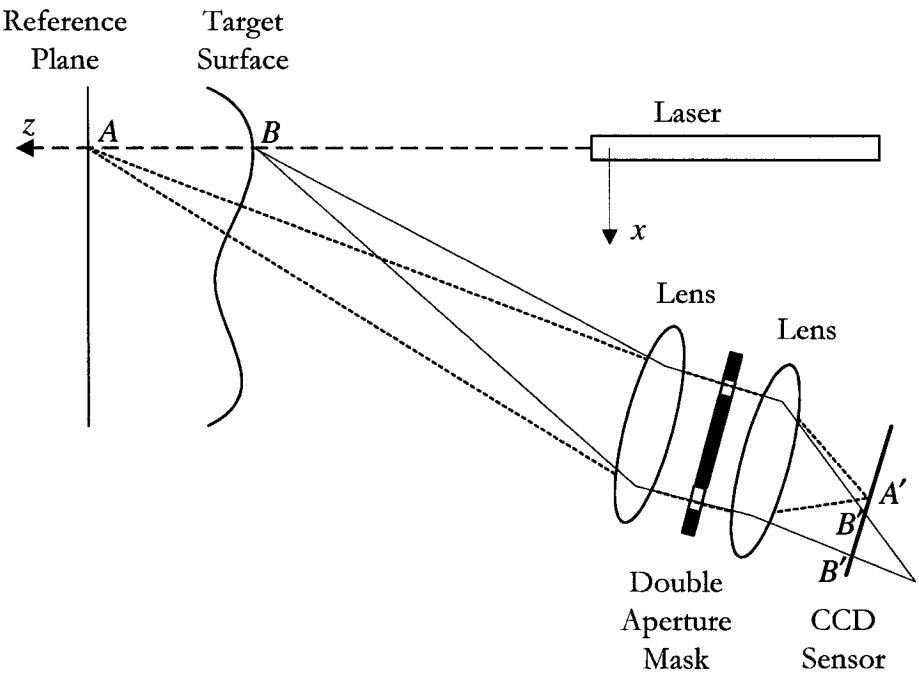

Figure 1-1 – Double aperture active triangulation laser range finder geometry [24]

The synchronized spot-scanning principle also known as the auto-synchronized scanning principle is the next evolutional step of active triangulation. In comparison to the double aperture mask principle, the synchronized spot-scanning principle requires very little optical

head size. This configuration shown in Figure 1-2 is set such that the laser beam and the detection system's optical axis are rotated synchronously through a double-sided mirror and controlled by a high accuracy galvanometer. Other implementations of synchronized spot-scanning systems have used pyramidal rotating mirrors to maintain synchronicity. The three-dimensional profile of the scanned surface is captured by emitting a laser beam forming a spot onto the object by the way of the oscillating double-sided mirror and then collecting the light that is scattered by the scene in synchronism with the projection mirror and focusing this light onto a linear position detector.

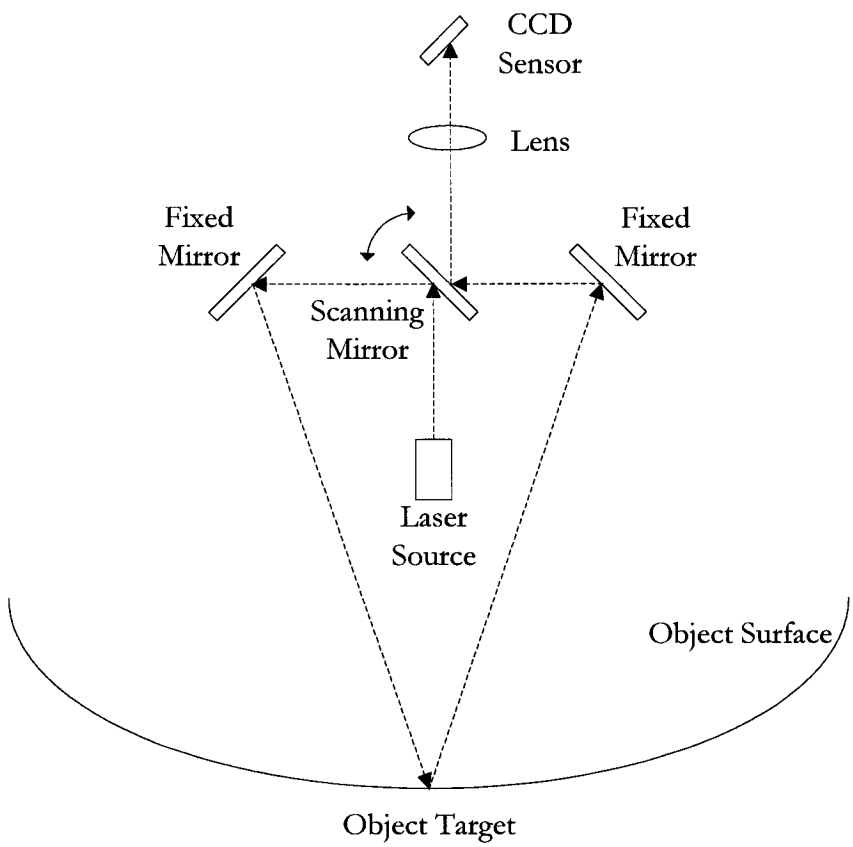

Figure 1-2 – Synchronized spot scanning active triangulation geometry [24]

The instantaneous field of view of the position detector follows the spot as it scans the scene. The lens' focal length is therefore related only to the desired depth of field or measured range and not to the total lateral field of view. With the small instantaneous field of view, ambient light immunity is greatly improved and speckle noise is reduced through spatial filtering [24].

In comparison to TOF and continuous wave active optical sensors that are capable of detecting objects with distance in the kilometre range, active triangulation methods are limited to detection ranges reaching from 0.2 to 2 metres [23].

Laser range finders are not immune to the need for calibration. As in most optical components, analysis of wavelength of the emitted laser light, its phase shift, and its amplitude are a few factors that must be analyzed to determine how the laser range finder detector should be configured to clearly isolate the reflectance of the laser light and reducing detection of light scattering and crosstalk [17]. For active triangulation laser range finders systems additional calibration is require besides optical calibration. Synchronous spot-scanning systems yield two quantities per sampling interval: one is for the angular position of the mirror and one for the position of the laser spot on the position detector. Owing to the shape of the coordinate system spanned by these variables, the resultant images are not compatible with the coordinate systems used by most geometric image processing algorithms. A re-mapping of these variables to a more common coordinate system like a rectangular system that accounts for distortion laws is therefore required [25, 26].

1.5 Problem Statement

Each reconstruction method tackles scene modelling in different approaches. For example, passive stereo is dependent on the environment lighting and the ability to extract intensity differences between foreground and background objects. Structured lighting is dependent on being able to detect a unique structured light pattern within the scene. Modalities such as structured lighting and laser range finding are dependent on the material surface reflectivity

and light absorbance of the scanned object. In the case of active triangulation and continuous wave laser range finders, the trade-off exists between accuracy and distance. Therefore the effectiveness of each model reconstruction method is specific to an environmental condition and to the type of object being detected.

Ideally, the ultimate solution would be to create a single modal range sensing system to withstand dynamic environment and recognize objects in a scene regardless of the object's light reflectance/absorbance qualities. But this approach as been researched *ad nauseam* with no single effective product.

However, another interesting approach has been taken that still promotes the positive qualities of various range sensing modalities, which is to combine similar or various range sensing technologies together to provide multiple modelling strategies. These systems are recognized as multi-modal range sensing systems.

This thesis will address a common problem in the multi-modal range sensing systems, which is the ability to correlate range between each modality. This is accomplished by the use of reconstruction and calibration methods for stereovision, structured lighting, and laser range finder modalities and calibration between each modality. The main advantage of such a correlation is the ability to provide an unbiased dataset where each modality can either confirm or deny the existence of an object within the scanned scene.

With the hopes of improving each individual reconstruction method mentioned above, the scope of this thesis is to investigate a strategy of combining various reconstruction methods to produce a more effective and more accurate reconstruction imaging system by combining stereo, structured lighting and laser range finder abilities.

1.6 Thesis Overview

In Chapter 2, the basic background of stereovision is introduced. Mathematical notation and common catchphrases are introduced as well as the benefits and weaknesses of this method. The literature review continues on the subject of structured lighting, which is a major component in the joint sensing strategy.

The problem definition that this thesis tries to resolve is qualitatively explained in Chapter 3 by analyzing and introducing similar joint sensing strategies. A new resolution is introduced to outline how the proposed joint strategy alleviates problems addressed in other strategies. The proposed joint strategy will concentrate on an approach involving calibration, which will then be applied to the design and implementation of an in-house multi-modal prototype.

Laboratory simulations, environment parameters, and results of the proposed joint sensing strategy through a simulated multi-modal system are presented in Chapter 4 to demonstrate the proof-of-concept of the multi-modal calibration. The limitations of this strategy are also analyzed to provide an objective point of view.

A detailed explanation and implementation of an in-house multi-modal system, performance and reconstruction abilities are presented in Chapter 5. Lastly, the contributions of this thesis are highlighted in Chapter 6, as well as recommendations for future improvement.

Chapter 2

MODALITY MODELLING, RECONSTRUCTION, AND CALIBRATION

To understand how to construct a multi-modal system, understanding individual modalities is imperative. This chapter provides in-depth information on the modalities presented in Chapter 1 and reveals their intricate geometric properties as well as their methods of calibration and reconstruction. We start with camera modeling as the fundamental unit of stereovision and structured lighting to the investigation on how its intrinsic properties are essential to depth perception to these systems. As these fundamental properties are explained, it will also present the required knowledge in understanding how multi-modal systems are categorized and what each unit within the system has in common with each other. In addition, presenting this information will provide the foundation of calibration techniques the proposed multi-modal system will use as well as the theoretical concepts to design it.

2.1 Camera Model and Reference Frames

Before the camera model can be described, reference frames must be introduced to determine how three-dimensional objects are represented in two-dimensional images. The first reference frame is known as the *world reference frame*, which determines the location of the objects in the scenery and the location of the camera. Likewise a *camera reference frame* exists to determine the world scenery from the camera's point of view. A third reference frame, known as the *image reference frame*, is responsible for mapping the pixel coordinates of the image plane to the camera reference frame.

The world reference frame represents three-dimensional world objects by defining three orthogonal axes (X_W, Y_W, Z_W) and an origin point, O_W. The camera reference frame also uses three orthogonal axes, but uses it in context of the camera model. The camera model is

composed of centre point or focus of projection, which is the convergence point of all light and is responsible for inverting a projected image on a retinal plane. The distance of this plane to the optical centre, O_C, is known as the focal length, f_C. A non-inverted retinal plane is known as the image plane whose distance is equivalent to the focal length. Both the retinal and image plane have a common axis, named the optical axis that runs perpendicular between each other. The point of intersection from the optical axis to the image plane is known as the principal point or the image centre, denoted as o. To demonstrate the functionality of the camera model, a given point in the scenery can be mapped directly to the image plane. From Figure 2.1, we can see that image point p is the intersection between the image plane π and the line from object P to the centre point O_C. The image plane from the camera model is also known as the image reference frame and is only defined by two axes (X_I, Y_I).

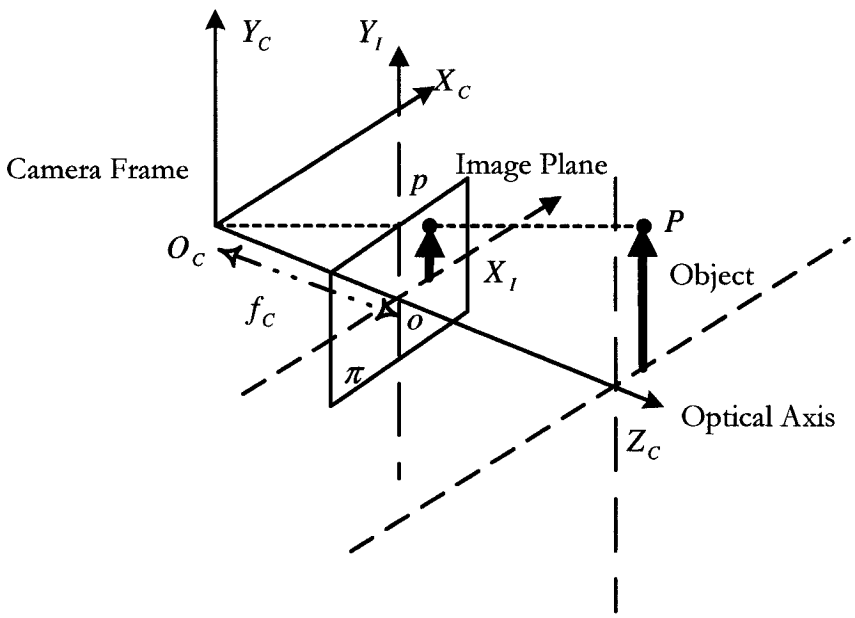

Figure 2-1 – Perspective camera model

In order to relate these three reference frames, a system of parameters, extrinsic and intrinsic, must be defined. The parameters can be determined through a process of measuring the distance and rotation of each reference frame called camera calibration. In most three-dimensional reconstruction systems, calibration parameters are estimated by identifying key feature points or by using *a priori* knowledge of the scenery. The process of calibration is a key field in machine vision and research has been quite extensive for various applications.

In order to relate the world reference frame and the camera reference frame, extrinsic parameters are used to define the location and orientation of the camera reference frame with respect to the world reference frame. This is accomplished by a set of geometric parameters that identify uniquely the transformation between the unknown camera reference frame and a known reference frame, which is usually the world reference frame. These parameters consist of a 3×3 orthogonal rotation matrix R that correlates the axes of the two references frames onto each other. A translation vector, T, is also required to describe the location of the origin between both reference frames. Thus the following equation can be derived:

$$P_C = RP_W - T \tag{1}$$

Where $P_W = [X_W, Y_W, Z_W]^T$ is any three-dimensional point within the world reference frame and $P_C = [X_C, Y_C, Z_C]^T$ are its coordinates with respect to the camera reference frame.

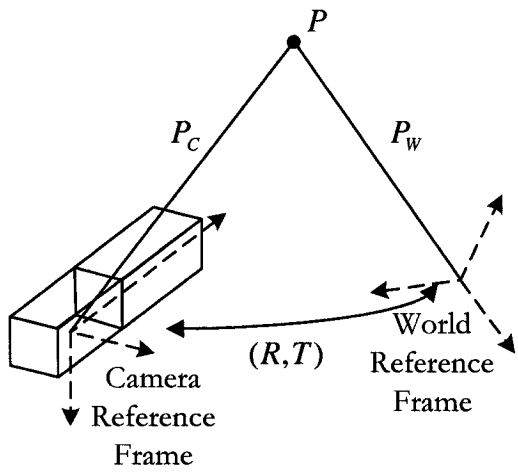

Figure 2-2 – Relation between camera and world reference frames

Intrinsic parameters define the perspective projection of a three-dimensional point onto a two-dimensional image on the image plane of the camera model. Three parameters are required to relate the camera reference frame and the image reference frame: the focal length of the camera, the transformation between the camera reference frame and the image reference frame two-dimensional axes, and the camera geometric distortion. If the relation is to be determined between the pixel coordinates in the image reference frame $[x_{IM}, y_{IM}]$ and the camera reference frame, the following terms and equations are used:

$$x_{IM} = \frac{f_C X_C}{s_X Z_C} + x_O \tag{2}$$

$$y_{IM} = \frac{f_C Y_C}{s_Y Z_C} + y_O \tag{3}$$

Where f_C is the focal length of the camera lens, s_X and s_Y are the effective pixel width and height of the camera, and $[x_O, y_O]$ is the principal point, where the principal axis intersects

the image plane of the camera in CCD elements and typically the coordinates of the centre of the image plane.

If radial distortion is neglected the two intrinsic equations can be represented in a matrix product,

$$M_{INT} = \begin{bmatrix} \frac{f_C}{s_X} & 0 & x_O \\ 0 & \frac{f_C}{s_Y} & y_O \\ 0 & 0 & 1 \end{bmatrix} \tag{4}$$

where

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = M_{INT} \begin{bmatrix} X_C \\ Y_C \\ Z_C \end{bmatrix} \tag{5}$$

then

$$x_{IM} = \frac{x_1}{x_3} \tag{6}$$

$$y_{IM} = \frac{x_2}{x_3} \tag{7}$$

One key criterion to stress with regards to intrinsic parameters is its transformation from three-dimensional information to two-dimensional. The depth information that is available from the camera reference frame is missing from the image reference frame.

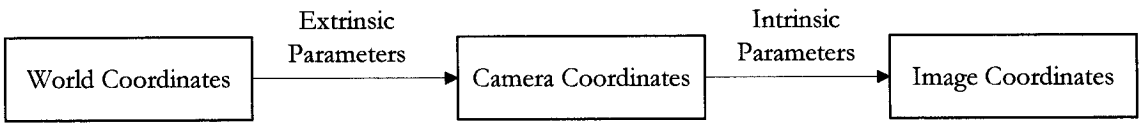

Figure 2-3 – Parameters relating world, camera, and image coordinates

Since the perspective projection is a non-linear process the recovery of the depth information is difficult. In order to alleviate this problem other camera models have been introduced such as the orthographic [27], weak-perspective [28, 29], and the paraperspective [30, 31] camera models.

2.2 Visualization from Stereo

Depth perception in human visual system (HVS) is a phenomenon that has been analyzed and modeled from mathematical, physiological, and psychological angles. In machine vision, a structured approach is implemented to extract information from the stereoscopic images. Thus the depth information is taken from images at different viewpoints. An example of this system, the typical binocular stereo camera system is illustrated in Figure 2-4.

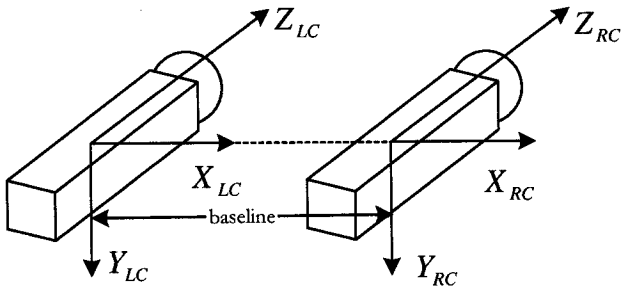

Figure 2-4 – Typical stereo camera configuration used for capturing stereo images

These two cameras are configured in such a manner that their optical axes are coplanar and aligned in parallel, yet not always in practice. Some beneficial properties of this configuration is the elimination of vertical disparity; the projections of a rectangular object within the image planes of both left and right cameras have the same area and aspect ratio, and this parallel optical axes geometry is very similar to the HVS model [32].

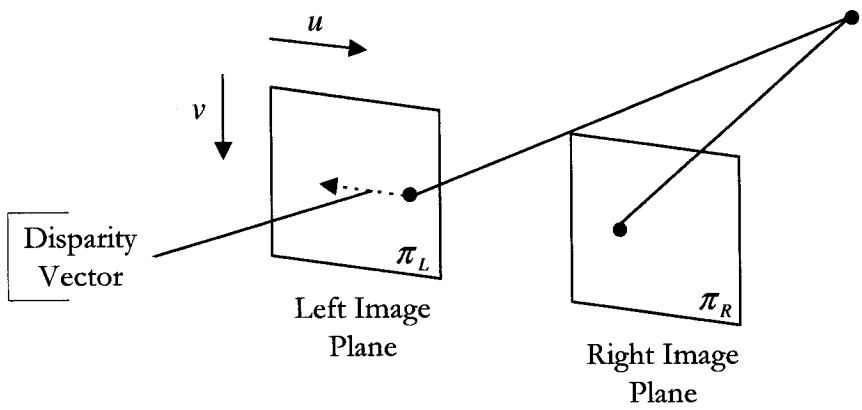

Figure 2-5 – Typical stereoscopic image geometry

The left and right cameras in the stereo system capture simultaneously in a fixed position and produce a pair of images along their image planes. The difference in the projected positions on the left $\pi_L(u,v)$ and right images $\pi_R(u,v)$ is referred to as the disparity. A disparity

vector can correlate each common point between both images to each other as shown in Figure 2-5. The collection of disparity vectors for a whole image is known as the disparity map.

2.2.1 Stereo Geometry

The first key advantage of stereo geometry deals with positioning of the cameras in relation to each other. Since most stereo models have their cameras in a fixed location or a fixed distance between each other, the intrinsic and extrinsic properties remain the same. This geometric system is known as epipolar geometry. In order to relate the coordinate systems of both cameras specific vectors must be defined. The vectors $P_L = [X_L, Y_L, Z_L]^T$ and $P_R = [X_R, Y_R, Z_R]^T$ refer to the same point, P, in the world reference frame seen from the left and right camera reference frame perspectives. The vectors $p_L = [x_L, y_L, z_L]^T$ and $p_R = [x_R, y_R, z_R]^T$ refer to the projection of P onto the left and right image reference frames respectively. For the depth coordinates of p_L and p_R, $z_L = f_L$ and $z_R = f_R$, which are dependent upon the focal lengths of the respective cameras.

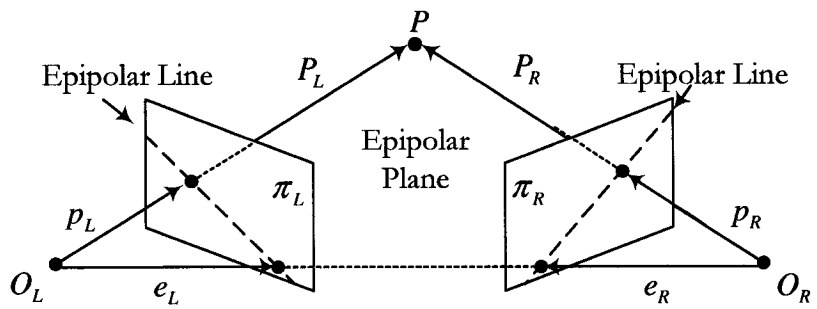

Figure 2-6 – Epipolar geometry [9]

Of the two available camera reference frames, one is designated as a reference frame to represent an object at a specific location. In this thesis the left camera is assigned as the reference frame. With this in mind a relation must be defined between the coordinate spaces of the left and right cameras.

From the known equation that relates the world reference frame to the camera reference frame in each camera

$$P_L = R_{WL} P_W - T_{WL} \tag{8}$$

$$P_R = R_{WR} P_W - T_{WR} \tag{9}$$

where (R_{WL}, T_{WL}) and (R_{WR}, T_{WR}) are the extrinsic parameters of the left and right cameras respectively. When common variable is isolated:

$$R_{WL}^{-1}(P_L + T_{WL}) = P_W = R_{WR}^{-1}(P_R + T_{WR}) \tag{10}$$

At a given point in the world reference frame that is perceivable by both camera reference frames of the left and right cameras the relation can be built to determine any point in the left camera reference frame in terms of the right camera or vice versa.

One can form a relation between the camera perspectives

$$P_L = R(P_R - T) \tag{11}$$

where R is the rotation $R_{WL} R_{WR}^{-1}$ and T is the translation vector defined as $R_{WL} R_{WR}^{-1} T_{WR} - T_{WL}$.

Other relations required are the intrinsic parameters relating the camera reference frame to the image reference frame. This is simply done by defining an arbitrary matrix, M, for the

translation and scaling. Here, p_{IL} and p_{IR} are the points in pixel coordinates in the image reference frame and p_L and p_R are the camera coordinate vectors.

$$p_L = M_L^{-1}\begin{bmatrix} p_{IL} \\ 1 \end{bmatrix} \tag{12}$$

$$p_R = M_R^{-1}\begin{bmatrix} p_{IR} \\ 1 \end{bmatrix} \tag{13}$$

To relate the world point from the camera perspectives to their image plane we rely on the equation of the perspective projection such as:

$$p_L = \frac{f_L}{Z_L} P_L \tag{14}$$

$$p_R = \frac{f_R}{Z_R} P_R \tag{15}$$

The last important vector required in epipolar geometry is the epipolar vector which is defined as e_L and e_R, that is the image projection of the optical centre on its counterpart camera, O_R and O_L respectively. The epipolar vectors and the projection vectors define the epipolar plane, which is confined between the points O_R, O_L and P as shown in Figure 2-6.

2.2.2 Reconstruction by Triangulation

If the intrinsic and extrinsic parameters are determined and the geometry of the system does not change, reconstruction is simply done by determining corresponding coordinates from both images of a featured projection point. Using knowledge of the extrinsic parameters, R and T, that define the rotation/scaling and the translation from the right camera reference

frame to the left camera reference frame, and reusing epipolar geometry where we defined a pair of corresponding vectors of a point P, p_L and p_R, the assumption is that the rays are known and the intersections can be computed. However the extrapolated vectors may not intersect exactly due to errors in feature extraction and camera calibration. Therefore their intersection is estimated to be the midpoint between the minimal distance of p_L and p_R segments.

Let a, and b be scalar variables for vectors p_L and p_R. Let the ray l from the point O_L through vector p_L to target P be represented as ap_L. Let the ray r from point from O_R through vector p_R to target P expressed in the left reference frame be represented as $T + bR^{\mathsf{T}} p_R$. In ideal conditions, as shown in Figure 2-7, the rays l and r would be equivalent to each other such that the following equation would be true:

$$\begin{aligned} l - r &= 0 \\ ap_L - (T + bR^{\mathsf{T}} p_R) &= 0 \\ ap_L - bR^{\mathsf{T}} p_R &= T \end{aligned} \tag{16}$$

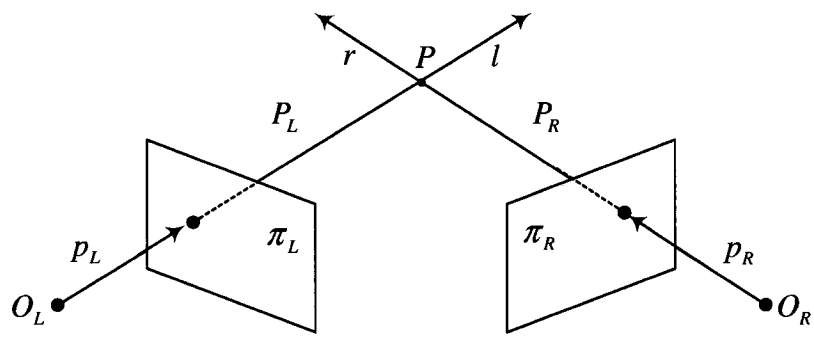

Figure 2-7 – Ideal reconstruction by triangulation [9]

Yet as explained above, eq. (16) is hardly true and thus only an estimate of P can be determined, denoted as P'. It is known that both p_L and p_R are non-parallel and thus can produce an orthogonal vector with the use of the cross product. With this in mind, eq. (16) must incorporate this new orthogonal vector.

Let c be the scalar coefficient for the orthogonal vector of p_L and p_R. Let the ray w represent the distance between ap_L and $T + bR^{\mathsf{T}} p_R$, whose magnitude is equivalent to the magnitude of the scaled orthogonal vector of p_L and p_R, shown in Figure 2-8 and defined as such:

$$\begin{gathered} l - r + w = 0 \\ w = c(p_L \times R^{\mathsf{T}} p_R) = ap_L - (T + bR^{\mathsf{T}} p_R) \end{gathered} \tag{17}$$

or in its more familiar form:

$$ap_L - bR^{\mathsf{T}} p_R + c(p_L \times R^{\mathsf{T}} p_R) = T \tag{18}$$

$$\begin{bmatrix} p_L & -R^T p_R & p_L \times R^{\mathsf{T}} p_R \end{bmatrix} \cdot \begin{bmatrix} a \\ b \\ c \end{bmatrix} = T \tag{19}$$

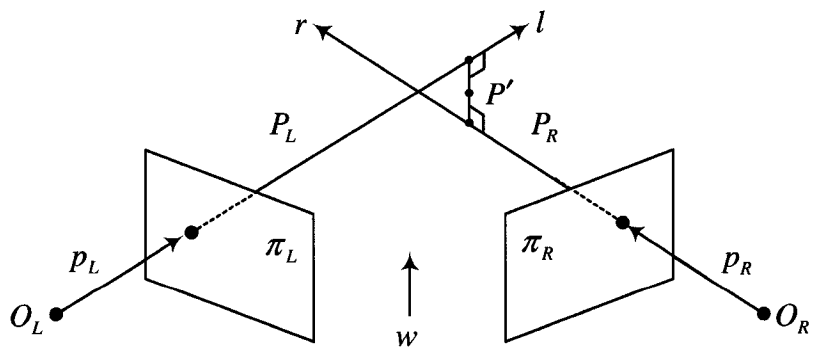

Figure 2-8 – Reconstruction by triangulation [9]

Determining the coefficients a and b by solving the linear system in eq. (19), provides the point on the rays of l and r respectively where they intersect each other or are closest to each other. In either case, a line segment can be produced from these two points, which can be named s, which is parallel to the ray w and joins rays l and r, ap_L and $T + bR^T p_R$. To determine the estimated target coordinate, P' from the left camera reference frame, the midpoint of line segment s is determined between the two points, simply defined as:

$$P' = \frac{ap_L + bR^T p_R}{2} \tag{20}$$

2.2.3 Epipolar Constraint

If the process of feature extraction was performed to obtain two sets of features, one from the left and right camera, the next process would be to find the corresponding feature in the left image for each detected feature in the right image. Without considering occlusion, it is theoretically possible to match every feature in the left image to every feature in the right image, which is a taxing process bounded in a two-dimensional region. To efficiently execute feature matching, additional constraints have been imposed. One such constraint is the epipolar constraint [9], which reduces the search from a two-dimensional region to a one-dimensional region.

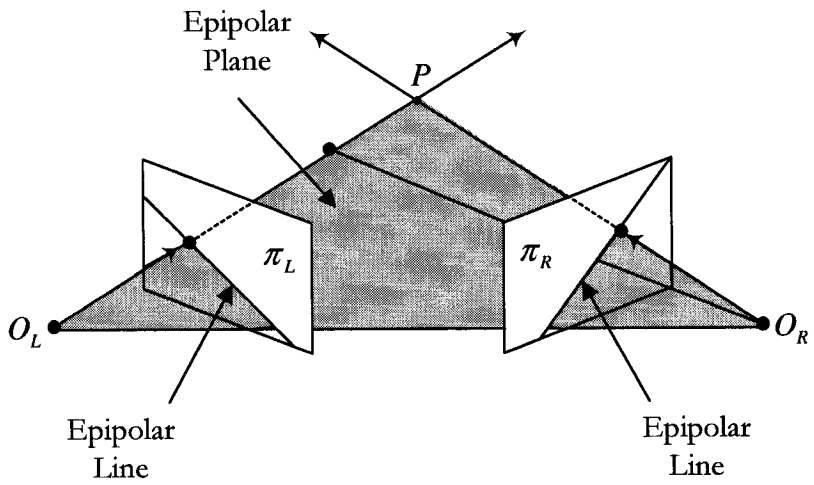

Figure 2-9 – Epipolar plane [9]

As shown in Figure 2-9, the epipolar plane is bounded by a point in space, P, and the optical centres of the left (O_L) and right (O_R) cameras. The lines that are formed where the epipolar plane intersects the image plane are known as epipolar lines, and the projections of the point must lie on these lines.

When the epipolar lines are determined in both left and right images, a scan-line comparison between the epipolar lines can be made to detect feature points. This feature comes in handy for images with sparse corresponding feature points albeit that the features are reliably extracted. However, in images where more than one feature point is extracted and lies in the proximity of the epipolar line, the depth of the point of interest cannot be uniquely determined. Therefore the epipolar constraint alone cannot guarantee that it can solve feature correspondence problems.

2.2.4 Stereo Matching Algorithm

To constrain the feature-matching problem and provide constraints to the epipolar geometry, disparity estimation algorithms are implemented to match these corresponding features. Among these algorithms are three major types of methods: feature-based, intensity-based, and phase-based.

Featured-based methods refer to the use of image features that have special properties that may correspond to parts of object or scene of structural significance. For example the use of edge properties can be used to estimate the disparity map. The choice of features to extract for reconstruction very often depends on the properties of the objects in the scene. Some important factors to consider are invariance, ease of detection, how they are eventually used, and whether they are sensitive to noise and calibration errors. When feature-based method disparity estimation is used to encode the stereo images, both the coordinates of the features and their corresponding disparities must be transmitted together, which introduces encoding complications.

In extracting specific features, which can entail the detection of edges, line, intensity variations, zero crossing of gradients, or high level template structuring, a distance metric is used to assess the similarities between the two images. The primary advantage to the feature-based method is its ability to be more invariant than actual image intensities under large viewpoint variations. Yet, due to its dependency upon the number of features in the images, sparse disparity maps may be obtained due to the lack of discernable features [32, 33].

In intensity-based methods, also known as area-based methods, points within a local region of the reference frame are used to match similar corresponding areas. Each pixel within the local region is assigned a common disparity vector based upon matching criteria with the reference frame by using either the cross correlation, sum of squared differences, the minimization of the mean absolute error (MAE) or mean square error (MSE) [34, 35]. The greatest advantage of this method is its ability to provide dense disparity vector maps, but

tradeoffs exist between the performance of the algorithm and its dependency upon scenes with large amounts of non-redundant textured regions. A commonly used method is to define a window size to limit the search algorithm depending upon the scanned scene. If the depths of objects within the scene are stationary, block-by-block disparity estimation is sufficient within the processing limitations.

Another disadvantage of this method is its assignment of disparity values to a frame based upon the search criteria. Depending upon the frame size and the windowing constraints, the granularity of precision of the disparity is inversely proportional. Even with a small frame size, the disparity value of the region does not reflect the true disparity value of every pixel. To compensate, the smoothness-constrained regularization is a commonly used intensity-based method which smoothes the disparity values at every point. This causes problems such as under-smoothing at object boundaries and over-smoothing in smooth regions.

One remaining common method for disparity estimation through stereovision is known as the phase-based method, which relies upon the Fourier phase information from both sets of images. The differences between the Fourier-phase images are used to compute dense disparity maps. The main advantage from this procedure is its avoidance of spatial feature matching. Yet a few problems can arise from phase discontinuities and unstable phase wrapping [36, 37].

2.2.5 Stereo Matching using Discontinuities

One particular stereo matching algorithm that is of interest, belongs to Birchfield and Tomasi [38] who have proposed an alternative approach of the traditional stereo algorithm by computing a rough disparity map to generate crisp discontinuities as opposed to using discontinuities to compute disparities. The basis of this algorithm depends on its ability to match epipolar scanlines independently and the ability to detect occlusions and discontinuities to produce a dense disparity map. This is accomplished by attempting to match each pixel in

the left scanline to the each pixel in the right scanline. For each pixel, two possible outcomes can be attained: *match* and *occluded.* For a match between the ordered pair of pixels, the intensity of the pixel in the left scanline, $I_L(x)$, is equivalent to the intensity of the pixel in the right scanline, $I_R(y)$, where x and y are the pixel location in the left and right scanline respectively. Likewise, unmatched pixels are *occluded* and adjacent occluded pixels bordered by non-occluded pixels form an occlusion.

Each match detected is then encoded in a match sequence, M, which contains only successful matches within the scanline as shown in Figure 2-10. In this particular figure, the match sequence is $M = \langle (1,0),(2,1),(6,2),(7,3),(8,4),(9,5),(10,6),(11,9),(12,10) \rangle$ and where the five middle pixels correspond to a near object.

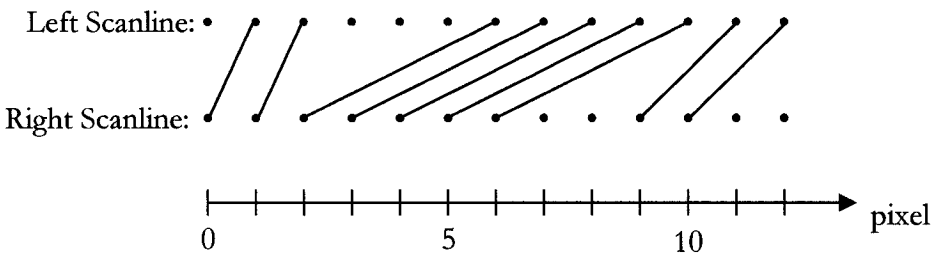

Figure 2-10 – Match sequence (reproduced from Birchfield *et al.* [38])

The disparity, $\delta(x)$, of a pixel in the left scanline at position x that matches a pixel y in the right scanline is defined as $x - y$, where as the disparity of occluded pixels are assigned disparities farther of the two neighbouring objects or the lowest possible disparity value. When all the disparities matched or occluded have been determined, a dense disparity map is produced and using the stereo reconstruction by triangulation process described in Section 2.2.2, depth or more specifically three-dimensional coordinates from the camera perspective can be determined. Before this can be achieved, the most important question is how the stereo algorithm will determine what pixel is considered matched or occluded.

To determine the likelihood of a match between individual pixels a cost function $\gamma(M)$ is defined by using a constant penalty for each occlusion, a constant reward for each match, and a sum of dissimilarities between matched pixels.

$$\gamma(M) = N_{occ}\kappa_{occ} - N_m\kappa_r + \sum_{i=1}^{N_m} d(x_i, y_i) \tag{21}$$

where κ_{occ} and κ_r are the occlusion penalty and match reward respectively, N_{occ} and N_m are the number of occlusions and matches respectively, and $d(x_i, y_i)$ is the dissimilarity between the pixels all within an epipolar scanline match sequence, M . The dissimilarity function is simply a measure to determine the minimal intensity variance between a point in the left epipolar scanline $I_L(x_i)$ and its matching point in the right epipolar scanline $I_R(y_i)$.

$$d(x_i, y_i) = \min\left|I_L(x_i) - I_R(y_i)\right| \tag{22}$$

The cost function $\gamma(M)$ works extremely well to produce piecewise-constant disparity maps. Therefore if a given object has contours or varying depths, it is assigned a single disparity value, such as in the case of a cylindrical surface. This algorithm sacrifices accurate scene reconstruction, but facilitates precise localization of depth discontinuities by emphasizing the change in disparity at the object's boundaries.

2.3 Visualization from Structured Lighting

Structured lighting is considered to be the basic principle of active range sensing. Structured lighting has been widely used as a range sensor due to its simplicity and the ease of calibration. The concept of structured lighting is based upon the projection of a known light pattern onto

a scene and extrapolating the features of the projected pattern from a single camera. This projected pattern can be in generic forms such as a projected point, linear stripe [39, 40], and grid [41, 42] or as a complex projection such as a varying pattern [43] or even a colour pattern [44, 45]. Based upon the observed location of the pattern in the captured camera image it can be correlated against a transformation matrix to determine the depth of the object at the key feature.

2.3.1 General Structured Lighting Geometry

The entire geometry of the structured lighting system is shown in Figure 2-11. Two major components are required, the light projector and the intensity camera (camera capable of capturing the projected pattern), which are placed at a distance known as the baseline starting from the centre of projection of the camera. For a given observed point in the scene, $P(X,Y,Z)$, its origin reference frame is the centre of projection of the camera. Thus all observed and extracted points from the system are given with respects to the reference of the camera.

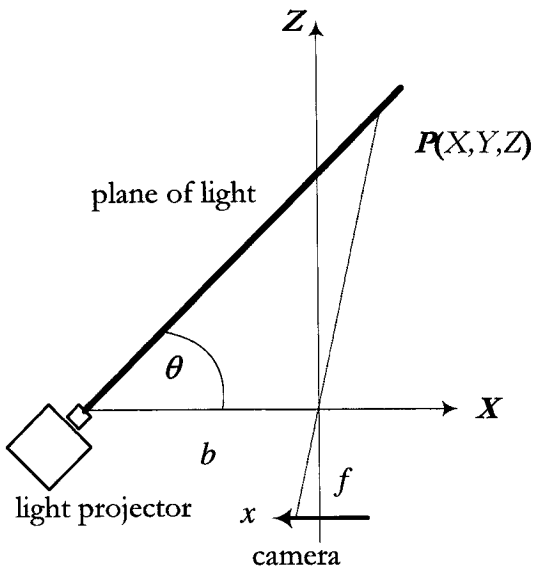

Figure 2-11 – Basic geometry of a structured lighting system [9]

Assume that the projected plane of light is perpendicular to the ***XZ*** plane at a projection angle of the plane of light θ with respects to the ***XY*** plane. The camera observes the intersection of the projected pattern with the scene surface. The depth of this detected intersection can be related by:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \frac{b}{f \cdot \cot\theta - x} \begin{bmatrix} x \\ y \\ f \end{bmatrix} \tag{23}$$

where b is the baseline separation between the projected light and the camera optical centre, f the focal length of the camera lens, x and y the location of the intersection detected in the image, and θ the projection angle of the laser. If the entire structured lighting stripe pattern can be extracted from the captured camera image, the above equation can be used to build a depth profile of a single slice of the scene [9]. To capture a full scene depth profile,

multiple slice acquisitions can be performed by either moving the scene/object under the structured stripe or by moving the light projector to sweep the scene.

The most important factor in structured lighting is the detection of the projected pattern on the scene. This helps correlate the relationship between the scanned object point and the detected coordinates of the point within the image. To ensure that the structured lighting system accurately observes the pattern many approaches can be taken. The most commonly used approaches are by changing the illumination conditions that affect the environment lighting and the brightness of the stripe allowing the camera to differentiate environment lighting to the stripe.

2.3.2 Structured Lighting Direct Calibration

The simplest method to calibrate a structured lighting system is to use direct calibration as proposed by Trucco and Fisher [46]. Direct calibration's main advantage is its efficiency and simplicity that allows the system to be calibrated without the use of mathematical equations. Just the use of image features is required for this calibration.

The direct calibration procedure builds a lookup table (LUT) linking image and three-dimensional coordinates. To effectively use the LUT a large amount of calibration data must be provided. To do this, a dense grid of workspace points (*a priori* points which form a calibration grid) with known world coordinates is measured accurately and stored in the LUT. If the camera detects an illuminated feature within the workspace, its position in the world reference frame can be obtained by inverting the resulting LUT and by interpolating between the surrounding points.

The calibration procedure is broken down into two distinct stages. The first stage consists of building the calibration grid. A calibration block is constructed consisting of a specific number of steps (stair-like) with incrementing heights and is placed under the structured

stripe of light such that one single entire step intersects completely with the plane of light. A camera observes the stripe as linear segments and its position in the image is recorded for the height of the step. The calibration block is then repositioned such that the adjacent step of the calibration block is the next target for capturing and reprocessing. When sufficient amounts of samples are taken, the LUT table presents itself as a calibration grid.

The second stage of direct calibration is the process of constructing image-world maps. By using the LUT collected in the first stage, the calibration grid is inverted and interpolated using linear interpolation to obtain a complete LUT where any captured image pixel is associated to a three-dimensional point within the calibrated workspace.

Although it would be possible to determine the fixed coefficients of the structured lighting system above, a simpler and more efficient method of projector calibration can be used to determine the mapping matrix from image reference frame points to world reference frame points from an initial position of the system. For example, Trucco and Fisher requirement for accurate calibration sampling by the placement and orientation of an *a priori* calibration target is far too idealistic. In addition, two deficiencies in this approach are that the LUT mapping is only valid for detected structured lighting features within the workspace and the fact that dense sampling must be fed to the LUT in order for interpolation to be effective. Even interpolation becomes ineffective in improving three-dimensional world coordinates from features detected outside of the LUT. The following section introduces another structured lighting calibration method similar to direct calibration where low-level details of the aiming vectors for the camera and structured lighting projector are not required and three-dimensional estimates of features beyond the workspace can be determined.

2.3.3 Structured Lighting Projective Calibration

Another popular method of structured lighting uses the concept of projective geometry [47, 48, 49, 50], whose main objective is to find a formula that converts points from the image

plane of a camera into world coordinate points of the corresponding object point. The origin of world coordinate points can be any given world-coordinate point as long as it remains consistent and constant in the world reference frame. Another key objective of projective geometry is the ability to avoid the arduous calibration of determining the aiming vectors of the camera and laser projector.

To understand projective calibration it is important to understand the basis of projective geometry, as it will play an integral role in defining the calibration process for the multi-modal sensor. Fortunately, this basis can be explained in one-dimensional terms and easily expanded for two-dimensional projectivity [50].

2.3.3.1 One-Dimensional Projectivity

Consider a one-dimensional plane, where a defined point, P, is the centre of projection. Two unique lines on the same plane, s and r, are defined such that they do not pass through P and are collinear or non-collinear to each other. Denote the variable X to be a point on line s and it's projective image as X' on line r, such that the line segment PX intersects X' as shown in Figure 2-12.

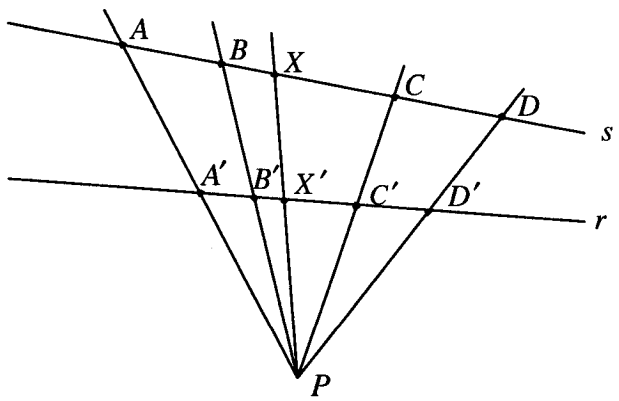

Figure 2-12 – One-dimensional projectivity [47, 52]

Four additional distinct points are added along line s, denoted as A, B, C, and D. The cross ratio (double ratio) of the points A and B with respects to points C and D is defined to be the ratio of the division ratios of segments AC/BC and AD/BD (also uses the notation (AB,C) and (AB,D) respectively), that is

$$(A,B;C,D) = \frac{AC}{BC}\Big/\frac{AD}{BD} = \frac{AC}{BC}\cdot\frac{BD}{AD} = \frac{AC}{AD}\cdot\frac{BD}{BC} \tag{24}$$

Likewise the projective points, A', B', C', D' form a cross ratio equivalent to $(A,B;C,D)$, which can be expressed as:

$$(A,B;C,D) = (A',B';C',D') \tag{25}$$

or

$$\frac{AC}{BC}\frac{BD}{AD} = \frac{A'C'}{B'C'}\frac{B'D'}{A'D'} \tag{26}$$

Since this relationship defines the projectivity between s and r lines, if the projectivity of X were to be evaluated, this could be accomplished by substituting X for D and X' for D'.

$$(A,B;C,X) = (A',B';C',X') \tag{27}$$

It is noted that the two corresponding sets of triplets (A,B,C) and (A',B',C'), assuming that they are from distinct points on distinct lines s and r can determine one and only one projectivity from point P [51]. This is the basis behind the fundamental theorem of one-dimensional projectivity, which states:

> Given three distinct collinear points on one line and another three distinct collinear points on a second line, there is one and only one projectivity which carries the first triple (A,B,C) respectively into the second triple (A',B',C') [47, 52].

In order to use one-dimensional projectivity and to determine the projectivity X' in applicable terms, a coordinate system must be defined to represent the position of the point on a line. The familiar coordinate system on the line is established by selecting a point of origin, O, from which all measurements along the line are to be made, a unit of measure, and a sense of (positive) direction along the line. This implies that two points O and U are required and are assigned the coordinates 0 and 1 respectively. The non-homogeneous coordinate x of any third point X on the line is then the directed distance of X from O. If a homogeneous coordinate system on the line is desired, this can be done by establishing coordinate (0, 1) to O, (1, 1) to U, and to any other point X with non-homogeneous coordinates x, any pair of coordinates (x_1,x_2) such that $x_1 / x_2 = x$ as shown in Figure 2-13 on line o.

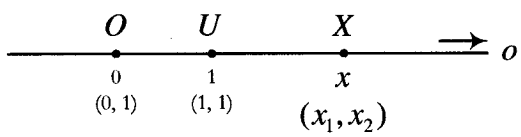

Figure 2-13 – One-dimensional coordinate system [47, 52]

Assuming that the origin O and the unit point U are used to define a coordinate system for a line s. In addition, assume that O' and U' have been used to define a coordinate system for another line r. It is not imperative that the unit length OU on line s be equal to the unit length $O'U'$ on line r. It is also not required that the points O' and U' be images of the points O and U under any projectivity. Thus eq. (27) is independent of the coordinate systems defined on either line in the projectivity and follows the theorem of cross ratio as stated:

> The cross ratio of any four points on a line is independent of the coordinate system established on the line [52].

Given a point x on line s, it is a simple matter to derive a formula for the corresponding point on line r. With respect to the coordinate system on line s, let the points A, B, C, X have coordinates, a, b, c, x respectively. Similarly on line r, let the points A′, B′, C′, X′ have coordinates a', b', c', x' respectively. Eq. (24) can then be rewritten as:

$$\frac{(a-c)(b-x)}{(b-c)(a-x)} = \frac{(a'-c')(b'-x')}{(b'-c')(a'-x')} \tag{28}$$

By substituting $\frac{(a-c)}{(b-c)} = \alpha$ and $\frac{(a'-c')}{(b'-c')} = \beta$, x' can be isolated in terms of x.

$$x' = \frac{(\alpha a' - \beta b')x + (ab'\beta - a'b\alpha)}{(\alpha - \beta)x + (a\beta - b\alpha)} \tag{29}$$

or

$$x' = \frac{a_{11}x + a_{12}}{a_{21}x + a_{22}} \tag{30}$$

where

$$a_{11} = \alpha a' - \beta b'$$
$$a_{12} = ab'\beta - a'b\alpha$$
$$a_{21} = \alpha - \beta$$
$$a_{22} = a\beta - b\alpha$$

In terms of homogeneous coordinates, x and x' are replaced with x_1 / x_2 and x_1' / x_2' respectively, such that:

$$\frac{x_1'}{x_2'} = \frac{a_{11}x_1 + a_{12}x_2}{a_{21}x_1 + a_{22}x_2} \tag{31}$$

or

$$\begin{cases} \rho x_1' = a_{11}x_1 + a_{12}x_2 \\ \rho x_2' = a_{21}x_1 + a_{22}x_2 \end{cases}, \rho \neq 0$$

In matrix form

$$\rho \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \tag{32}$$

Since a point in homogeneous coordinates does not have a unique expression, such that $x' = x_1 / x_2 = \rho x_1 / \rho x_2$, the introduction of the free variable ρ, helps ensure that regardless of the homogeneous coordinates chosen, the projectivity solution will always satisfy eq. (32). Additionally, the roles of X and X' can exchange such that another form of the matrix in eq. (32) is produced such as:

$$\rho \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} \\ e_{21} & e_{22} \end{bmatrix} \begin{bmatrix} x_1' \\ x_2' \end{bmatrix} \tag{33}$$

2.3.3.2 Two-Dimensional Projectivity

As for one-dimensional projectivity in two-dimensional space, a similar formalism can be used for two-dimensional projectivity in three-dimensional space. Consider two planes in

space denoted as s and r and a point P, the centre of projection, which does not lie on either planes. For every point X that lies on the s plane, it forms a line segment, PX, which intersects the r plane. This point of intersection of PX with the r plane forms the image point X' as in Figure 2-14.

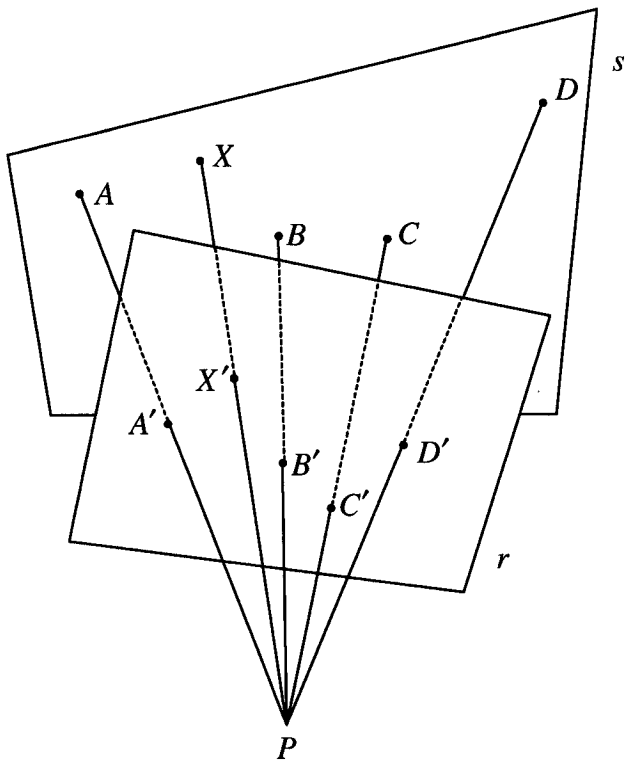

Figure 2-14 – Elements of two-dimensional projectivity [47, 52]

The invariance of the cross ratio is still valid in two-dimensional projectivity for any four collinear points on s and their image points on r. In addition for any collinear point on s, their image points on r are also collinear. Extending the fundamental theorem of one-dimensional projectivity for two-dimensional purposes:

> Given four distinct non-collinear points on a plane and another four distinct non-collinear points on the other plane, there is one and only one projectivity which carries the first quad of points (A, B, C, D) respectively into the second quad of points (A', B', C', D') [47, 52].

To establish a homogeneous coordinate system on a plane, the one-dimensional approach is simply extended. This is done by establishing a point of origin in a plane and using two orthogonal unit points. An example would be to designated (0, 0, 1) as the origin and (1, 0, 1) and (0, 1, 1) as the orthogonal points to build the coordinate frame in the plane. The homogeneous coordinates of any point in the plane are given by (x_1, x_2, x_3) with $x_3 \neq 0$. Similarly to the derivation of eq. (32), a 3×3 conversion matrix is used to convert any homogeneous point X on plane s to its homogeneous image point X' on plane r as shown in eq. (34).

$$\rho \begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{34}$$

If the roles of X and X' are exchanged, the form of generic equation in eq. (34) is still retained as follows:

$$\rho \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \cdot \begin{bmatrix} x_1' \\ x_2' \\ x_3' \end{bmatrix} \tag{35}$$

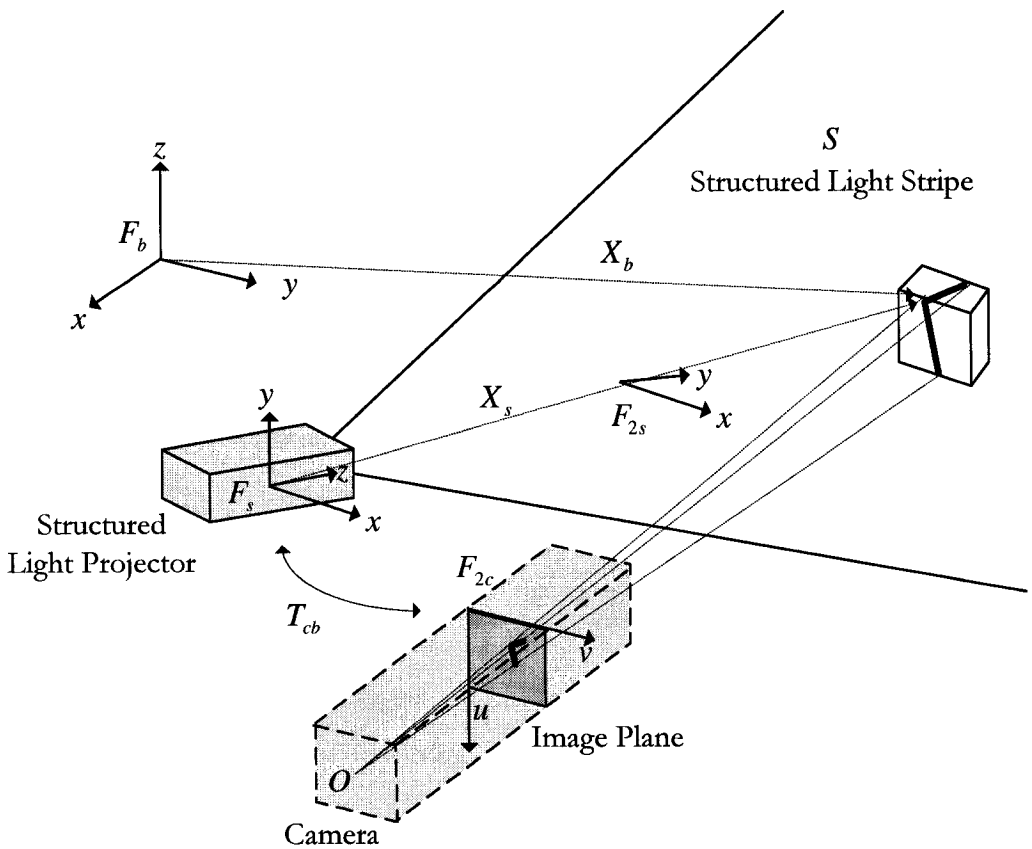

Figure 2-15 – Structured lighting two-dimensional projective model

The two-dimensional projectivity model can be applied to the structured lighting system by setting the camera-focus as the centre of projection O, the structured light stripe as the plane S, and the image plane as plane r. This construction also assumes that the classic pinhole camera model applies to the camera [47]. The coordinate system of the image plane can be arbitrary, however a convenient definition consists of using the row index u and the column index v and selecting the centre of the image plane as the origin. The coordinate frame on the image plane is denoted as F_{2c}. Therefore, any point on the image plane has the

coordinates (u, v) or has the coordinates $(u, v,1)$ if a homogeneous coordinates system is used with respects to F_{2c}. A coordinate system is also required for the light stripe plane, which is based upon any global frame of reference. This frame of reference is denoted as F_b and for every point, X_b, represented from this frame will have homogeneous coordinates $(x, y, z,1)$. To have this frame of reference pertinent to the structured light stripe, a translation and a rotation from the world reference point, F_b, to the structured light stripe plane S are defined. This new frame of reference, F_s, inherits the coordinates system defined on the xy plane of the frame F_b and retains information regarding the translation and rotation from F_b. Thus a subset coordinate system with these two axes can be formed for a new coordinate system F_{2s} specific to the plane S. Therefore a three-dimensional point X_s on plane S, which is assigned homogeneous coordinates (x_1, x_2, x_3) with respects to F_{2s}, where $x_3 \neq 0$, has the same homogeneous coordinates $(x_1, x_2, 0, x_3)$ with respects to F_s. The conversion of any point X_s from its two-dimensional frame of reference F_{2s} to its three-dimensional frame of reference F_s can be rewritten as:

$$\begin{bmatrix} x_1 \\ x_2 \\ 0 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \tag{36}$$

Determining the representation of X_s with respect to the base frame F_b, can be accomplished by multiply X_s with F_s, which is a 4×4 homogenous conversion matrix.

$$X_b = F_S \cdot X_S \tag{37}$$

By replacing (x_1', x_2', x_3') with $(u, v,1)$ in eq. (35) and combining eq. (36) and eq. (37), a 4×3

conversion matrix T_{cb} is produced that converts a point U in the camera image plane to the light stripe point X_b in the defined global coordinate frame.

$$X_b = T_{cb} \cdot U \tag{38}$$

or

$$\rho \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \\ t_{41} & t_{42} & t_{43} \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \tag{39}$$

Again the use of the free variable ρ accounts for the non-uniqueness of homogeneous coordinate expressions. In order to avoid T_{cb} becoming rank deficient the value of t_{43} is set to 1.

The key component to projective geometry structured lighting is to determine the conversion matrix T_{cb}. This requires knowledge of the positions and the orientations of the camera and of the structured light projector. However, to determine this conversion is far more complicated and over-idealistic, due to positioning errors and dealing with intrinsic properties of the structured light projector. The purpose of calibration is therefore to determine T_{cb} without actually measuring these positions and orientations.

2.3.4 Determining the Conversion Matrix

According to Chen and Kak, the two-dimensional projectivity exists between the camera image plane and the structured light plane [47]. The conversion matrix T_{cb} can be determined with knowledge of the positions and the orientations of the camera and light

plane projector, but their calibration procedure can determine this matrix without the need of this information. To determine this projectivity conversion matrix, a minimal set of four coplanar but non-collinear points in the light plane and their corresponding points in the image plane are required. Therefore at least four illuminated object points are needed as calibration points, their three-dimensional coordinates and the corresponding image coordinates are sufficient to solve the conversion matrix T_{cb}, which can be accomplished by using *a priori* knowledge of calibration target and *a priori* knowledge of the orientation of the calibration target placed in the base reference frame F_b.

Assuming that all required information from the four calibration points is determined the conversion matrix can be easily solved by a system of linear equations. Rewriting eq. (39) to replace T_{cb} row vectors and the input image coordinates into a single variable with $U = [u_i, v_i, 1]^T$ and T_j, $1 \leq j \leq 4$ as line vectors of the projection matrix:

$$\rho \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} = \begin{bmatrix} T_1 \\ T_2 \\ T_3 \\ T_4 \end{bmatrix} \cdot U \tag{40}$$

This leaves four equations:

$$\rho x = T_1 \cdot U \tag{41}$$
$$\rho y = T_2 \cdot U \tag{42}$$
$$\rho z = T_3 \cdot U \tag{43}$$
$$\rho = T_4 \cdot U \tag{44}$$

To eliminate the free variable ρ, eq. (41), eq. (42), and eq. (43) are divided by eq. (44) to produce:

$$x = T_1 \cdot U / T_4 \cdot U \tag{45}$$
$$y = T_2 \cdot U / T_4 \cdot U \tag{46}$$
$$z = T_3 \cdot U / T_4 \cdot U \tag{47}$$

Or equivalently in system equations:

$$T_1 \cdot U - xT_4 \cdot U = 0 \tag{48}$$
$$T_2 \cdot U - yT_4 \cdot U = 0 \tag{49}$$
$$T_3 \cdot U - zT_4 \cdot U = 0 \tag{50}$$

For each calibration point three linear equations are produced in terms of twelve coefficients of T_{cb}. Therefore only four calibration points are sufficient to determine twelve linear equations to solve for the twelve unknowns.

2.3.5 Structured Lighting Automated Calibration

To retrieve the calibration samples necessary to solve the coefficients of T_{cb}, Chen and Kak [47] have devised a method that relies on the precision of a robot end-effector. The structured lighting system is placed as the robot end-effector, such that the system can be moved in the workspace to collect calibration points. However, not all calibration points may be illuminated by the light plane due to the mobility of the robot limited by discrete steps. To avoid this difficulty, extended objects are placed in the work area, which have the ability to produce at least four non-collinear points illuminated by the light plane.

One approach to determine the world coordinates of these objects would be to first collect the vision data from the illuminated samples and then to move the robotic end-effector to each of the calibration points. Based upon the mechanical calibration of the robot the calibration points could be determined with respects to the robot's base frame of reference. However, this would imply that the robot calibration must be accurate and the end-effector must be able to precisely reach each calibration point.

The second approach is to create a calibration target where the robot knows the edges of the objects with respects to the world reference frame. This requires a set of equations that define the edge of the object. This edge, better represented as a line, can also be defined as the intersection of two planes such as:

$$\begin{cases} a_1 x + b_1 y + c_1 z = d_1 \\ a_2 x + b_2 y + c_2 z = d_2 \end{cases} \tag{51}$$

As the structured light projection intersects the target edge (calibration line), its intersection point can be determined in image coordinates, U, by identifying the endpoints that the structured light projection makes with the calibration target. From this point only the image coordinates are known and the three-dimensional world reference coordinates of the illuminated calibration points are unknown. However, there is no need to determine the three-dimensional world points since it can be replaced by the product of the conversion matrix with the image coordinate points as shown in eq. (52) where x, y, and z in eq. (51) are substituted by eq. (45), eq. (46), and eq. (47) respectively.

$$\begin{cases} a_1 T_1 \cdot U + b_1 T_2 \cdot U + c_1 T_3 \cdot U = d_1 T_4 \cdot U \\ a_2 T_1 \cdot U + b_2 T_2 \cdot U + c_2 T_3 \cdot U = d_2 T_4 \cdot U \end{cases} \tag{52}$$

Therefore each calibration line produces two equations in terms of the twelve coefficients of the conversion matrix T_{cb}. It would require a minimum of six calibration lines to form twelve equations and solve for the twelve unknowns. However, it is not necessary to produce six different calibration lines as the robot end-effector can position itself to a different viewpoint and sample from the same calibration line.

As the end-effector moves to a new location, the line equations change to compensate for the movement. Eq. (52) now includes the new displacement of the end-effector by (d_x, d_y, d_z) as shown:

$$\begin{cases} a_1 T_1 \cdot U + b_1 T_2 \cdot U + c_1 T_3 \cdot U = (d_1 - a_1 d_x - b_1 d_y - c_1 d_z) T_4 \cdot U \\ a_2 T_1 \cdot U + b_2 T_2 \cdot U + c_2 T_3 \cdot U = (d_2 - a_2 d_x - b_2 d_y - c_2 d_z) T_4 \cdot U \end{cases} \tag{53}$$

With two calibration lines, only three positions (including the initial position) are required to solve the conversion matrix T_{cb}.

2.4 The Importance of Calibration

Calibration methods introduced in this chapter present the foundation by which today's sensing modalities follow as unbreakable tenet. There is no exception when calibrating individual modalities in this thesis, however certain methods have been adapted for flexibility and the use of complementary data hastens mathematical operations with greater precision and the reduction of dependency on *a priori* points, as in the example of structured lighting. In multi-modal range sensing systems, whose main goal is to provide complementary data to reconstruct object models, complementary assets can also serve in expediting and simplifying individual calibration processes to the degree where the multi-modal system is capable of automated calibration. In the next chapter, various multi-modal solutions are investigated

and a new prototype for a multi-modal system is presented using the extension of calibration procedures presented in this chapter. Simulation of this prototype multi-modal system is also presented in Chapter 4 as well as experimental results of the system used in scene reconstruction to wet the appetite of the reader.

Chapter 3

MULTI-MODAL RANGE SENSING SYSTEM STRATEGY

Although the idea of combining multiple range sensing systems is not a recent proposal and there has been use of multi-modal sensing strategies in today's range sensing systems, there has been little classification or devoted literature on this subject. However, amidst the lack of explicit discussion of multi-modal range sensing strategies, such systems are visible in the literature. The reason for the lack of detailed publications is mostly due to the numerous possible methods that multi-modal range sensing can be achieved. Depending upon the objective and application of the system, whether used for scene scanning, robot path planning or object differentiation, there lie many combinations of multi-modal range sensors. Before the proposed strategy of multi-modal range sensing is introduced, an investigation of a few current multi-modal range sensing techniques is presented based on range sensing techniques discussed in Chapter 2. From the analysis of these existing multi-modal sensors, we build upon their concepts to produce a solution that improves on these methods. To prove that the proposed solution is applicable to multi-modal system a prototype multi-modal sensing system is designed and implemented using the solution guidelines. Theory presented in Chapter 2 is used to elaborate the design of this multi-modal system and to determine the required steps in building a successful model, its physical construction, and its calibration acquisition process.

3.1 Omnidirectional Laser Range Finder and Panoramic Video Camera

The article by Laurent *et al.* [53] discusses an approach of robot localization based on multisensor cooperation. This prototype multisensor, named SARAH, presents the use of two range sensing systems: an omnidirectional system made of a panoramic video camera

coupled with a rotative laser range finder. The main reason for using a multisensor approach is to improve safety of the system's algorithm of localization pattern matching and to reduce the position determination computing time. Therefore localization can be determined using complementary data.

The SARAH perception system consists of two omnidirectional sensors. The first builds from a CCD camera and a conic mirror and the second from a laser and a rotative mirror. Each sensor works independently of each other and simultaneously. The laser range finder system that utilizes a rotative mirror to reflect the beam in any direction is placed on the base of the cone. Instead of using multiple cameras to generate an omnidirectional view of the scene, a single camera is used with a conic shaped mirror, which reflects the entire surrounding scene to the CCD camera as shown in Figure 3-1.

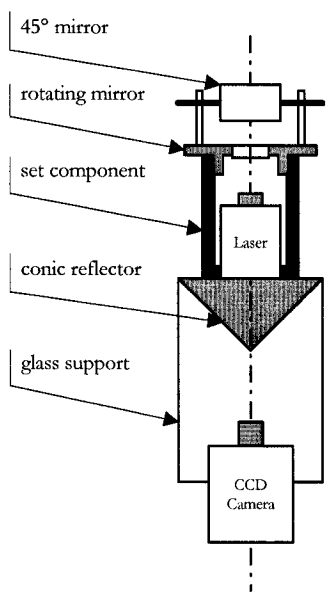

Figure 3-1 – Schema of the omnidirectional system (reproduced from Laurent *et al.* [53])

The panoramic range finder is composed of a fixed laser range finder and a mobile component, which is a 45-degree mirror that reflects the laser outwards. This rotative mirror

is moved by a step engine, which permits a one degree maximum angular precision. Therefore one complete rotation of this system can map a slice of the scenery.

In order for the camera system to extract radius data from the omnidirectional image, which is a ring between two concentric circles produced by the conic reflector, Sobel edge detection and threshold filtering is performed to keep the strongest values of gradients. Finally a segmentation algorithm with a linear regression method is executed to extract a series of straight lines and determine their equation following the linear form, $Y = aX$. A straight line that intersects the centre of the image can easily characterize a vertical object that is placed within the environment. Each beacon's location is distinguished by the angle of the intersection line.

To correlate objects observed by both sensing systems, natural beacons are placed strategically in the environment such that the projection of these beacons are detectable by the omnidirectional camera. These beacons can consist of objects, pillars, or hallway entrances, such that the image of these beacons is identified as contrasting lines that converge toward the centre of the image. Regardless of the remaining surrounding environment, these beacons should be easily distinguishable.

SARAH permits the association of two complementary datasets, both of which consists of lined up points characterized by the partitions in the room and a series of vertical lines (lines that intersect the centre of the dataset) generated by the division between these partitions. By using the characteristics of the divisions, both datasets can be used to correlate common environmental features in the room and extract a position of the perception system within the generated global map without odometer errors.

3.2 Sonar Ring Sensor and Hybrid Structured Lighting Stereovision

Another example of a multi-modal system is the McGill QUADRIS sensor platform, which utilizes a hybrid structured lighting stereovision system also known as the BIRIS coupled with a sonar sensor ring for robot navigation [24, 54]. The QUADRIS presents an efficient range sensing system by introducing selective processing with the dual structured lighting system. This approach called "just in time sensing", allows the QUADRIS to use the structured lighting system only when the preliminary interpretation of sonar data appears to be insufficient. The use of sonar sensing appears frequently on mobile robots due to their low cost, their simplistic processing, and rapid data acquisition ideal for observing large regions of space. However, sonars possess limitations when used as range sensors. Measurements tend to have low spatial resolution and the effects of multi-bounce specular reflections typically corrupt the observed data. On the opposite end of the spectrum, structured lighting sensors are able to obtain relatively accurate data with fewer artefacts when compared to simple sonar sensors.

The BIRIS system used in this multi-modal platform is based upon the active triangulation double aperture mask principle, which uses a special lens with two pinholes near the nodal point. This produces a double image of objects in the scene with a disparity that depends directly on the distance of the object from the focal plane of the lens. Therefore a stereo image is captured using only one CCD camera. When a projected laser stripe is emitted onto the scene, the target's depth can be readily computed. However, like all sensing technologies, the BIRIS sensor also has disadvantages. Its primary shortcoming is that although the accuracy of range measurements is reasonable on detected objects over short distances, the accuracy degrades rapidly with detected objects at longer distances. Furthermore, obtaining either a dense range image or range data over a wider field of view than the 25 degrees or so covered by a typical camera lens implies physically sweeping the camera and structured light pattern across the scene, and hence requiring a time delay.

The second modality of the QUADRIS perception system is comprised of a ring of 12 sonars positioned equidistantly around it such that redundant sonar data lessens the susceptibility to artefacts. However, multiple bounces or echoes associated with sonar data may give rise to errors. The first of these errors are phantom objects that produce a "third" wall when the system is near two walls forming a corner. A second typical error may be the false measurements that would indicate an object detected beyond the working range of detection. The latter of these two types of error is easily suppressed by eliminating objects that exceed a "logical" distance from the sensor or under the assumption that they are not reliable. In this system, detected objects that exceed the system limitation of one-meter radius are assumed to be errors and therefore discarded.

To improve the results obtained solely from the sonar system, the BIRIS system is judiciously called into play according to "just-in-time" sensing strategy. Results from the BIRIS provide additional criterion for eliminating errors in sonar data. When a line segment is detected by the BIRIS system, the assumption is made that there is no object present between the line segments of the wall to the position of the robot.

The judicial use of "just-in-time" BIRIS sensing is primarily used to complement sonar sensing. The data generated by BIRIS is only acquired when sonar data is ambiguous. To determine when sonar data is unreliable, the world environment is modeled into two categories: "terminal" for confirmed objects and "non-terminal" for objects to be confirmed. When the QUADRIS mobile robot approaches a wall, it attempts to navigate around the wall while increasing the length of the current line segment in the world map until it detects a non-terminal endpoint of the current line segment. An example of a non-terminal endpoint would be where two walls meet to form a corner. When the QUADRIS reaches such a non-terminal endpoint and detects inconsistencies BIRIS is then applied. As QUADRIS approaches, sonar data detects the "third" wall between the two real walls due to sonar reflections, hence a non-terminal unconfirmed object is detected and BIRIS is activated to scan the region. BIRIS then accurately maps the regions determining that there

are two walls and conclude that the ends of the two walls are two endpoints. QUADRIS then continues to map the new wall. Another situation where "just-in-time" BIRIS sensing would be required is for scanning physical extremity of a wall. The BIRIS would observe the endpoint of the object converting it from "non-terminal" points to the exact positions and marking the data as "terminal".

To correlate sonar and BIRIS data together, Dudek *et al.* have used incremental construction of dataset maps from each range sensing devices. Accounting for the position and the orientation of the line segments from the datasets, and determining their confidence measure as defined by Elfes *et al.* who model range sensing data by probabilistic grids [55], these dataset can be merged into a single dataset map.

This single dataset map is partitioned into three types of data according to the origin of the range sensor: *Sonar* segments originating from the sonar system, *Biris* segments originating from the BIRIS system, and *Complex*, for segments after merging *Sonar* and *Biris* line segments.

Two criteria exist for QUADRIS range data merging. The first criterion is based upon the distance threshold, where if two segments from *Sonar* and *Biris* are nearly parallel and do not exceed a distance threshold then they are merged immediately. If the two segments intersect each other at a point, then a *Complex* line segment is formulated such that it intersects the same point, but is given a slope with the largest confidence measure. If the two lines do not intersect, then their endpoints are identified and a calculated weighted midpoint using the same weights is used for the slope. This produces a new line segment passing through the weighted midpoint with a slope that reflects the largest confidence measure.

Overall the QUADRIS system describes an innovative approach in combining range data from two different sensor sources to efficiently take advantage of the characteristics that each type of device provides with exploring and mapping unknown terrain. The "just in time

sensing" provides a constrained used of the BIRIS sensor, which provides a "second opinion" when range datasets are insufficient or ambiguous.

3.3 Omnidirectional Stereo and Laser Range Finder

In another article by Miuri *et al.* [56], two range sensing technologies, passive stereo and laser range finder, are concurrently used for robot map generation. Each modality produces individual spatial grids of probabilistic regions of the location of obstacles and free space. These individual sets of data are combined to form a joint spatial map of known and unknown data, thus producing a reliable map.

The passive stereo system is built from a pair of vertically aligned omni-directional cameras that provide circular stereo images. Each image is "unrolled" to form stereo panoramic images where each vertical column becomes an epipolar line. These images are then used with conventional disparity algorithms to produce a panoramic disparity image. If the disparity is less than or equal to one pixel, the mapping for this region is considered as unused space. The laser range finder, which is based upon the TOF principle, is vertically aligned below the omni-directional cameras and scans a 180 degrees field of view. A single range map of a single horizontal plane is generated for each scan.

The main aspect that differentiates Miura *et al.*'s publication from other reported works is that it addresses a technique of merging the resultant data from two range sensing techniques together. A problematic situation could occur when each modality recognizes the same object differently. For example if each modality is to observe a table, the passive stereo method may observe the table top and the visible legs while the laser range sensing may only observe the legs. Immediately a discrepancy is detected between both modalities. If the passive stereo technique insists that an object exists within the grid and the laser range finder insists that the position is free space, an unbiased approach must be taken such that any obstacle reported by either modality results as an obstacle in the final mapping.

A probabilistic model is used to determine the reliability of both modalities. A forward sensor model, defined as a model that describes the physics of the environment, from causes to events, is used to determine the occupancy of objects within a certainty grid and is applied along with a Bayesian probabilistic approach, which is classically used by Elfes [57]. Each occupancy grid is characterized by three possible attributes: *occupied*, *free* and *unknown*, and is dependent on the reliability of the range sensor's capacity to detect an object at a grid point and represented as a probability value (0% - 100%). The *free* attribute is defined as the known free space that exists between the scanner and the object, the latter being defined as *occupied*. Occlusion, which can be identified by the stereo disparity algorithm, cannot provide a deterministic occupancy attribute, thus this region is defined as *unknown*. In addition regions of space that have not been defined or are beyond the limits of the laser range finder and stereo system field of view are also defined as *unknown* as shown in Figure 3-2.

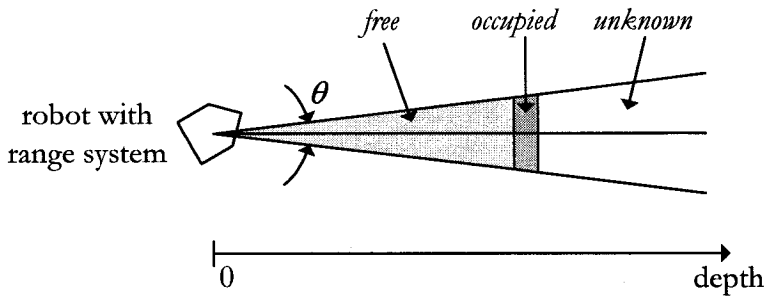

Figure 3-2 – Determination of grid attributes (reproduced from Miura *et al.* [57])

The occupancy probability is then defined by the Bayes theorem:

$$P(E \mid O) = \frac{P(O \mid E)P(E)}{P(O \mid E)P(E) + P(O \mid \overline{E})P(\overline{E})}$$
$$P(E \mid \overline{O}) = \frac{P(\overline{O} \mid E)P(E)}{P(\overline{O} \mid E)P(E) + P(\overline{O} \mid \overline{E})P(\overline{E})} \quad (54)$$

where O is an event where the obstacle is detected and its inverse $\overline{O}$ is an event where a *free* grid is detected. The variable E is denoted as an event where an obstacle exists and its inverse $\overline{E}$ is an event where an obstacle does not exist. Therefore, $P(E)$ and $P(\overline{E})$ represent the probability that an obstacle exists and does not exist respectively. The conditional probabilities $P(E \mid O)$ and $P(E \mid \overline{O})$ reflect the likelihood that an obstacle exists in an *occupied* and *free* grid space and are dependent on the probability model defined by $P(O \mid E)$ and $P(O \mid \overline{E})$.

These observational models for $P(O \mid E)$ and $P(O \mid \overline{E})$ are defined as the probability that an obstacle is observed when it actually exists or does not. In passive stereo, there is the likelihood that a correct match is not obtained due to environment changes or limitations in passive stereo where the object is too far away from the scanner. Thus a simple model is devised where the probability of an occupied space, $P(O \mid E)$, is inversely dependent upon the distance of this defined space. Hence the closer and larger the object appears, a higher probability is assigned to that grid space. Likewise, an object that is smaller or further away decreases the probability that the obstacle is observed. For scenarios where a false object is detected, $P(O \mid \overline{E})$, a low probability is assigned. Since the laser range finder provides fairly reliable measurements within its field of view and is not dependent on the distance or size of the obstacle, the observational model for $P(O \mid E)$ and $P(O \mid \overline{E})$ is set to a high and low probability respectively.

With the two probabilistic grid maps created from each modality, integration is applied to produce a final probabilistic grid. First a threshold is applied on the two grids to determine and categorize what grid sample is designated as *free* space (low probability), *occupied* (high probability), or *undecided* (mid-range probability). For example, if the probability of a grid is larger than 0.7 then the grid is characterized as an *obstacle*, if the probability of a grid is below 0.3 then the grid is characterized as *free space*, and in between these thresholds the grid is

characterized as *undecided.* The two probabilistic grid maps now become two deterministic grid maps and can be easily merged together to produce a final grid map of only defined *obstacle* and *free space* entities as computed in Table 3.1.

		Passive Stereo		
		Obstacle	*Unknown*	*Free Space*
Laser Range Finder	*Obstacle*	Obstacle	Obstacle	Obstacle
	Unknown	Obstacle	Obstacle	Free Space
	Free Space	Obstacle	Free Space	Free Space

Table 3.1 – Integration of probability occupancy grid (reproduced from Miura *et al.* [57])

3.4 Concepts and Design Considerations – Proposed Heterogeneous Multi-Modal Sensing Strategy

The term "multi-modal" has not been widely used in the field of range sensing. However different implementations of multi-modal systems have been built to integrate range-sensing datasets in hopes to improve modeling of a scanned region. For example, the three examples of range sensing techniques presented in the previous section are all working examples of multi-modal range sensors. Each system delivers a unique approach of integrating diverse range sensing equipment together and integrating complementary data to create a final model of the scanned scene.

From a high level view, multi-modal systems can be defined by their multiple and diverse modes of range sensing used to perceive scenery. Systems that use multiple yet identical modes are not considered multi-modal since only a single mode is used. With this mind, the coined definitions of active and passive range sensing [8, 9] can be extended for the purpose of multi-modal systems into *homogeneous sensing* and *heterogeneous sensing.* A homogeneous sensing system is defined as the application of range sensing technologies that are built from all active or all passive subsystems. Likewise, heterogeneous sensing systems use both active and passive subsystems in tandem.

A working example of a multi-modal homogeneous range sensing system is the application of two active range sensing systems: a laser range finder and a sonar/acoustic sensing system as proposed by Laurent *et al.* [53] and Dudek *et al.* [54]. With these two methods, the laser range finder and sonar system sample separately without knowledge of each other's extrinsic locations. Once the individual scans have been completed by each subsystem the two datasets are merged together to provide a single map of the environment. The success of this multi-modal system is dependent upon the environment in which the system operates. For example if the environment is simply a maze where walls are the only objects, the system performs admirably. However, in a complex environment where there are objects of varying height there is no mechanism that correlates the measurement of what the laser range finder perceives to that of the sonar system.

A common example of a multi-modal heterogeneous range sensing system is that of the omnidirectional stereo and a rotating laser range finder system, as proposed by Miura *et al.* [56]. In these systems, a passive sensor, omnidirectional stereo, is merged with the active laser range finder. The complexity of merging the datasets of both range sensing technologies is clearly defined by Miura *et al.*, who unlike Laurent *et al.* [53] and Dudek *et al.* [54] outline the dilemma of different possible perceptions of an object. As a solution, Miura *et al.* propose the use of probabilistic grids to aid in classifying each subsystem based upon their strengths, weaknesses and limitations, which are all important factors when merging datasets.

Inspired by these various approaches, a different heterogeneous sensor named VIVA M^2S-SSL (**M**ulti-**M**odal **S**ensor – **S**tereovision, **S**tructured lighting, **L**aser range finder) is introduced, which combines three active range sensing systems (one laser range finder system and two structured lighting systems) and one passive range sensing system (stereovision system), as shown in Figure 3-3. The key aspect considered in the design of the proposed multi-modal system is to provide complementary data, instead of having the system able to cover different range depths. It is expected that a combination of these three common range sensing technologies will provide a high robustness to various environments where lighting,

textures, reflectivity and other characteristics influencing range data collection might vary considerably.

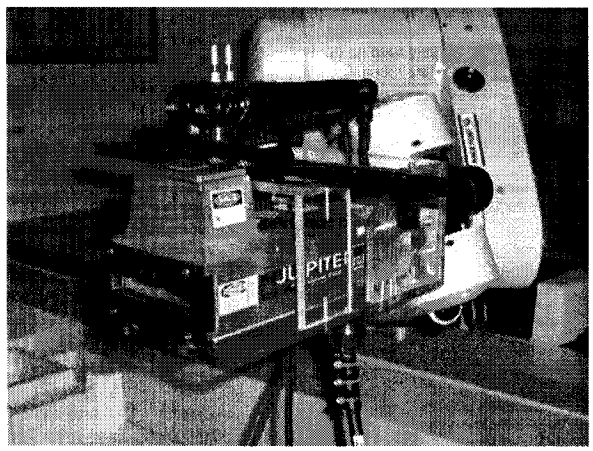

Figure 3-3 – Multi-modal chassis attached to robotic end-effector

We propose a heterogeneous multi-modal range system that combines the strengths of Miura *et al.* where the dilemma of different possible perceptions of an object is considered with those of Laurent *et al.* [53] and Dudek *et al.* [54] where feature extraction can be used to correlate one sensing system data to another. However, unlike these three multi-modal approaches, we introduce a calibration process that minimizes the use of run-time feature extraction. Feature extraction becomes an essential dynamic to the calibration process that eliminates the future "guess work" of feature correlation used to determine which features from one system matches the other.

Calibration of range sensing system is completely unavoidable and at times a lengthy process. For example, in stereovision and structured lighting systems where the user has the ability to position the structured light projector or cameras in different orientations to each other, a

calibration procedure is required for any positional change between the cameras or light projector to ensure consistent three-dimensional reconstruction. Laser range finders, such as those that use active triangulation principles also require calibration, but are typically enclosed systems where laser emitter and detector are in fixed positions to each other. Thus a single calibration procedure at factory is most likely performed. The following sections describe the physical design and construction of the prototype multi-modal sensor and the calibration procedure that has been developed in this thesis work to make it fully operational.

3.4.1 Proposed Multi-modal Concept

The proposed multi-modal range sensing system consists of four subsystems combining three different range sensing techniques. The first subsystem is a laser range finder, which provides two-dimensional data along a scan-line marked by a visible red line projected on the scene. The second subsystem is a stereovision system built from two CCD cameras mounted in close proximity to the laser range finder. The third and fourth subsystems are structured lighting systems that use the left and right stereovision cameras independently to detect the projected structured light emitted from the laser range finder. Figure 3-4 outlines the modalities used in the proposed multi-modal range sensing system.

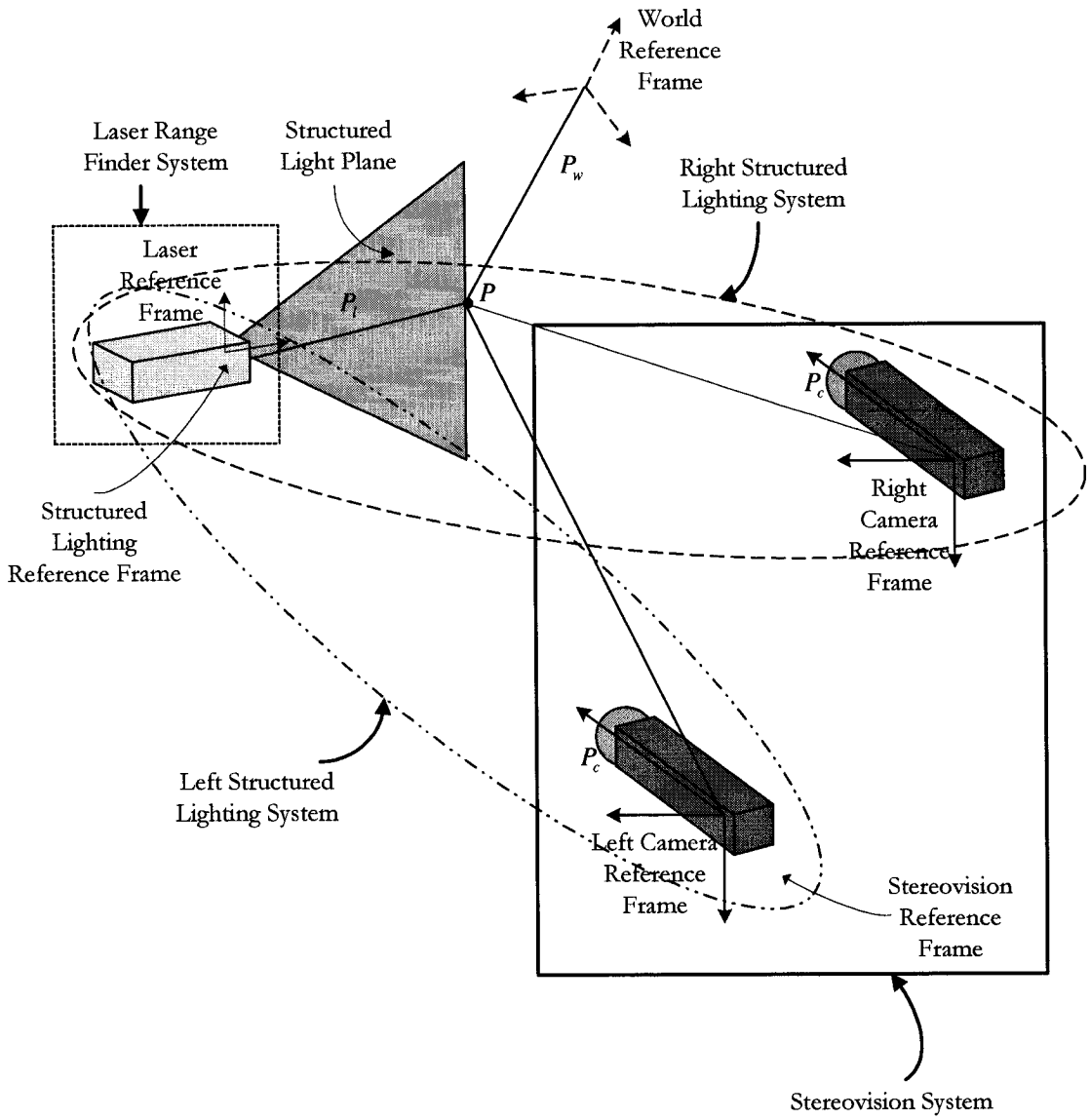

Figure 3-4 – Perceptions from individual modalities in the multi-modal range sensor

Three major design steps are defined for constructing a multi-modal system that can provide effective range sensing data for three-dimensional reconstruction:

Step 1: Define a multi-modal infrastructure and strategically place sensing systems where correlation between systems can be achieved. This first step is crucial since it defines a rigid physical infrastructure between each modality that remains constant. This first step is commonly used in multi-modal design as seen in the works by Laurent *et al.* and Miura *et al.* where strict guidelines are imposed to have the camera system and the laser range finder placed vertically aligned. The motive of this alignment is to ensure that systems monitor a shared field of view region and thus collaborative sensing is achieved. The same must be applied to our proposed multi-modal range sensor, where each modality must monitor a shared field of view region and also must be placed strategically to do so. Stereovision and laser range finder have no particular stringent configuration demands, just as long as the field of view is shared. Therefore the only requirement between these two modalities is that they both scan within a shared field of view as shown in Figure 3-5 and Figure 3-6.

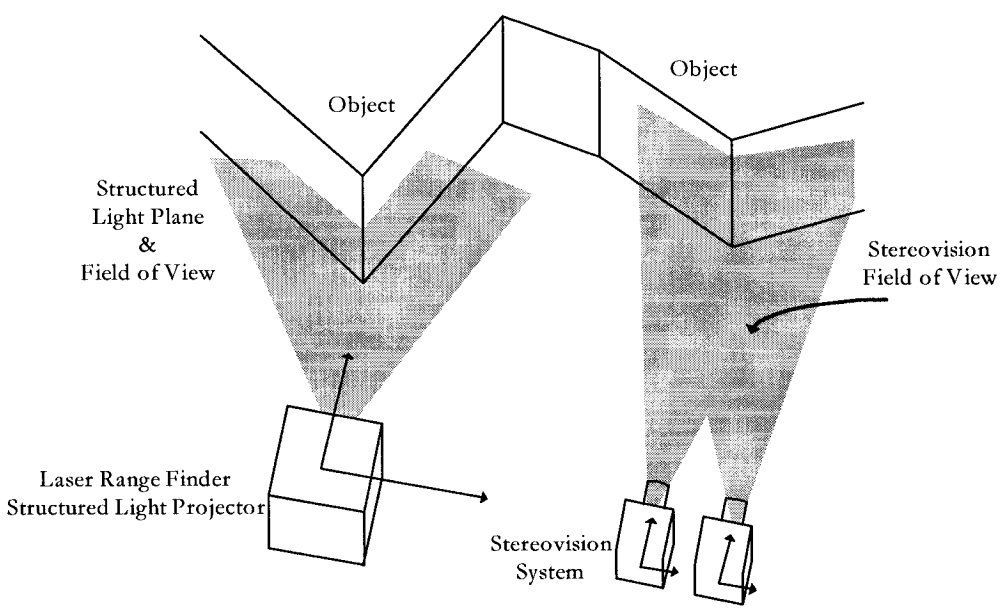

Figure 3-5 – Laser range finder and stereovision unable to produce joint data due to non-shareable field of view

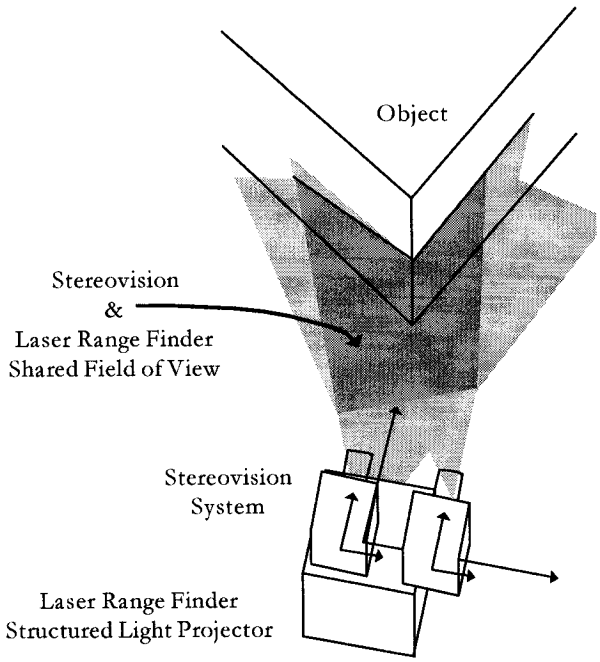

Figure 3-6 – Laser range finder and stereovision sharing field of view

A quick assessment of the range sensing technologies indicates that in structured lighting the camera must also be placed in a shared field of view with the laser range finder. More specifically, the camera must be positioned in non-coplanar position to the laser range finder structured light. Structured lighting is highly dependent upon the location of the laser pattern detected in the image. If the cameras were placed co-planar to the projected line strip, the detected line in the system would remain fixed regardless of the depth of the object in the path of the structured light, as shown in Figure 3-7 and Figure 3-8.

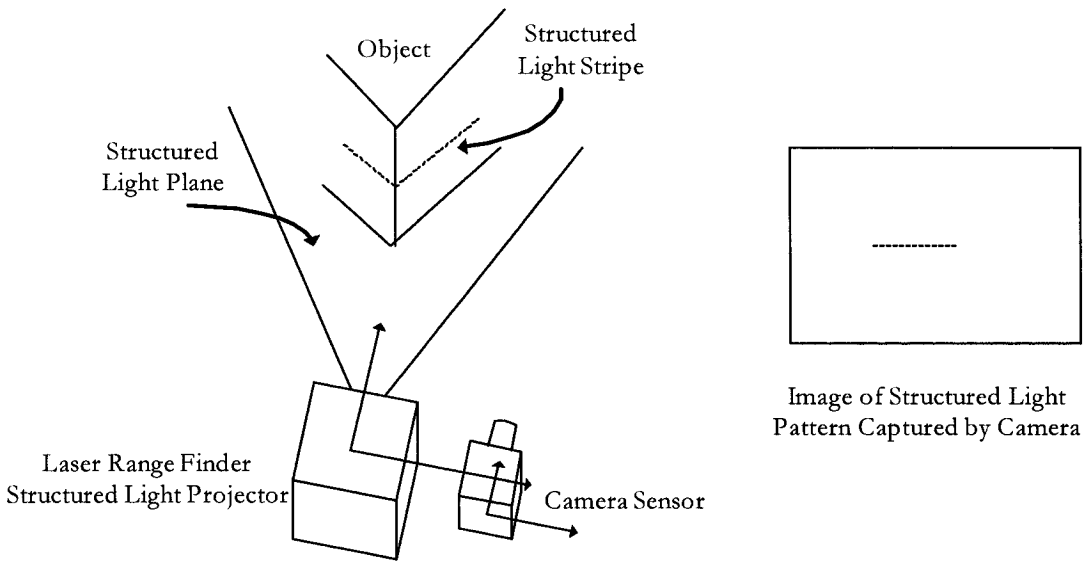

Figure 3-7 – Structured lighting system with camera placed coplanar to the structured light plane

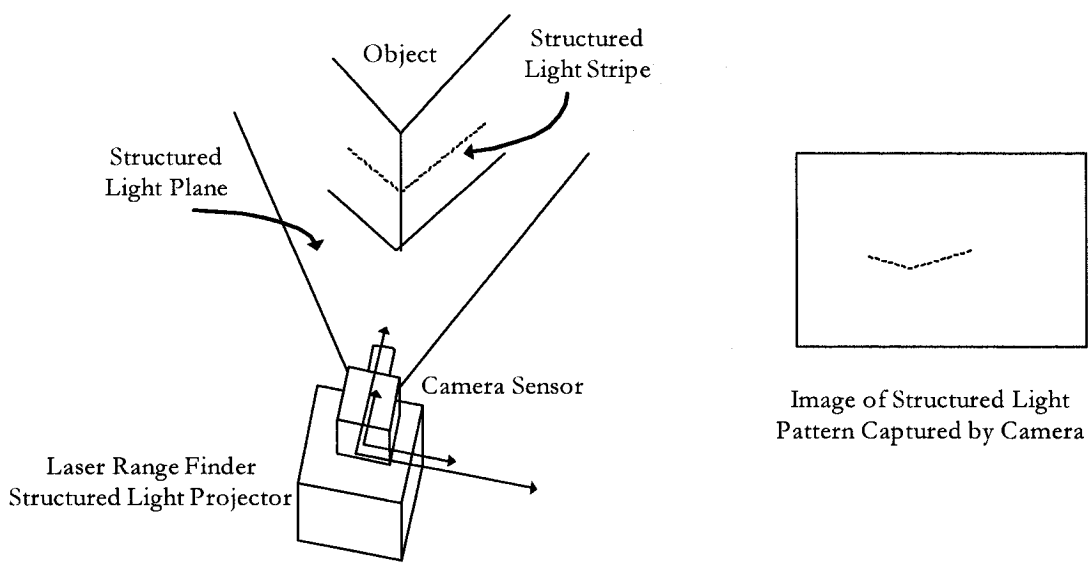

Figure 3-8 – Structured lighting system with camera placed non-coplanar to the structured light plane

This would make it impossible for the structured lighting system to extract range data. For the stereovision system, the cameras must have an appropriate baseline, selected such that objects are visible from both cameras, and disparity is detectable between the acquired images.

Step 2: Define a common calibration mechanism that is cost-effective, uses minimal workspace and is usable in automated calibration. The design of a multi-modal system must also define a calibration mechanism that may be used for each subsystem. Unlike the approaches from Miura *et al.* and Laurent *et al.*, where collaborative sensing data requires data-fit post-processing, the use of calibration will reduce or eliminate the use of guesswork and post-processing. Evaluating the proposed multi-modal system has determined that various types of calibration approaches can be taken to achieve calibration. However, two main features of this step are desired: firstly the use of automated calibration and the minimal dependency on the user intervention and secondly the use of well-known calibration algorithms that have proven to be effective and known to work. The classical Tsai's camera calibration technique [14] used to determine intrinsic camera parameters and extended for extrinsic stereovision calibration has been the mainstream accepted standard used for stereovision calibration. Structured lighting calibration has gone through many permutations, and among the best methods are the direct calibration approach [46] which requires minimal mathematical processing yet potentially requires large amounts of stored data, and Chen and Kak's projective calibration approach [47] for structured lighting systems which requires a transformation matrix to be calculated. Each calibration technique is suited for the proposed multi-modal system such that automated calibration is easily achieved. However one particular troublesome aspect is the direct calibration target defined in the direct calibration by Trucco *et al.*, which uses a staircase-like target that requires the structured light stripe pattern to be aligned at each step [46]. The projective structured lighting calibration approach is far more flexible as discussed in Section 2.3.5 with the only constraint that the structured lighting pattern intersects the calibration target edges to generate a visible and extractable feature.

The latter half of this step requires that the designer consider defining a workspace area that can be used for calibration purposes and normal operation. For example the direct calibration system by Trucco *et al.* is dependent on a moving track where the calibration target was placed. The structured lighting system remains fixed in the world reference frame and objects are placed on a conveyor belt track and pass in the field of view of the system. If the structured lighting system requires calibration, user intervention is required to place the aligned calibration target onto the track, thus defeating the purpose of automated calibration.

For the proposed stereovision and projective structured lighting calibration, a design approach has been introduced that facilitates automated calibration. This is accomplished by designating a calibration space and a normal operation space within the robotic workspace. The calibration workspace is solely used to calibrate the multi-modal system when calibration is required either when inconsistencies are detected, failure to reconstruct large number of feature points, or on a periodic basis. Since this space is solely dedicated for calibration, a calibration target can be placed within the space, without fear that it will be detected during normal operation. Likewise, the normal operation space is used only for object and scene reconstruction. To move the multi-modal system between these two designated space is possible due to the flexibility of the two systems attached to a robot end-effector, where the system can freely move within a bounded area in either calibration or normal operation space as shown in Figure 3-9. In comparison to the implementation by Trucco *et al.* automated calibration can be easily achieved by having the multi-modal system moved into the calibration space and its sampling process started. This arrangement is much preferred than a system fixed in a permanent position and waiting for the calibration target to be passed within its field of view by an operator.

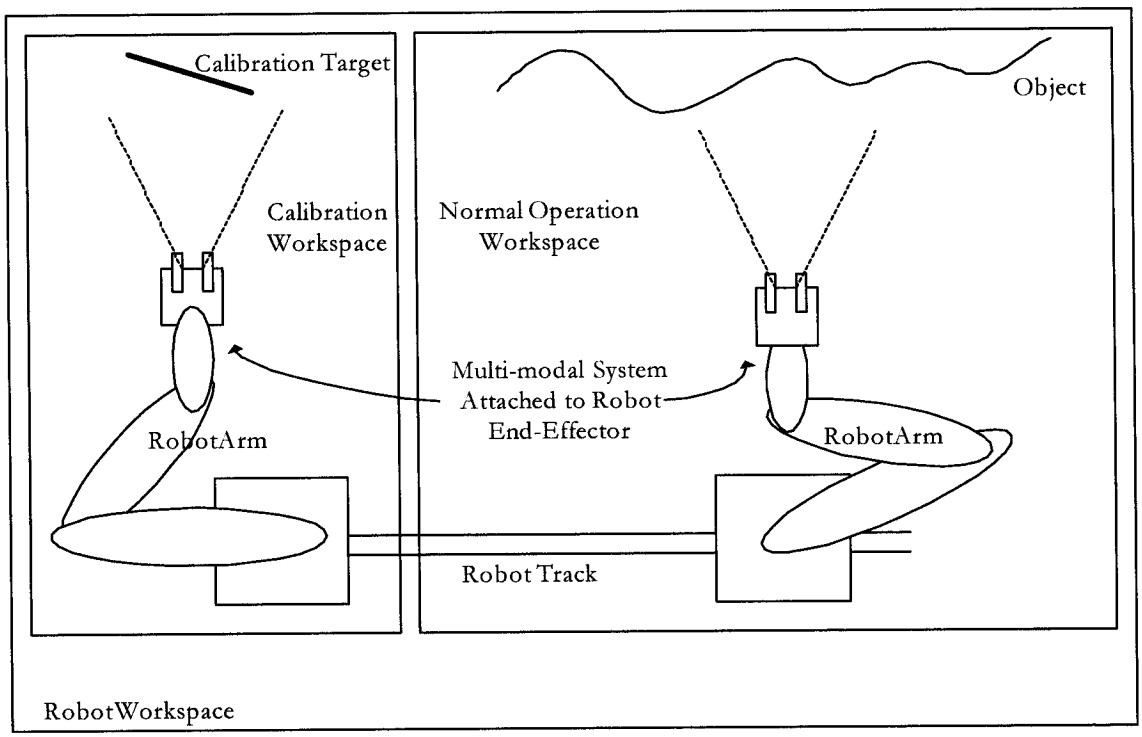

Figure 3-9 – Dedicated workspace for calibration and normal operation

Step 3: Determine a global frame of reference with respect to which all three-dimensional datasets are defined. Each modality employs its proper reference frame to define the three-dimensional world coordinates. Since one key objective is to reduce the amount of post-processing after acquisition, it is desirable that the various multi-modal data models be represented from a common reference frame. Whether this frame is a world reference frame within the workspace or a reference frame from a single modality, additional extrinsic calibration parameters must be defined to achieve this association. Determining where this common reference frame should reside can be achieved by identifying a reference frame that is shared by the majority of sensing systems.

This however is not the case for the multi-modal range sensor. Stereovision and laser range finder modalities both have their own frame of reference located in their respective optical centres. Structured lighting systems using the Chen and Kak's projective calibration approach correlate three-dimensional world points on the world reference frame that defines the edges of the calibration target. Assuming that both structured lighting subsystems use the same geometric properties as the structured light calibration target, hence using the identical world frame of reference, there is a total of three difference frames of reference within the proposed multi-modal range sensor, each sharing a common field of view, but generating different three-dimensional results.

To successfully complete this step to define a common reference frame and transform each three-dimensional model into a common frame of reference, we propose in Section 3.5 a modified projective structured lighting calibration approach that advantageously uses the laser range finder reconstruction precision as well an extrinsic calibration procedure that converts reconstructed stereovision to the laser range finder frame of reference. The end result is a flexible structured lighting system that can report depth information from the laser range finder frame of reference, eliminating the need to calculate additional extrinsic parameters. In addition, the stereovision system will be capable of transforming reconstructed three-dimensional data into two-dimensional data from the laser range finder frame of reference using an innovative extrinsic calibration closed-form solution approach proposed by Pless and Zhang [64].

In Figure 3-10, the steps required for multi-modal construction are illustrated where each shape is considered as a range sensing modality to be assembled in a multi-modal range sensing system.

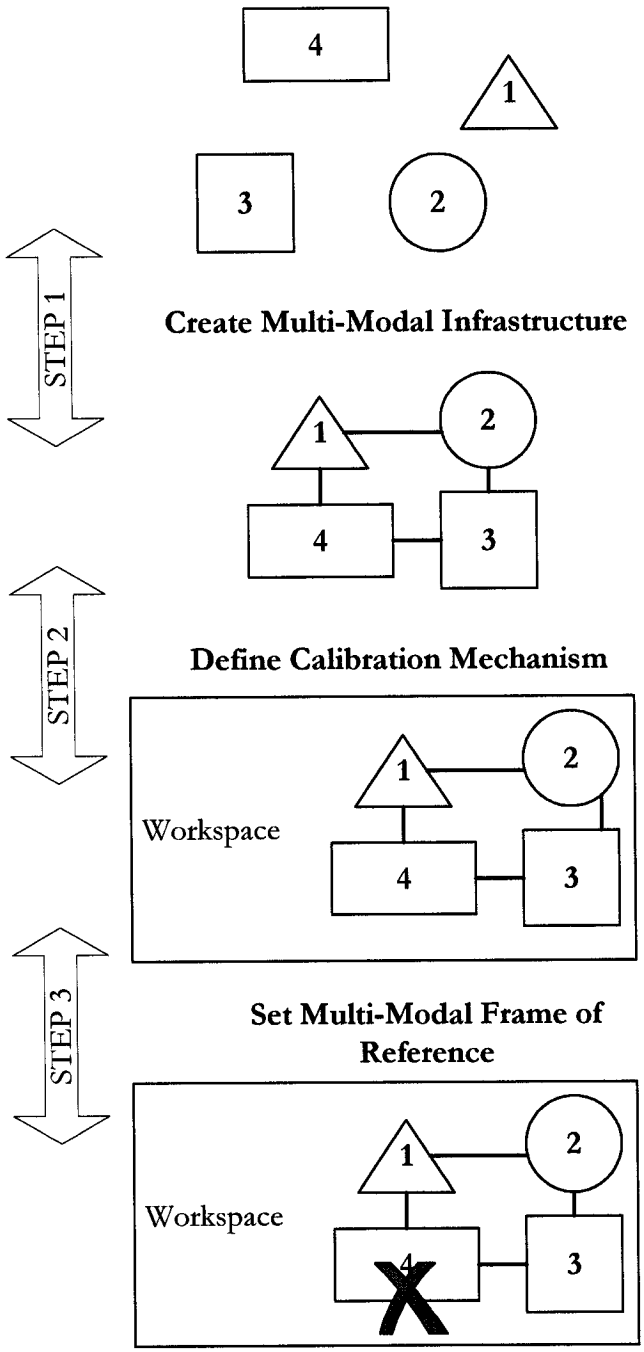

Figure 3-10 – Step layout for multi-modal construction

3.4.2 Physical System Implementation

The Jupiter laser range finder, manufactured by Servo-Robot Inc, is operating upon the auto-synchronous active triangulation principle and is capable of acquiring a maximum of 512 sampling points per scan through the emission of a single 150mW laser spot oscillated by a two-sided mirror. The field of view depth ranges from 0.3 to one metre, which is not atypical for this type of active triangulation laser range finder device. Control of the Jupiter is handled by the Servo-Robot Cami-Box and offers an RS-232 asynchronous link to an external interface. A feature of interest is the frequency scan rate that controls the oscillating period of the two-sided mirror. At low frequencies, the projection of the laser beam can be seen as a slow moving dot moving back and forth across the scene. This setting allows for high precision sampling of the scene. At higher frequencies, the movement of the projected spot becomes much faster and to the eye the appearance of a laser stripe is projected on the scene. This stripe becomes a visual marker allowing the user to recognize what area of the scene is currently being scanned.

A pair of Sony XC-999 CCD cameras, the key component of the VRex CAM-3000C [58] product is used as the stereovision system as shown in Figure 3-3. These cameras are mounted on a metal sliding bracket that allows a maximum 30-centimetre baseline and can effortlessly encompass the same field of view and depth as the Jupiter laser range finder [59, 60] by adjusting the distance between the cameras along the bracket. To synchronize stereovision image acquisition, the deployment of the VRex VRMUX2N [58] is coupled to a Matrox Orion video card [61] permitting real-time image sensing. The VRex VRMUX2N unit ensures simultaneous stereo image acquisition within a period based on an internal or external clock source. This eliminates the necessity of constructing software for synchronization between individual cameras and eases the processing required for automation purposes. One particular interest of VRMUX2N design is its ability to multiplex either colour or black and white stereovision camera signals into a single interlaced video feed [58]. For video and image processing, a Matrox Orion video frame grabber easily de-interlaces multiplexed feed using its on-board processor thus alleviating consumption of resources by

other real-time processes. The de-interlaced processes produce two raw 320×240 pixels dimensioned images.

When the Jupiter is configured for high frequency sampling, the structured spot of light becomes a structured stripe of light visible to the stereovision cameras. This structured stripe provides a third range sensing technique based on structured lighting. In addition, since two cameras are required for stereovision, each camera can act as an individual structured lighting system, increasing the total number of range sensors in the multi-modal system to four.

To maintain a fixed positional relationship between the CAM-3000C stereovision system and the Jupiter laser range finder, the two systems are held together by a custom-made Plexiglas chassis as shown in Figure 3-3. The chassis is then mounted to an end-effector of a six degree-of-freedom [62] F3 manipulator installed on a CRS two-metre track manufactured by CRS Robotics as shown in Figure 3-11. To control the end-effector, a CRS C500C digital programmable controller communicates via an in-house driver through an RS-232 asynchronous link built from the RAPL-3 language, which provides an interface to high application software by translating requests for translations and rotations [7].

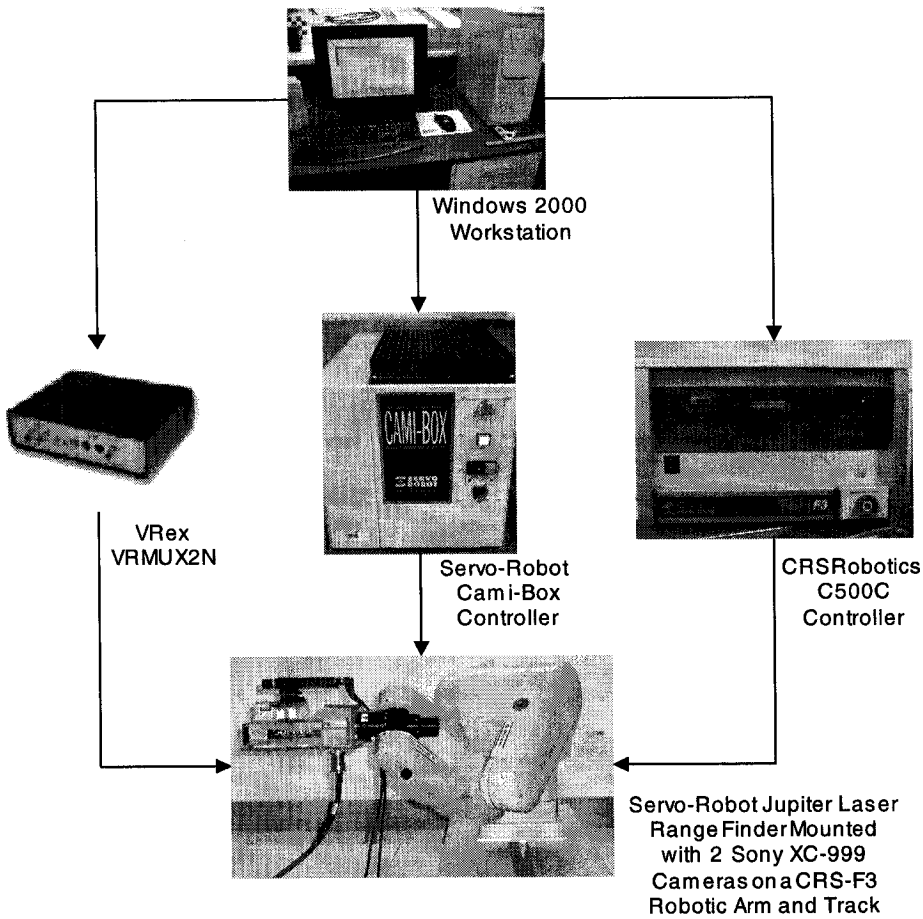

Figure 3-11 – VIVA M²S-SSL robotic integrated system

3.5 Multi-modal Calibration

Although there are many different approaches to system calibration for both intrinsic and extrinsic properties, the classical technique of Tsai's camera calibration model for stereovision calibration [14, 63], a refined version of Chen and Kak's structured lighting subsystem calibration [47], and Pless and Zhang's closed-form solution for camera to laser range finder

relationship [64] were selected. These methods are combined in an original way in order to minimize manipulation and data collection.

Chen and Kak's model of structured lighting system calibration could be replaced by other well-known calibration methods as noticed by Trucco and Fisher [46]. However, Chen and Kak provide the interesting idea of using a simple calibration target, which relies upon the movement of a robot end-effector carrying the laser projector and camera. The use of two-dimensional projectivity theorem as introduced in Section 2.3.3.2 to derive a transformation matrix that converts detected structured light emitted on defined edges in the world space provides a simpler calibration approach that can be easily used in automated calibration.

The closed-form solution defined by Pless and Zhang to relate a laser range finder and camera system [64] is appropriate for inter-subsystem calibration between the laser range finder and the stereo system. But instead of using a planar pattern placed in different poses from the camera and laser range finder, the same calibration target used for the extended Chen and Kak calibration procedure is exercised such that the stereo system can detect the defined edges of a known object in the world space.

The largest difficulty in designing and operating a multi-modal system is the organization required to calibrate each subsystem and interoperating subsystems. For the scenario of using a laser range finder, stereovision, and two structured lighting systems, calibration can seem more problematic. However, with a systematic approach and the reduction of transformations, we propose a resilient automated calibration mechanism.

For the proposed combination of sensing technologies, advantage can be taken of widely known calibration procedures such as Tsai's camera calibration technique [14] or internal self-calibration functionalities integrated in device controllers. In the present case, Tsai's approach is used to calibrate individual CCD cameras and to determine intrinsic and extrinsic parameters between the pair of camera reference frames. Likewise, the Jupiter laser range finder used in the experimental setup has a fixed internal calibration provided by the

manufacturer [60]. The original part of the proposed calibration procedure is related with the estimation of the registration for the dual structured lighting system and the extrinsic calibration between the stereovision pair and the laser range finder. These aspects are detailed in the following sections.

3.5.1 Structured Lighting Subsystem Calibration Revisited

Since stereovision and the structured lighting from the laser range finder are available, the visible trace of the laser on the scene is exploited to create two structured lighting subsystems, one for each CCD camera. The proposed calibration mechanism for these subsystems is based upon Chen and Kak's structured lighting calibration technique [47] where a recovery conversion matrix can be built from a minimal set of four known points on the surface of a given object as described in Section 2.3.3.2 [47,52]. The equations defined in eq. (48), eq. (49), and eq. (50) illustrate that the transformation matrix defined by line vectors of the projection matrix T_1, T_2, T_3, and T_4 need to be determined.

Chen and Kak propose an acquisition process that requires the projected structured light pattern to fall on known points as described in Sections 2.3.4 and 2.3.5. This acquisition process, which is easily automated, also requires that a simple object, the calibration target, be placed in the path of the structured light such that its edges would form discontinuities in the structured light and would be detectable by the camera system. The main limitation of Chen and Kak's method is the requirement that the 3-D edge points of the calibration target need to be fully known and characterized by the intersection of planes defined in the world coordinates frame before being substituted in eq. (48), eq. (49), and eq. (50). The structured lighting system is then positioned at different locations by movement of a robotic arm such that the mounted system moves away from the calibration target and edges are extracted from the images such that U is sampled. A minimal set of four unique coplanar but non-collinear

calibration points is required to produce a set of twelve linear equations to solve for twelve coefficients that define the calibration matrix.

By expanding eq. (39) and isolating the world coordinate points a set of linear equations based upon the number of sampled points used for calibration generates the conversion matrix. The expansion of eq. (39) for N samples becomes:

$$\begin{aligned} t_{11} \cdot u_1 + t_{12} \cdot v_1 + t_{13} &= \rho x_1 \\ t_{11} \cdot u_2 + t_{12} \cdot v_2 + t_{13} &= \rho x_2 \\ &\vdots \\ t_{11} \cdot u_N + t_{12} \cdot v_N + t_{13} &= \rho x_N \end{aligned} \tag{55}$$

$$\begin{aligned} t_{21} \cdot u_1 + t_{22} \cdot v_1 + t_{23} &= \rho y_1 \\ t_{21} \cdot u_2 + t_{22} \cdot v_2 + t_{23} &= \rho y_2 \\ &\vdots \\ t_{21} \cdot u_N + t_{22} \cdot v_N + t_{23} &= \rho y_N \end{aligned} \tag{56}$$

$$\begin{aligned} t_{31} \cdot u_1 + t_{32} \cdot v_1 + t_{33} &= \rho z_1 \\ t_{31} \cdot u_2 + t_{32} \cdot v_2 + t_{33} &= \rho z_2 \\ &\vdots \\ t_{31} \cdot u_N + t_{32} \cdot v_N + t_{33} &= \rho z_N \end{aligned} \tag{57}$$

$$\begin{aligned} t_{41} \cdot u_1 + t_{42} \cdot v_1 + 1 &= \rho \\ t_{41} \cdot u_2 + t_{42} \cdot v_2 + 1 &= \rho \\ &\vdots \\ t_{41} \cdot u_N + t_{42} \cdot v_N + 1 &= \rho \end{aligned} \tag{58}$$

When substituted into eq. (48), eq. (49), and eq. (50), the end matrix is formed such that:

$$Aq = b \tag{59}$$

where:

$$A = \begin{bmatrix}
u_1 & v_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -u_1 \cdot x_1 & -v_1 \cdot x_1 \\
u_2 & v_2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -u_2 \cdot x_2 & -v_2 \cdot x_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
u_N & v_N & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -u_N \cdot x_N & -v_N \cdot x_N \\
0 & 0 & 0 & u_1 & v_1 & 1 & 0 & 0 & 0 & -u_1 \cdot y_1 & -v_1 \cdot y_1 \\
0 & 0 & 0 & u_2 & v_2 & 1 & 0 & 0 & 0 & -u_2 \cdot y_2 & -v_2 \cdot y_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & u_N & v_N & 1 & 0 & 0 & 0 & -u_N \cdot y_N & -v_N \cdot y_N \\
0 & 0 & 0 & 0 & 0 & 0 & u_1 & v_1 & 1 & -u_1 \cdot z_1 & -v_1 \cdot z_1 \\
0 & 0 & 0 & 0 & 0 & 0 & u_2 & v_2 & 1 & -u_2 \cdot z_2 & -v_2 \cdot z_2 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & u_N & v_N & 1 & -u_N \cdot z_N & -v_N \cdot z_N
\end{bmatrix}$$

$$q^{\mathsf{T}} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{21} & t_{22} & t_{23} & t_{31} & t_{32} & t_{33} & t_{41} & t_{42} \end{bmatrix}$$
$$b^{\mathsf{T}} = \begin{bmatrix} x_1 & x_2 & \cdots & x_N & y_1 & y_2 & \cdots & y_N & z_1 & z_2 & \cdots & z_N \end{bmatrix}$$

which can be solved for q. Theoretically the minimal number of calibration points required is four coplanar but non-collinear points to fulfill the fundamental theorem of two-dimensional projectivity as described in Section 2.3.3.2.

3.5.2 Structured Lighting Calibration and Acquisition

The actual calibration process as defined by Chen and Kak's procedure is designed such that only a single active triangulation range sensing system is to be calibrated. There are two constraints required for valid calibration. First a world frame of reference must be defined in the workspace, such that the geometric properties of the calibration target can be defined. Secondly, the intersecting planar equations defining the edge properties of the calibration target must be exact to the actual calibration target and vice versa. We propose a method of extracting structured lighting calibration samples without direct knowledge of these two geometric constraints.

One key feature that is inherited by Chen and Kak's calibration procedure is the use of an object that can form a discernable alteration on the structured lighting pattern. Originally, the calibration procedure described the use of an object whose edge can be defined in the robotic workspace and is represented as an intersection between two planes. As the structured lighting system projection intersects the edge, an expected alteration of the pattern is formed such that the camera system is capable of determining the point of intersection within its image. What we propose is the use of a triangular planar shaped calibration target that is easily constructed from a piece of cardboard and whose face is placed directly in front the path of the striped structured light, as shown in Figure 3-12.

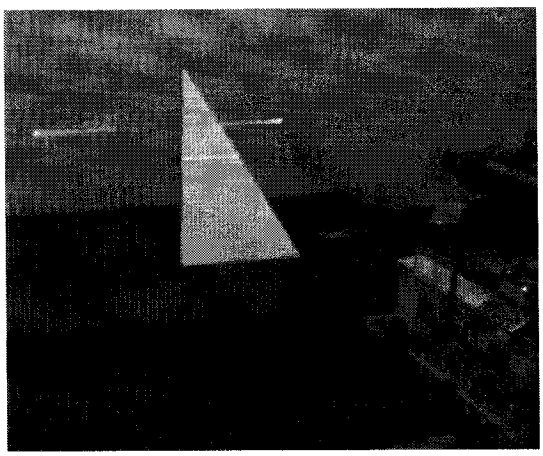

Figure 3-12 – Multi-modal system calibration

Using the simple triangular shaped calibration target placed in front of a planar background, the structured light pattern is extracted by determining structured light segments emitted on the target and on the background. The striped light from the laser projector emits a discernable red light and is easily detectable by a CCD camera when projected on a non-red coloured object. A simple filter, either by applying a threshold to the image or by filtering for intensity gradient changes, can be used to isolate the structured light pattern in the image. This produces an image similar to that shown in Figure 3-13, which outlines the structured light pattern projected on the calibration target and on the background.

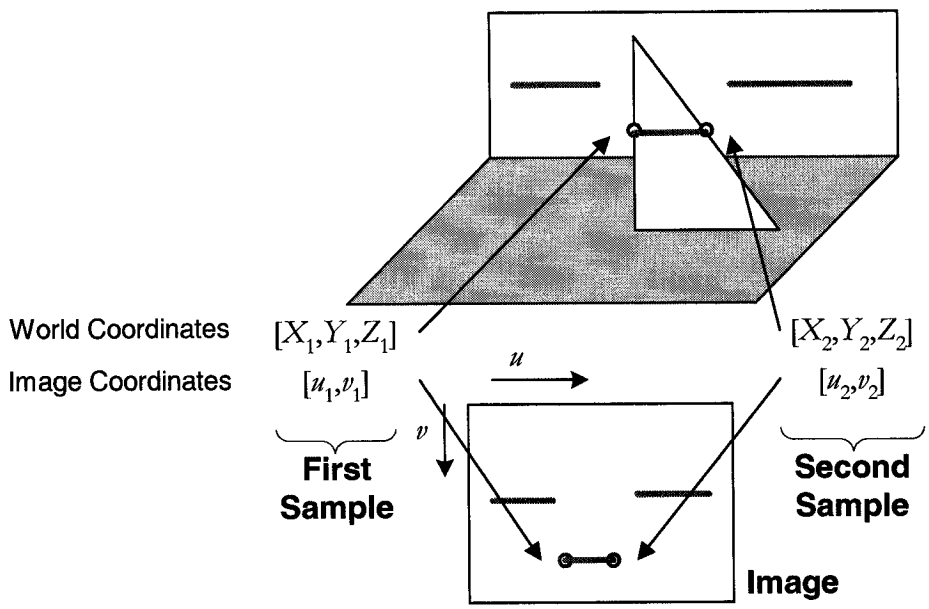

Figure 3-13 – Structured light sampling

To validate whether the structured light pattern from the image was correctly filtered a pattern is expected from the filtered image. The start and end of the structured light pattern is valid only when the following pattern is detected: background light segment – foreground light segment – background light segment as shown in Figure 3-12 and Figure 3-13. Although the background segments do not contribute to the sampling data, what it indicates is that the structured light stripe has crossed the width of the calibration of the target. This guarantees that the endpoints of the foreground light segment in the image are the edges of the calibration target in the image, in contrary to Figure 3-14. Both left and right structured lighting systems (based on stereovision cameras) sample the same feature points produced by the calibration target. In addition, since the structured light pattern is the by-product of the auto-synchronous active triangulation laser range finder, a similar pattern of foreground and background objects should be reflected in its scanned results. The laser range finder would distinguish a discontinuity of the distance along the edges of the calibration target.

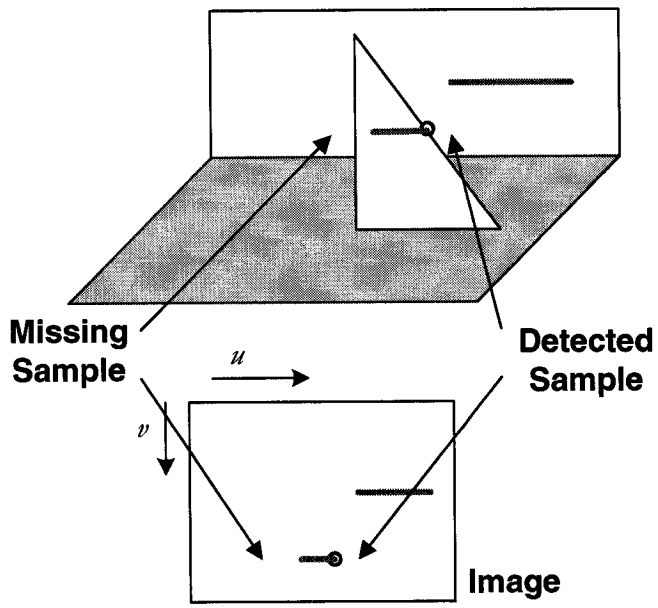

Figure 3-14 – Invalid structured lighting calibration sample - missing sample

Note that the specific shape of the calibration target is not required to be triangular. Any object that produces two discontinuities of the striped structured light can be used as a calibration target. The only restriction is that for the projective matrix to converge, unique non-collinear calibration points must be sampled. The benefits of using a triangular shaped calibration target as opposed to a rectangular object, which can also produce a similar discontinuous light pattern, comes into effect when three background light segments are detected. This phenomenon occurs when the calibration target itself occludes partially one of the background light segment patterns as shown in Figure 3-15. To eliminate the "third" structured light segment the structured lighting system can be lowered such that it deliberately occludes this segment from the camera's view.

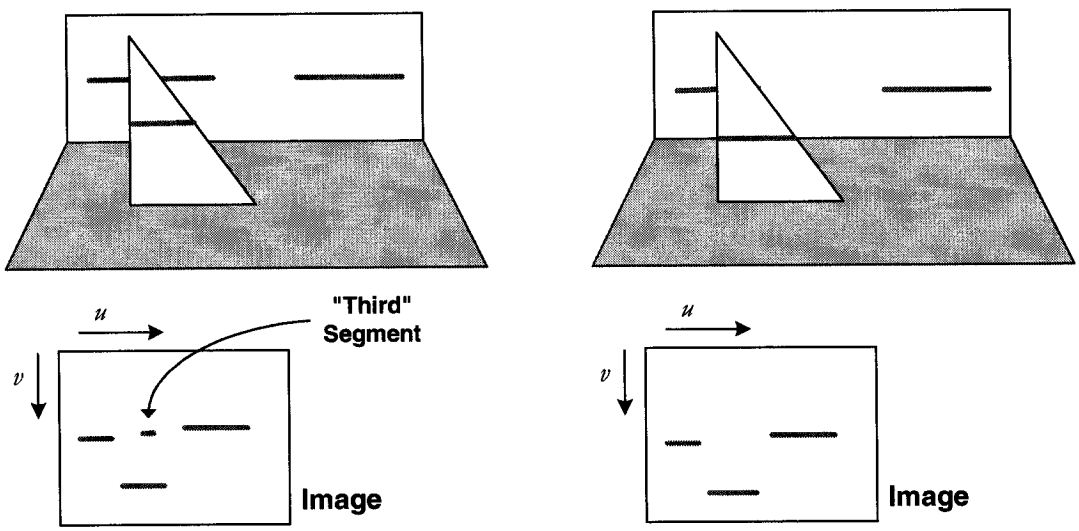

Figure 3-15 – Invalid structured lighting calibration sample – "third" segment and its effect by lowering the structured lighting system

To reiterate the requirements for completion of the series of linear equations defined in eq. (59) that are used to determine the structured lighting projective matrix, two components are required: image coordinates of the intersection between the structured light stripe projection and the calibration target and its world coordinate. Instead of determining the geometric intersecting planes within the world coordinate frame as defined by Chen and Kak, these inaccurate manual measurements are replaced with high-resolution samplings from the auto-synchronized active triangulation laser range finder $[X_i, Y_i, Z_i]$. By extracting the edge points that the structured light plane forms with the calibration target, $[u_i, v_i]$, both the laser range finder and the camera systems can easily detect matching discontinuities in the line scan. As a result, this adaptive method of calibration requires no *a priori* knowledge of the calibration target and of its position in space.

These modifications to Chen and Kak's approach not only reduce error coming from human intervention but also simplify the procedure by eliminating the need to determine the complicated intersecting planes defining the edges of the calibration target. Full advantage is

taken of the multi-modality of the sensing system, even during the calibration phase. In addition, this adaptation holds the ability to directly estimate three-dimensional coordinates with respect to the laser range finder reference frame.

In order for the projectivity matrix to converge consistently, a higher number of calibration samples would minimize some error criterion by forming an over-determined system of linear equations. Figure 3-16 shows that, using the revisited Chen and Kak approach, the maximum percentage error of reconstructed features gradually decreases over the increase of calibration samples. In this figure five calibration sets containing respectively 6, 10, 16, 22, 30, 60, and 106 samples were used to sample an *a priori* known object placed in the center of the shared field of view. The maximum error achieved from each calibration set is displayed. Linear interpolation between these error levels is used to highlight the trend on the effect of a variable number of calibration samples. Although this figure only uses a handful of selected *a priori* points within the field of view, it does not fully benchmark the behaviour of all points achieved within the shared field of view. What this figure does demonstrate is a clear decrease of errors when the number of calibration samples is increased from 4 to 15 samples. Beyond the use of 15 calibration samples, the percentage error seems to asymptotically reduce. Therefore to ensure a converging structured lighting calibration matrix, it is recommended that 15 or more calibration samples be used.

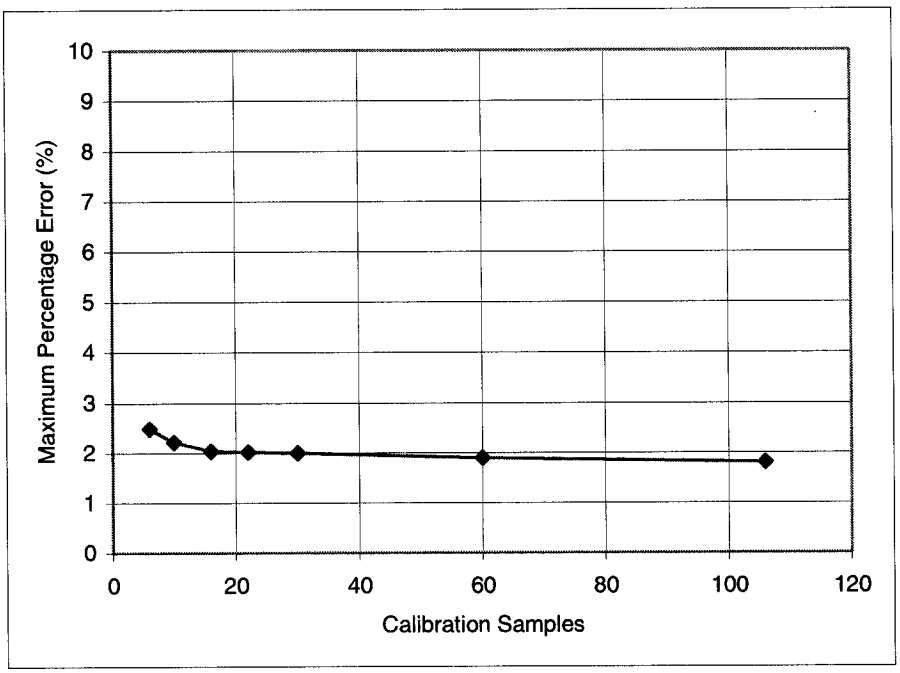

Figure 3-16 – Structured lighting maximum percentage error over a number of calibration samples

These observations were determined by collecting non-collinear calibration samples within a bounded region in the robot workspace. The calibration samples then determined a unique calibration matrix and were used to determine two three-dimensional *a priori* coordinates within the same bounded region in the robot workspace. The distance between these *a priori* points is known and when determined by the structured lighting modalities it can be compared and evaluated using eq. (60).

$$PercentageError = \frac{\left|L_{actual} - L_{experimental}\right|}{L_{actual}} \cdot 100\% \tag{60}$$

where:

$L_{experimental}$ = Length of object from subsystem three-dimensional reconstruction

L_{actual} = Actual length across the *a priori* known object

3.5.3 Stereovision and Laser Range Finder Intersystem Calibration and Acquisition

Only one subsystem-to-subsystem calibration is required to complete the design of the proposed multi-modal sensor that is a calibration between the stereovision and the laser range finder reference frames. A calibration mechanism is proposed which is similar to that used for structured lighting system calibration and which utilizes the same calibration target.

Unlike Pless and Zhang's approach for camera to laser range finder calibration, which uses a linear closed-form solution and a regressive non-linear optimization approach to correlate between the camera frame of reference and the laser range finder [64], the proposed strategy requires only the linear closed-form solution as advantage can be taken of the structured light trace on the scene. The complexity of stereovision and laser range finder subsystem calibration is reduced when the structured light is projected on the scene.

To correlate points between the laser range finder subsystem and the stereovision subsystem, the same approach as discussed in Section 3.5.2 is used where a triangular face target is placed in the direct path of the structured light pattern that the auto-synchronized active triangulation laser range finder produces. The stereovision system samples the scene and generates a disparity map of the calibration target using the Birchfield *et al.* disparity algorithm [38]. The disparity map is the same size as an image of the scene, except that each pixel value represents the disparity instead of the intensity. Similarly to structured light calibration acquisition, using a threshold or an intensity filter, the structured light segments are isolated. Once again a validation procedure that determines the sequence of foreground and

background line segments is used. This ensures that the laser range finder can properly detect the edge points of the calibration target. At the same time, it warrants that the same features are observed by the stereovision system. One additional feature of the isolated structured light segments is that it acts as an image mask to the disparity map generated by stereovision.

By masking the disparity map, the end result is an image of detected segments and their disparity values. When the foreground segment is isolated, the three-dimensional coordinates of the endpoints from the stereovision frame of reference can be determined as shown in Figure 3-17.

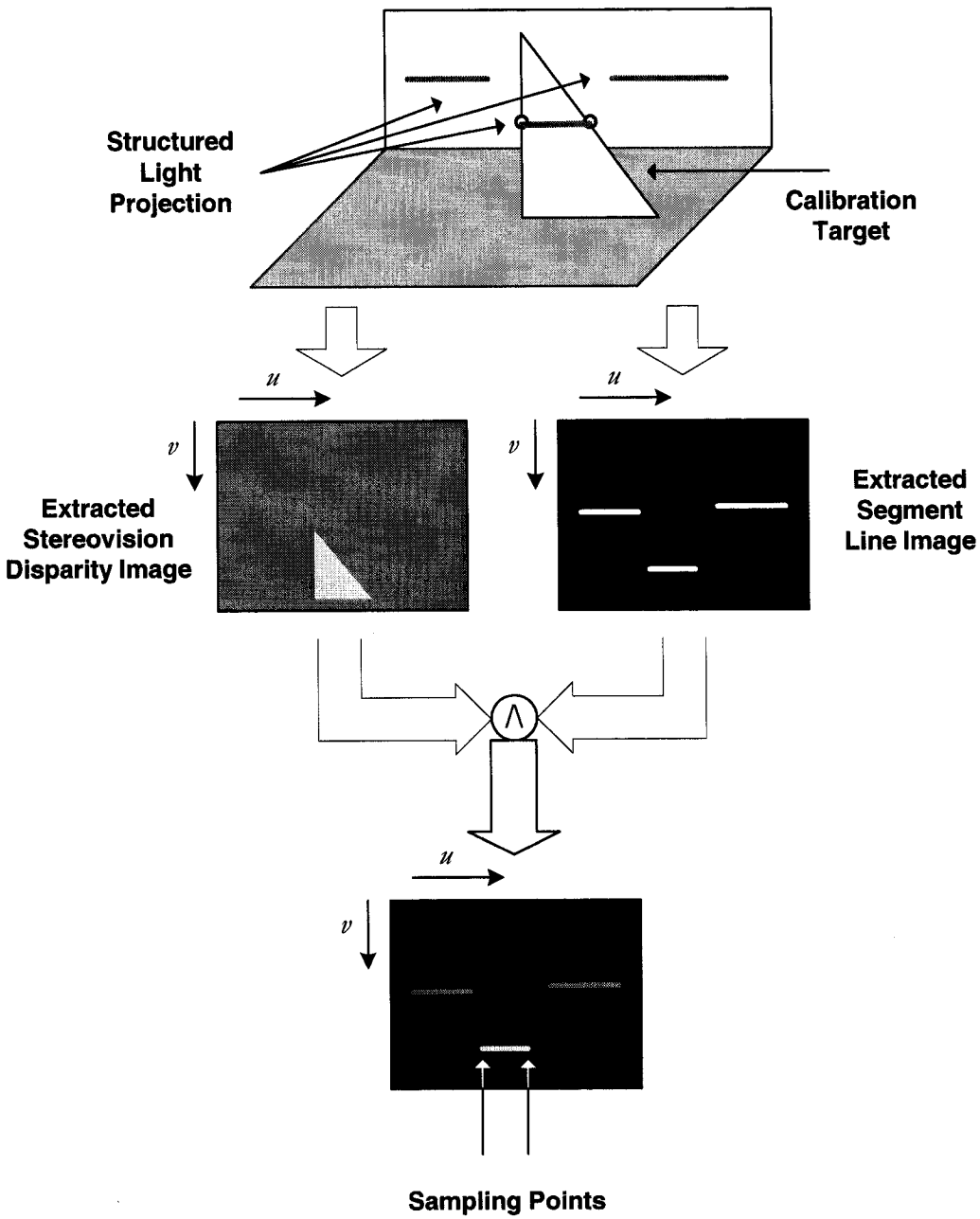

Figure 3-17 – Isolating feature points for stereovision and laser range finder intersystem calibration

Meanwhile, as in structured lighting calibration, the active triangulation laser range finder system can detect the same edge features, by extracting the endpoints of the calibration target. Thus three-dimensional reconstructed results from stereovision can be used to correlate the laser range finder extracted features as shown in Figure 3-18.

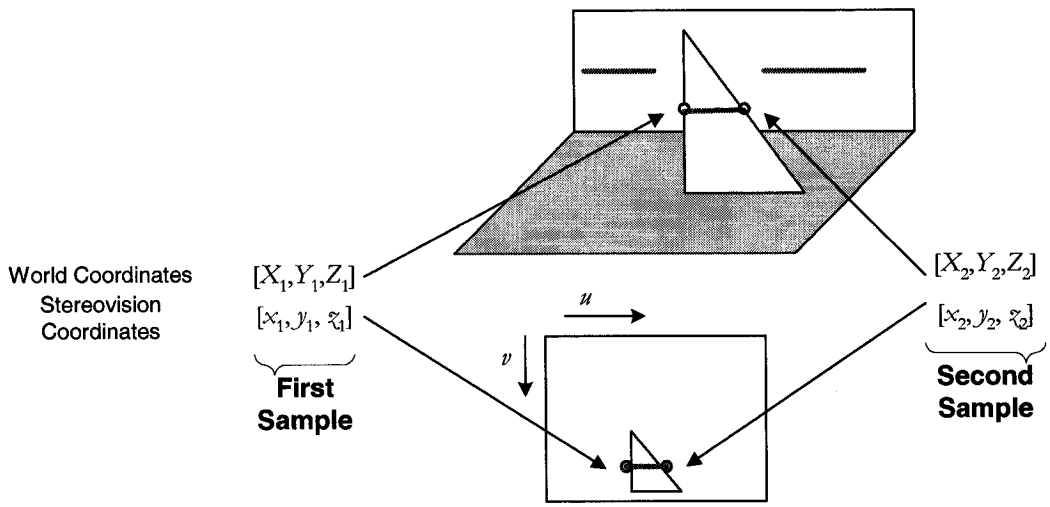

Figure 3-18 – Inter-subsystem calibration: stereovision and laser range finder

The following standard transformation equation allows to relate the stereovision reference frame to the laser range finder reference frame as long as their extrinsic properties remain unchanged, which is ensured by the fixed assembly of sensor's components on a robot's end effector. A three-dimensional coordinate observed by the laser range finder and the stereovision subsystem is denoted respectively by P_{LRF} and P_{SV}. The stereovision system is related to the laser range finder by a rotational matrix and a translation vector denoted respectively by R_{S2L} and T, which transform coordinates from the stereovision frame of reference whose origin is on one camera optical centre to the laser range finder frame of reference, such that:

$$P_{LRF} = R_{S2L} P_{SV} - T \tag{61}$$

Since the laser range finder depth perception points are within the field of view of the stereovision system, the equation to relate the laser range finder to the stereovision system is rewritten.

$$P_{SV} = R_{L2S}(P_{LRF} + T) \tag{62}$$

where $R_{L2S} = R_{S2L}^{-1}$.

Given the fact that the laser range finder used in the proposed setup scans a single line, the laser range finder depth values are two-dimensional such that all points are found along a common plane, $Y = 0$, defined in laser range finder frame of reference. Therefore all sampled depth values can be denoted as $P_{LRF} = [X, Z, 1]^T$ in homogeneous coordinates. Eq. (62) becomes:

$$P_{SV} = R_{L2S}\begin{bmatrix} 1 & 0 & \\ 0 & 0 & T \\ 0 & 1 & \end{bmatrix} P_{LRF} \tag{63}$$

where $T = [t_X, t_Y, t_Z]^T$ and R_{L2S} is a (3×3) rotational matrix.

Solving eq. (63) is then simplified by the identification of the calibration matrix M.

$$P_{SV} = M \cdot P_{LRF} \tag{64}$$

Or in a more formal representation:

$$\begin{bmatrix} X_{SV} \\ Y_{SV} \\ Z_{SV} \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{bmatrix} \cdot \begin{bmatrix} X_{LRF} \\ Z_{LRF} \\ 1 \end{bmatrix} \tag{65}$$

From a series of correlated sample points obtained from the stereovision and laser range finder subsystems respectively, the series of linear equations can be formed as:

$$Ar = b \tag{66}$$

where:

$$A = \begin{bmatrix} X_{1LRF} & Z_{1LRF} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ X_{2LRF} & Z_{2LRF} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{NLRF} & Z_{NLRF} & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_{1LRF} & Z_{1LRF} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & X_{2LRF} & Z_{2LRF} & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & X_{NLRF} & Z_{NLRF} & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & X_{1LRF} & Z_{1LRF} & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & X_{2LRF} & Z_{2LRF} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & X_{NLRF} & Z_{NLRF} & 1 \end{bmatrix}$$

$$r^{\mathsf{T}} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{21} & m_{22} & m_{23} & m_{31} & m_{32} & m_{33} \end{bmatrix}$$

$$b^{\mathsf{T}} = \begin{bmatrix} X_{1SV} & X_{2SV} & \cdots & X_{NSV} & Y_{1SV} & Y_{2SV} & \cdots & Y_{NSV} & Z_{1SV} & Z_{2SV} & \cdots & Z_{NSV} \end{bmatrix}$$

This system can then be solved directly for r, from which T and R_{L2S} are extracted.

Theoretically, the use of three calibration sample points is sufficient in solving the series of linear equations defined in eq. (64). However in order to determine a converging relation it is suggested that a higher number of calibration points be used and the system of equations solved through the least squares technique. Once the transformation matrix, M, is determined, the relationship between the laser range finder frame of reference and the stereovision frame of reference is defined. The relationship from the stereovision frame of reference to the laser range finder frame of reference is simply obtained by inverting the transformation matrix.

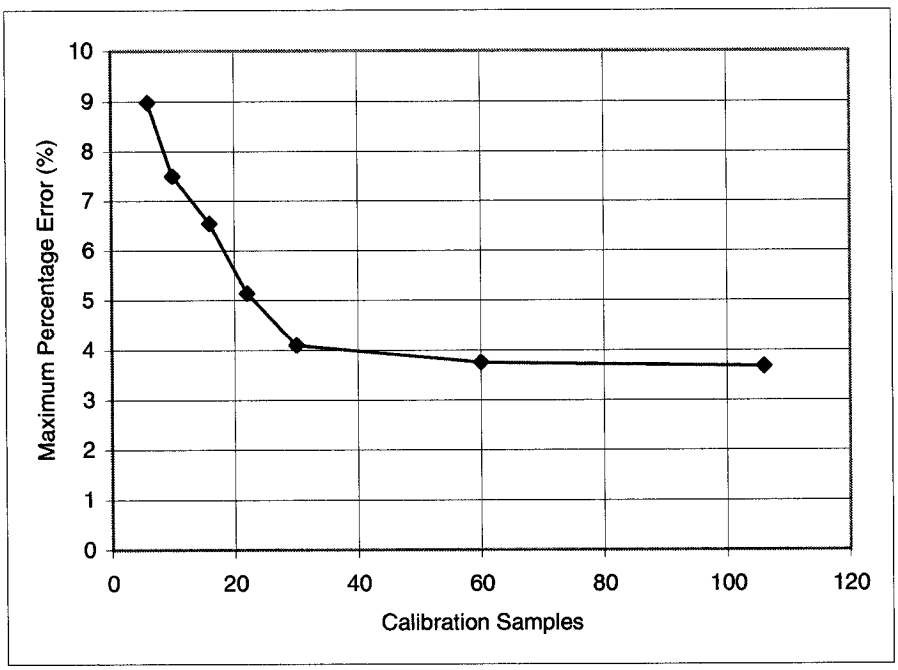

Figure 3-19 – Extrinsic stereovision to laser range finder calibration maximum percentage error over a number of calibration samples

Preliminary field-testing using the proposed multi-modal system demonstrated that increasing the number of calibration samples gradually improves the percentage error on reconstructed

features as determined from eq. (60) and shown in Figure 3-19. These tests were conducted respectively with 6, 10, 16, 22, 30, 60 and 106 calibration samples to determine the trend on the effect of a variable number of calibration samples on extrinsic stereovision to laser range finder calibration. Similarly to structured lighting, calibration samples were collected within a bounded region of the stereovision and laser range finder shared field of view. The calibration samples were then used to determine the extrinsic calibration matrix and tested for *a priori* known three-dimensional coordinates also within the same bounded region in the robot workspace. Similar to the results discussed in structured lighting calibration, when the number of calibration samples is increased from 3 to 25 samples there is a large decrease in the percentage error. Well beyond 25 calibration samples, it is observed that the percentage error only slightly reduces. Therefore the use of 25 or more calibration samples is recommended to be used to ensure that the extrinsic stereovision to laser range finder calibration matrix converges properly.

3.6 Overview of System Design and Calibration

Following the original design considerations that we have defined in Section 3.4.1, all three steps have been completed and resulted in a fully calibrated multi-modal system. In the first step, a flexible and well-built infrastructure has been defined that maintains the physical relationships between each modality. Although various physical configurations can be established with our system, each variation would have its own set of unique calibration parameters and would require the calibration process to be re-performed to determine these. The second step of importance uses ingenuity to merge possible calibration procedures that effectively define the essence of each modality. This requires the understanding of calibration of an individual modality within the multi-modal system and integrating common features of the procedure with calibration process features of other subsystems.

As observed, both the structured lighting calibration and the extrinsic stereovision to laser range finder calibration count on a common frame of reference: the laser range finder frame of reference. This observation is the intentional design process outlined in the third step, where a multi-modal global frame of reference is commonly defined for each modality. In the proposed system, this global frame of reference is the optical centre of the active triangulation laser range finder. The structured lighting frame of reference becomes the optical centre of the structured light projector, hence the laser range finder optical centre. The stereovision frame of reference is also the optical centre of the laser range finder, which facilitates the extrinsic calibration. Lastly, the laser range finder frame of reference is by default its own optical centre. The unison of these modalities' ability to provide three-dimensional datasets from the global frame of reference completes all essential steps that we have defined as a successful calibrated multi-modal range sensing system which eliminates need for data post-processing.

Figure 3-20 summarises the proposed multi-modal system's properties, from the modalities incorporated, the equipment used, to the intra-system (intrinsic) and the inter-system (extrinsic) calibration applied.

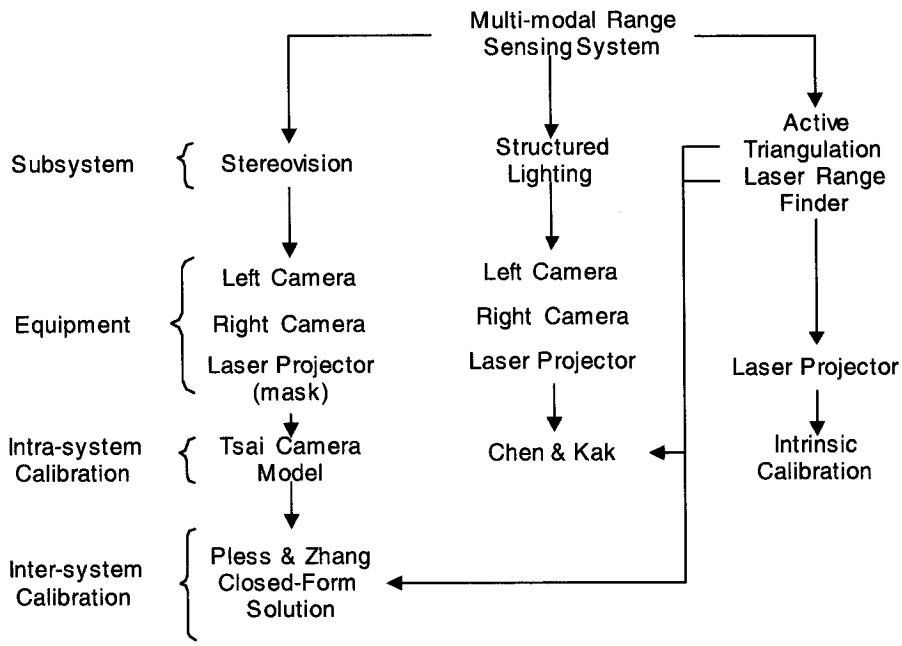

Figure 3-20 – Multi-modal system's structural overview

One key feature to note of the active triangulation calibration and extrinsic stereovision to laser range finder calibration is that its entire calibration process can be completed within a single instance. To clarify, calibration samples for both processes are acquired simultaneously. For this purpose, the calibration target used for this multi-modal system serves for both Chen and Kak's structured lighting approach and Pless and Zhang's closed-form extrinsic calibration. It is also noted that structured lighting calibration and extrinsic stereovision to laser range finder calibration uses a similar mechanism to acquire samples, thus permitting the reuse of these calibration samples for both procedures.

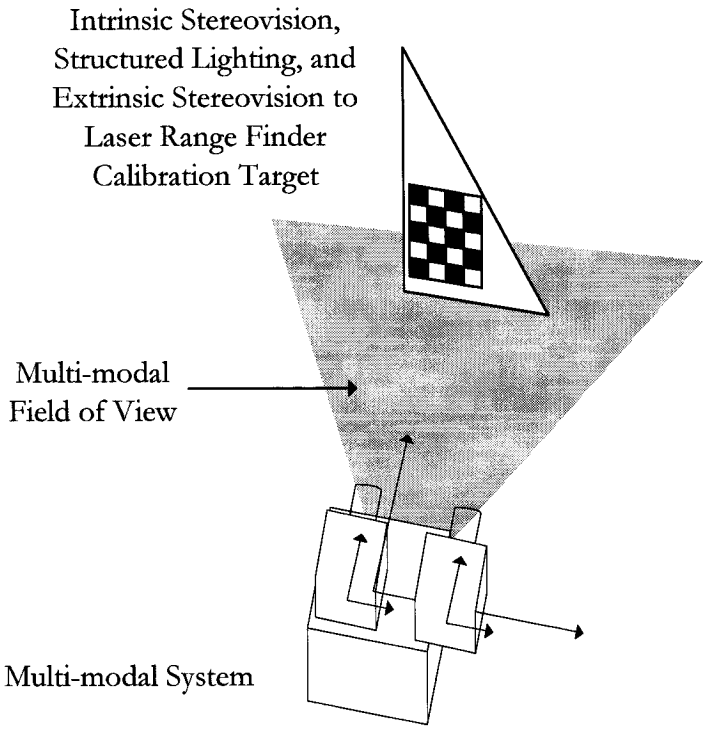

Figure 3-21 – Extended calibration procedure - merging all calibration targets into one

If we extend the proposed strategy one step further, intrinsic stereovision calibration can also be integrated within the same calibration space by adding a calibration pattern on the triangulation target used by structured lighting calibration and extrinsic stereovision to laser range finder calibration as shown in Figure 3-21. Once again, by positioning the end-effector in various locations in the robotic calibration workspace with visibility to the stereovision calibration pattern would ensure calibration of the stereovision system via Tsai's camera calibration method.

Chapter 4

SIMULATIONS OF MULTI-MODAL PROTOTYPE

To design and evaluate the proposed approach described in Chapter 3, simulation tests have been conducted. This chapter presents and discusses the results from these simulation tests. The key benefit of creating a simulated environment allows us to examine the ability and performance of the proposed multi-modal system in a cost-reduced method without potentially damaging the delicate instruments. In addition, these simulations provide a glimpse of any potential problems that might be encountered and observed in the real system prototype, which will be reported in Chapter 5.

4.1 Simulation Environment

From the proposed approach in Section 3.4.1, the three steps of building a multi-modal system were outlined. To recapitulate, the first step defined a multi-modal system infrastructure where each modality remains physically related to each other. The second step required a calibration procedure to be implemented for each modality. And the third step required that a multi-modal global frame of reference be defined. This simulation will approach each step in such a way that it may be used as a comparison or prediction of results observed in the proposed multi-modal system.

Before the simulated multi-modal system is built, there are several assumptions that can be made when using a simulated environment. One advantage that is inherited by all simulations is the ability to construct a flawless environment; however here are a few more benefits:

1. Features are easily and correctly extracted and can be modelled explicitly without concern of environment or lighting conditions;

2. Precision of intrinsic properties of subsystems is easily modelled with minimal error;

3. Occlusion of feature points due to viewpoint variation can be conveniently ignored;

4. Positioning of subsystems can be easily accomplished with minimal calibration processing;

5. Objects to be simulated are much easier to be built in a virtual environment, instead of searching for real physical objects to be used;

6. The ability to simulate a multi-modal system without actual use of specific equipment.

The ease of positioning subsystems virtually anywhere in the simulated environment eliminates any physical constraint that would be considered in a real multi-modal system. The size, weight, and packaging of the cameras and the active triangulation laser range finder are dynamics to be considered. Yet for a simulated multi-modal range sensor, the position and orientation of the frame of reference for each modality is only the minimal requirement. To effectively create an environment that will somewhat mimic realistic equipment, Matlab software is used to model the cameras and active triangulation laser range finder. Intrinsic and extrinsic camera properties of the VRex CAM-3000C have been pre-determined by the use of OpenCV [65], which is third-party open source software incorporating Tsai's camera calibration and stereovision application. These parameters, typical for a parallel stereovision system with a configuration as in Figure 2-4, are defined in Appendix A and hence used as

models of the simulated stereovision system with minor alterations. The image size of the simulated cameras are 320 × 240 pixels similar to the real camera system. Modelling of the active triangulation laser range finder on the other hand is noticeably basic where the assumption is that the laser range finder has flawless near infinite depth perception and a considerably large field of view, therefore no comparison and evaluation of the efficiency of the active triangulation laser range finder is made to other modalities. One particular feature that is preserved is the laser range finder's two-dimensional planar perception, coplanar to its frame of reference *XZ* axis, as it is central to the structured light and subsystem modality extrinsic calibration as shown in Figure 4-1.

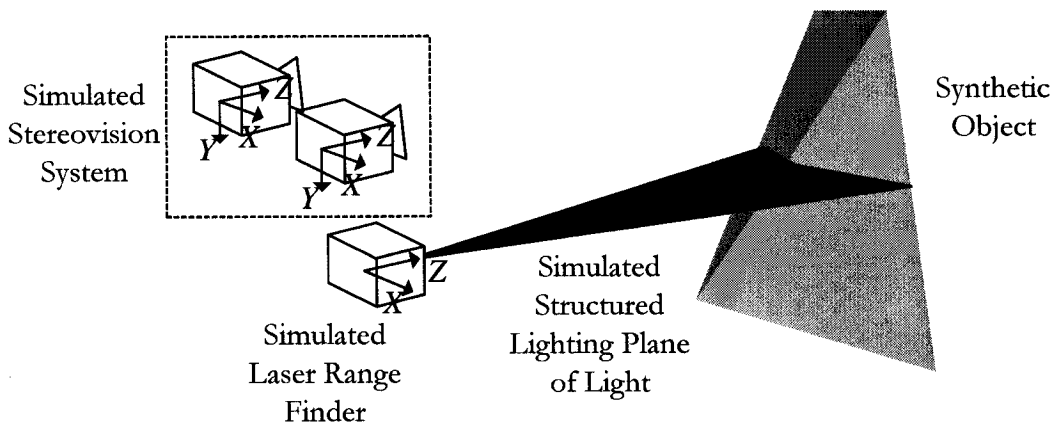

Figure 4-1 – Physics of the simulated environment

The physical relationship between each modality is structured with consideration of mentioned behaviour discussed in the latter half of the first step in Section 3.4.1, due to structured lighting's inability to effectively perceive depth when the camera's frame of reference is coplanar to the structured light stripe pattern. A similar configuration to that of Figure 3-3 and Figure 3-6 is utilized to remain consistent with the proposed multi-modal system.

One of two significant points of this simulated environment will be demonstrated as the theoretical adaptations made to classical calibration procedures that are used as a proof of concept. Calibration samples, either defined in the environment or perceived by the simulated laser range finder, will pioneer towards determining a relationship between each modality. This procedure will complete the objectives defined in the second and third steps discussed in the proposed multi-modal concept defined in Section 3.4.1. As the requirements that define a multi-modal system have been achieved, the latter contribution of this simulation emerges as the ability of observing the multi-modal system subjected to stimulus. Thus being how the multi-modal system accurately and precisely detects features in the environment.

4.2 Multi-modal Calibration

With each modality positioned in a fixed relative position to each other, simulation will determine whether or not theoretical calibration matrices can be constructed in the simulated environment. Since the intrinsic and extrinsic camera calibration matrices are provided as preamble to the simulation, there is no need for stereovision camera calibration. Therefore only two calibration steps are required to complete the simulated multi-modal calibration process: structured lighting (intra-system) and extrinsic stereovision (inter-system) calibration.

Instead of moving the multi-modal system to different positions in the simulated environment to resample additional calibration points from a single calibration target, a series of *a priori* calibration points are defined in the world reference frame that lies on the projected structured light as shown in Figure 4-2. This approach is similar to moving the calibration target to different positions as opposed to moving the multi-modal system. Since both structured lighting calibration and extrinsic stereovision to laser range finder calibration can be calibrated simultaneously using the same calibration points, only calibration points visible by both the left and right cameras are selected. These same calibration points are re-used again to test the stereovision reconstruction algorithm defined in Section 2.2.2.

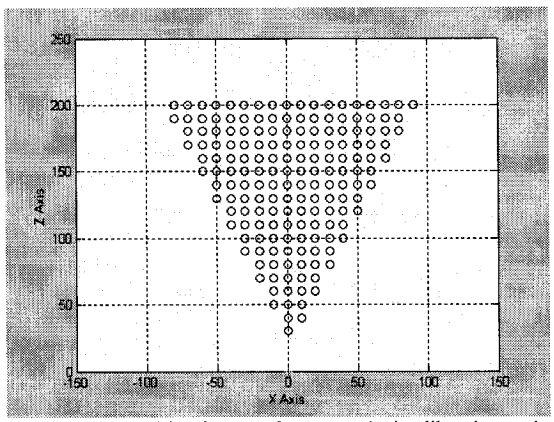

Figure 4-2 – World reference frame a priori calibration points

To ensure that the calibration matrix converges to a distinct solution and following the analysis from Figure 3-16 and Figure 3-19, 170 calibration points are used as sampling points, which is more than sufficient for both structured lighting and stereovision calibration matrices to converge. This provides 170 calibration points for each structured lighting system calibration and 170 calibration points for the inter-system stereovision and laser range finder calibration, which can provide reconstruction results for distances within the calibration workspace. After using calibration algorithms discussed in Section 3.5.1 and Section 3.5.3, the calibration matrices are formed as shown in Appendix B. Before approving the calibration matrices determined by the simulation, the calibration points captured by the camera system are applied to the calibration matrices to reveal the *a priori* global reference frame points.

The reconstructed calibrated points from 170 mm to 200 mm show the first sign of inaccurate reconstruction as shown in Figure 4-3, Figure 4-5, and Figure 4-7. Although the number of calibration points used was significant, this inaccuracy is due to the precision of the modelled camera system, which is initially providing a resolution of 320 × 240 pixels. If a higher pixel resolution camera was to be used, results in the reconstruction of the same depth would drastically improve as shown in Figure 4-4, Figure 4-6, and Figure 4-8 where a 640 × 480 pixel image is used for the simulated camera. In these figures, the same points in the field of view are reconstructed with better results. Reconstructed points between 170 mm to 200

mm in the 640 × 480 image clearly show four rows of sampling points as opposed to three rows of concatenated points when a lower camera resolution of 320 × 240 pixels is used.

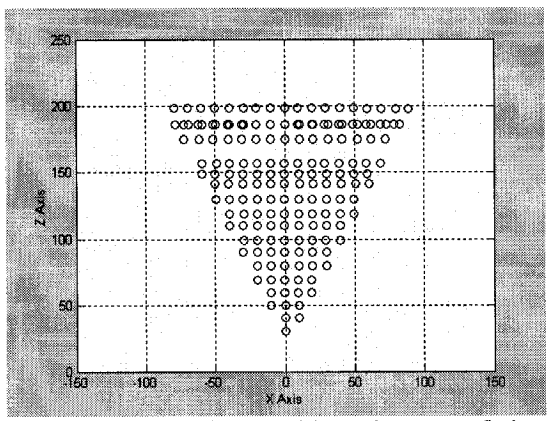

Figure 4-3 – Extrinsic stereovision to laser range finder calibration point reconstruction (320 × 240 pixel images)

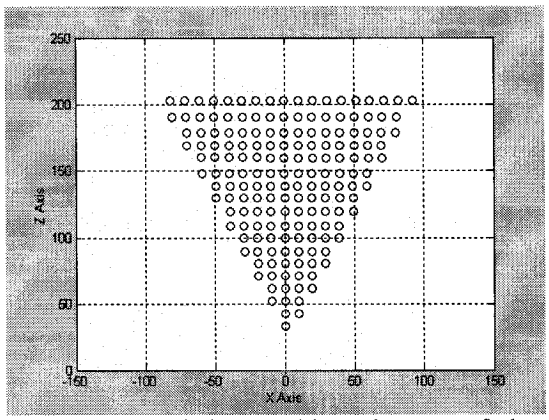

Figure 4-4 – Extrinsic stereovision to laser range finder calibration point reconstruction (640 × 480 pixel images)

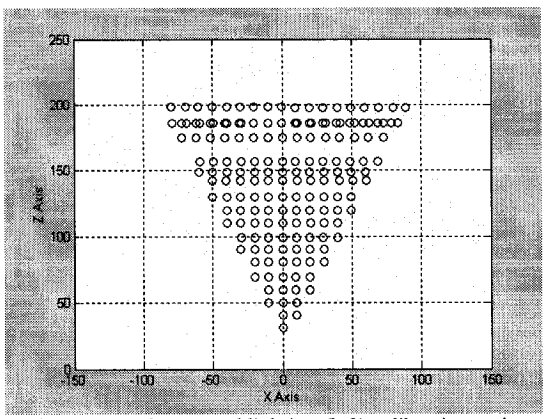

Figure 4-5 – Structured lighting (left) calibration point reconstruction (320 × 240 pixel images)

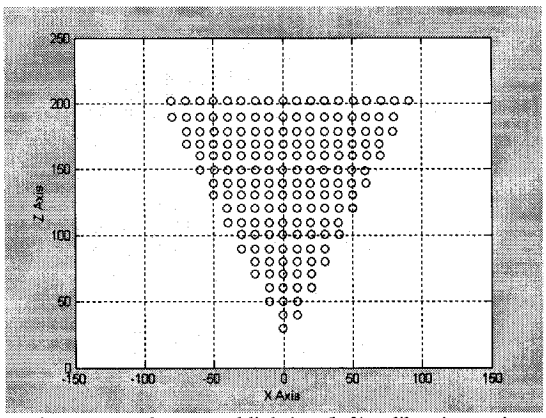

Figure 4-6 – Structured lighting (left) calibration point reconstruction (640 × 480 pixel images)

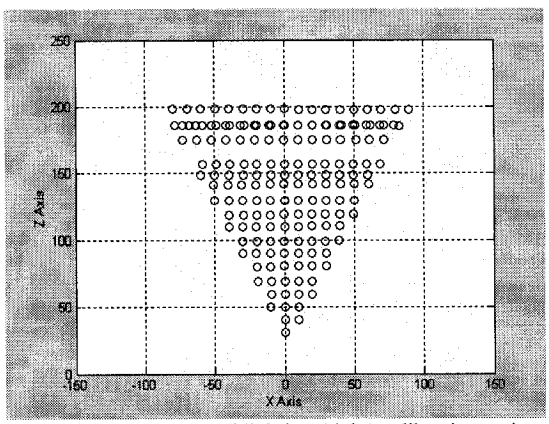

Figure 4-7 – Structured lighting (right) calibration point reconstruction (320 × 240 pixel images)

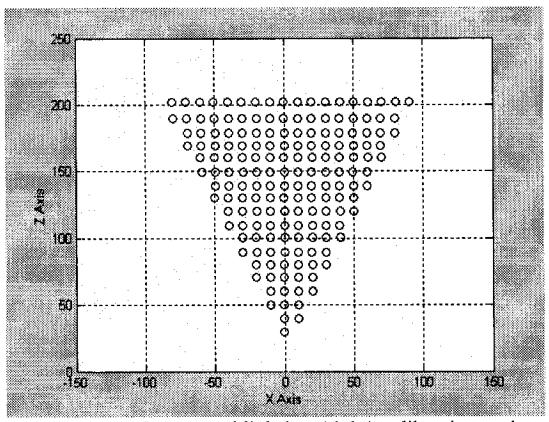

Figure 4-8 – Structured lighting (right) calibration point reconstruction (640 × 480 pixel images)

4.3 Scene Reconstruction

The second experiment examines the efficiency of the multi-modal calibration by performing multi-modal three-dimensional reconstruction. Since all subsystems reconstruct the scene in relation to the laser range finder reference frame, a comparison of the performance of each multi-modal subsystem is achieved by visual inspection. The basis of comparison will be made against the synthetic models constructed to test three common features found in environments or on objects such as flat surfaces, contoured surfaces and edges.

For this purpose, four simple models are constructed to imitate objects used with the prototype multi-modal range sensor. The first simple model is a planar surface in the middle of the scene to imitate a pillar in the scene as shown in Figure 4-9 and Figure 4-10. The second model is a convex surface placed in the middle of scene as shown in Figure 4-11 and Figure 4-12, and the third model is a concave surface as shown in Figure 4-13 to Figure 4-14. These two models will provide interesting results on how effectively the multi-modal range model can reconstruct varying contoured objects. The fourth and fifth synthetic models in Figure 4-15 to Figure 4-20 are composed of two planar surfaces that intersect with the intersecting edge being inwards exposed and outwards exposed. The use of this model is to

simulate what the multi-modal sensor might perceive in an indoor room environment such as wall corners, which is more relevant for robot path planning purposes.

Figure 4-21 to Figure 4-50 are the reconstructed results from the synthetic models using the simulated 320×240 pixel resolution images.

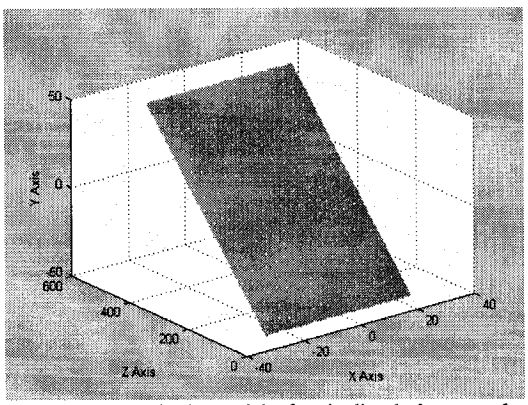

Figure 4-9 – Synthetic model of an inclined planar surface

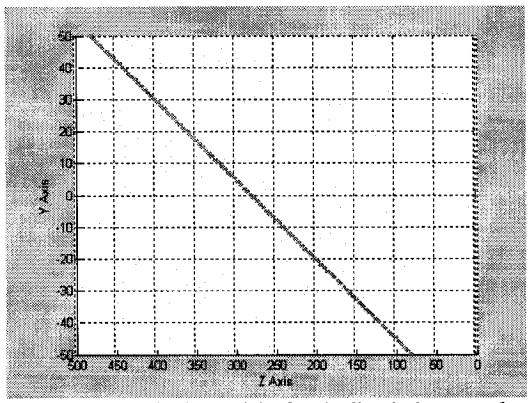

Figure 4-10 – Synthetic model of an inclined planar surface – cross section view

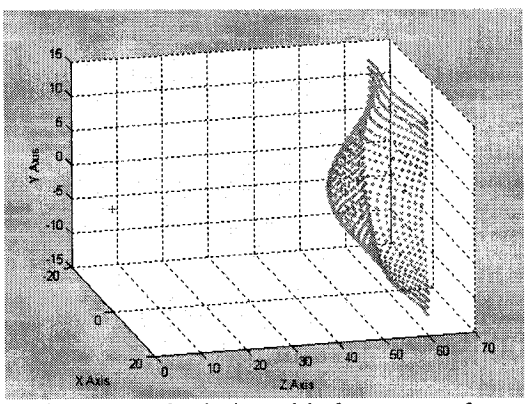

Figure 4-11 – Synthetic model of a convex surface

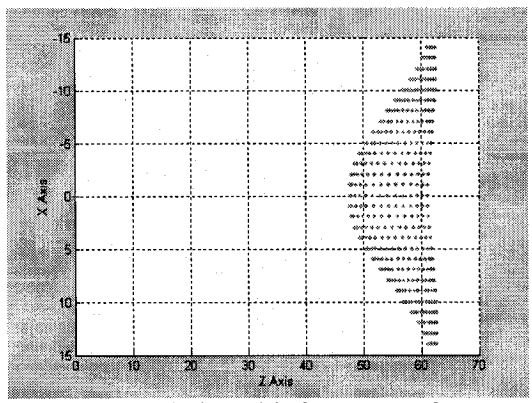

Figure 4-12 – Synthetic model of a convex surface – cross section view

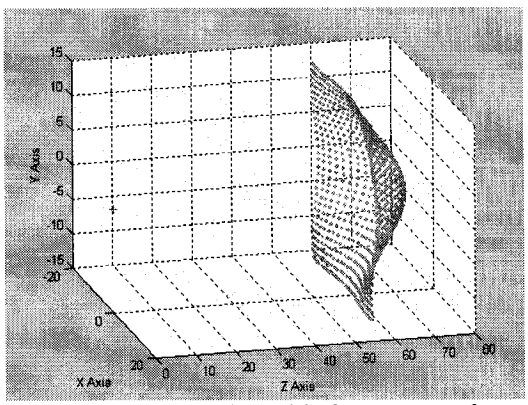

Figure 4-13 – Synthetic model of a concave surface

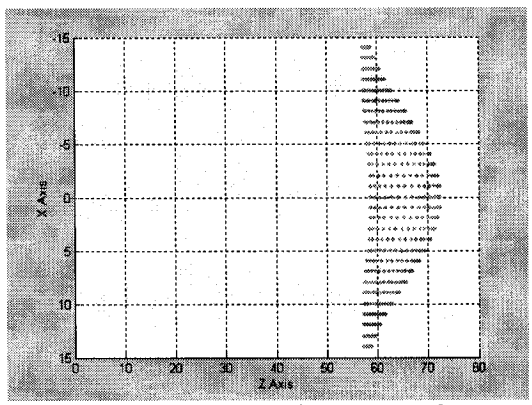

Figure 4-14 – Synthetic model of a concave surface – cross section view

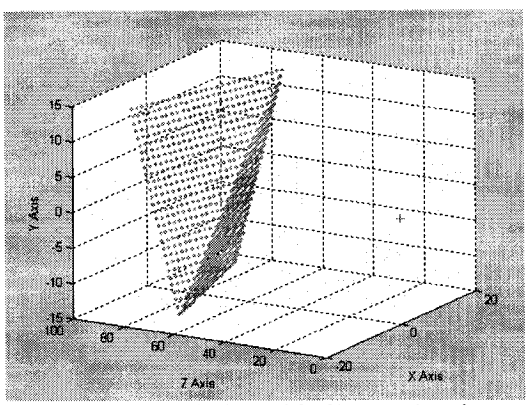

Figure 4-15 – Synthetic model of intersecting planar surfaces with edge facing inwards – left perspective

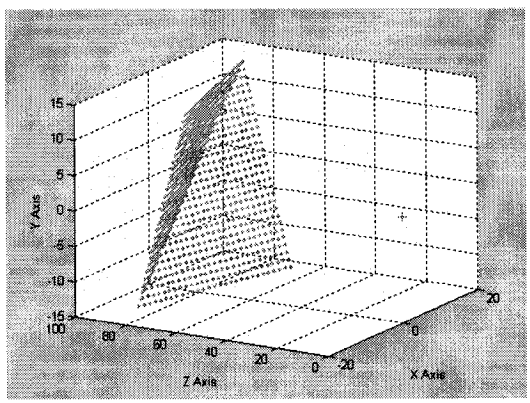

Figure 4-16 – Synthetic model of intersecting planar surfaces with edge facing outwards – left perspective

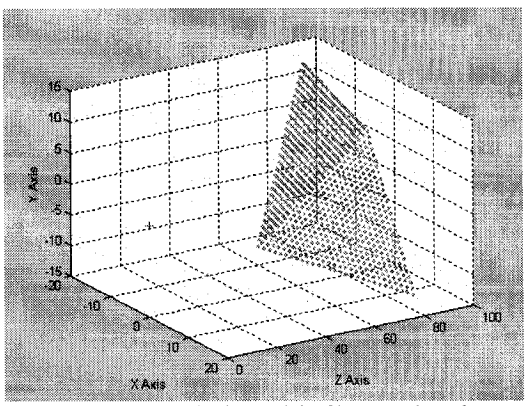

Figure 4-17 – Synthetic model of intersecting planar surfaces with edge facing inwards – right perspective

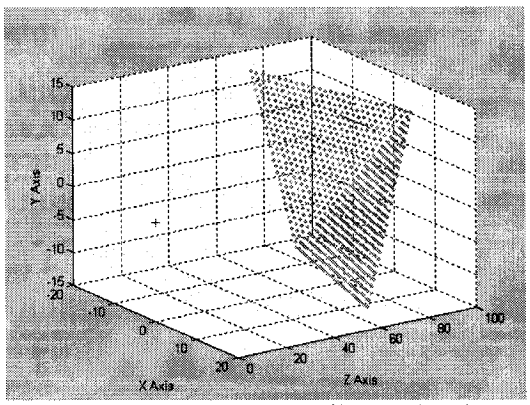

Figure 4-18 – Synthetic model of intersecting planar surfaces with edge facing outwards – right perspective

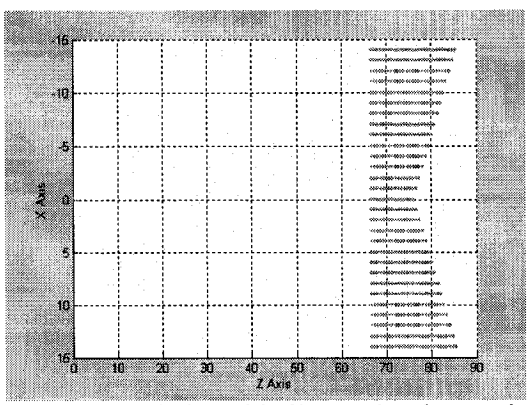

Figure 4-19 – Synthetic model of intersecting planar surfaces with edge facing inwards – cross section

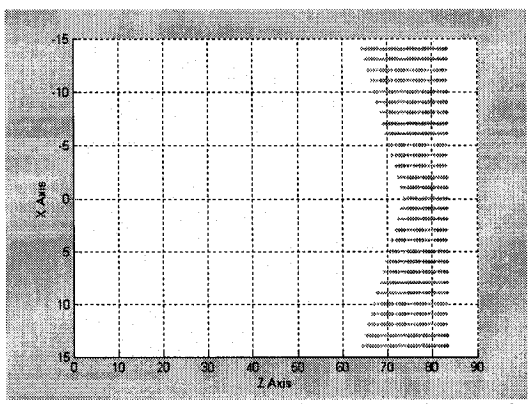

Figure 4-20 – Synthetic model of intersecting planar surfaces with edge facing outwards – cross section

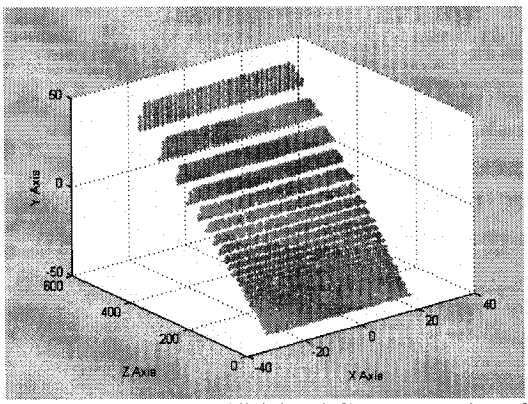

Figure 4-21 – Structured lighting (left) reconstruction of planar surface

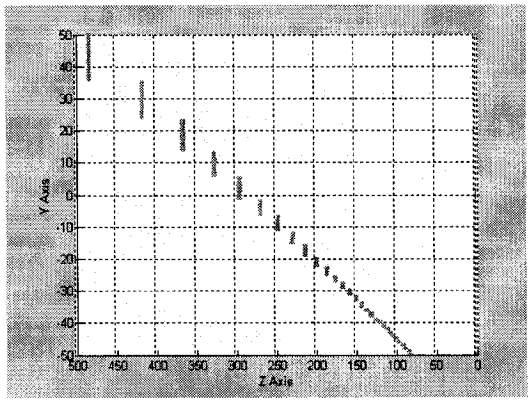

Figure 4-22 – Structured lighting (left) reconstruction of planar surface – cross section view

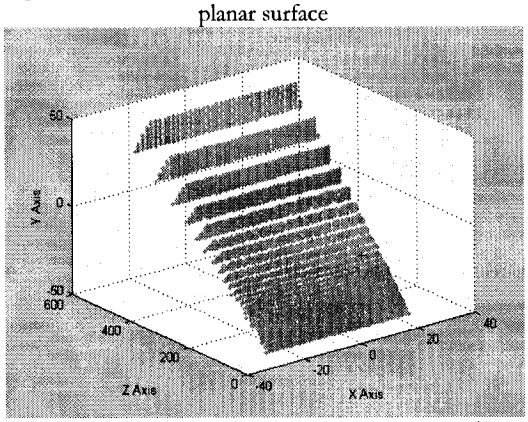

Figure 4-23 – Structured lighting (right) reconstruction of planar surface

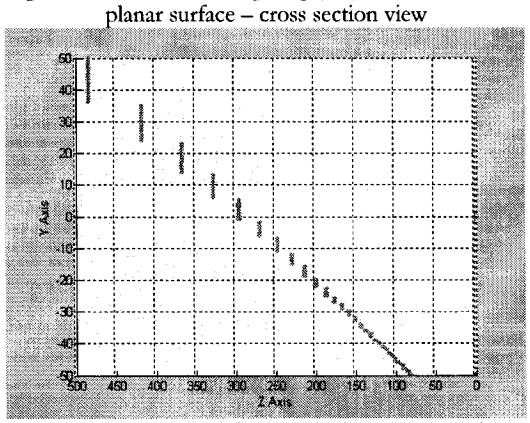

Figure 4-24 – Structured lighting (right) reconstruction of planar surface – cross section view

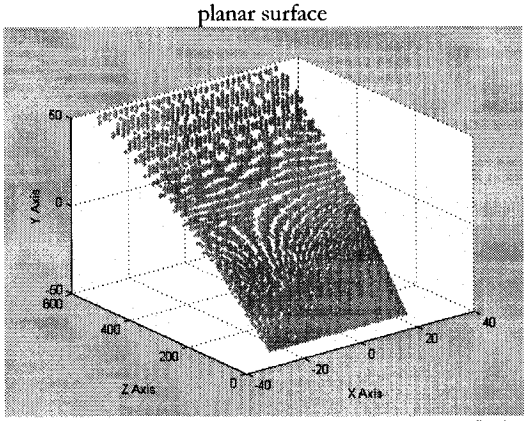

Figure 4-25 – Extrinsic stereovision to laser range finder reconstruction of planar surface

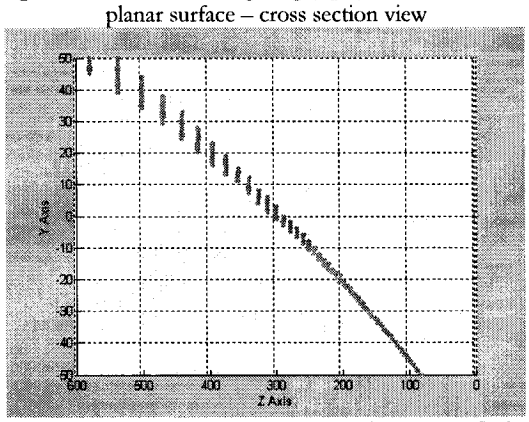

Figure 4-26 – Extrinsic stereovision to laser range finder reconstruction of planar surface – cross section view

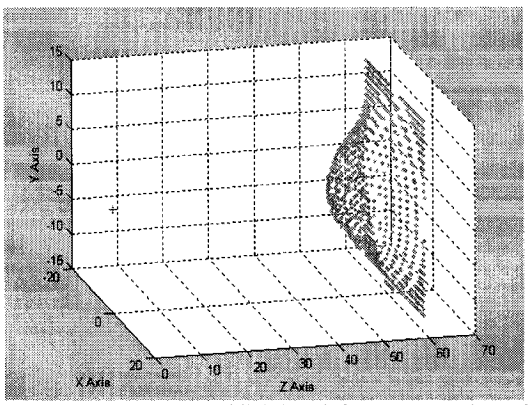

Figure 4-27 – Structured lighting (left) reconstruction of convex surface

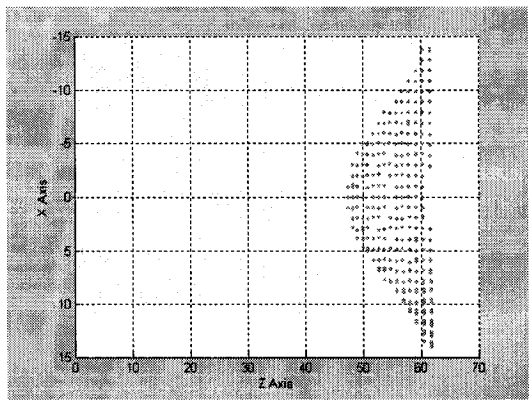

Figure 4-28 – Structured lighting (left) reconstruction of convex surface – cross section view

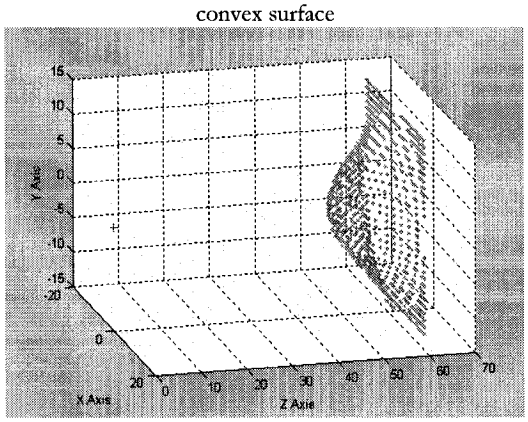

Figure 4-29 – Structured lighting (right) reconstruction of convex surface

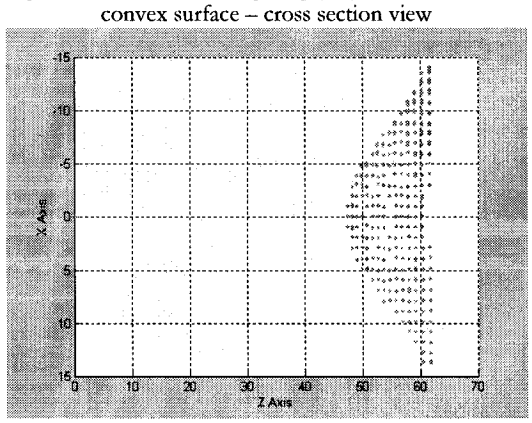

Figure 4-30 – Structured lighting (right) reconstruction of convex surface – cross section view

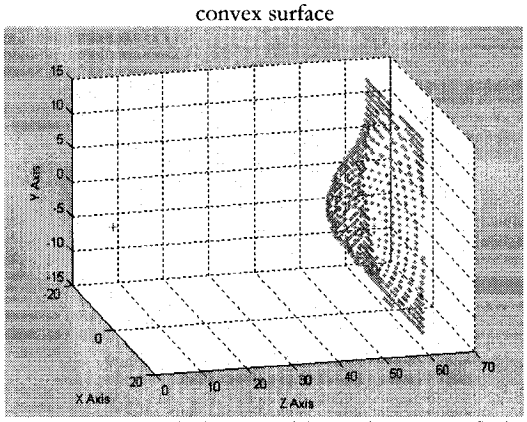

Figure 4-31 – Extrinsic stereovision to laser range finder reconstruction of convex surface

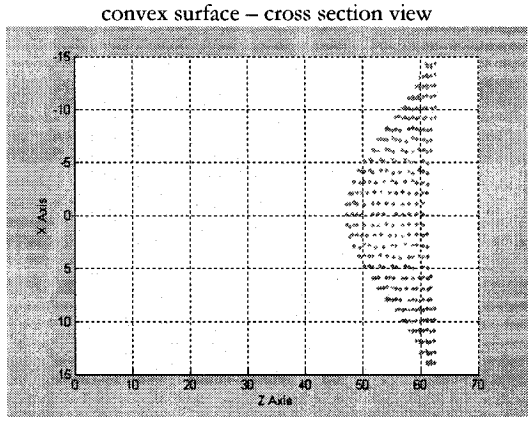

Figure 4-32 – Extrinsic stereovision to laser range finder reconstruction of convex surface – cross section view

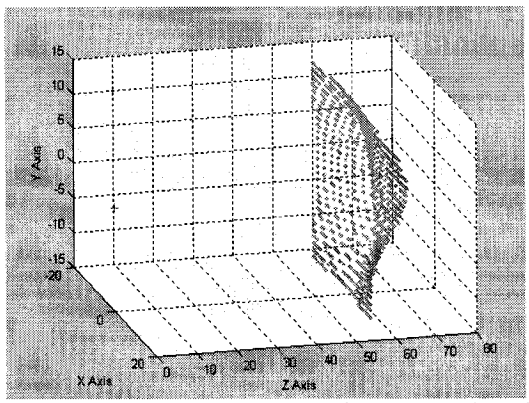

Figure 4-33 – Structured lighting (left) reconstruction of concave surface

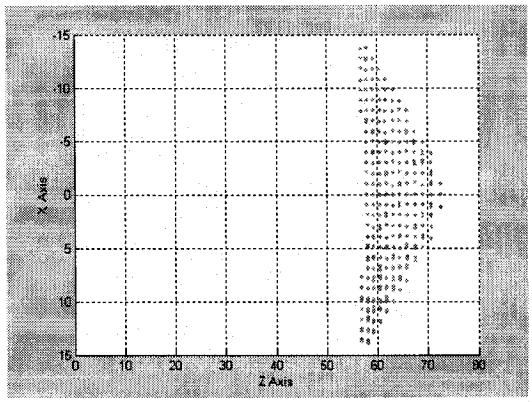

Figure 4-34 – Structured lighting (left) reconstruction of concave surface – cross section view

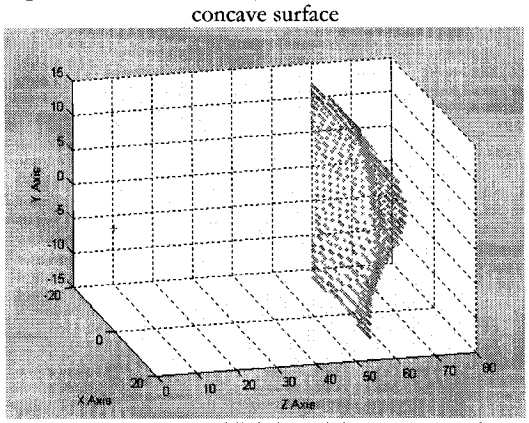

Figure 4-35 – Structured lighting (right) reconstruction of concave surface

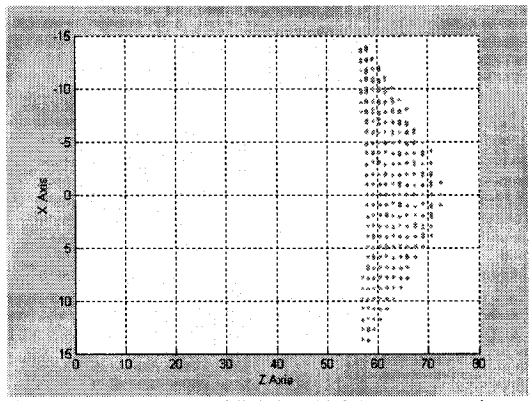

Figure 4-36 – Structured lighting (right) reconstruction of concave surface – cross section view

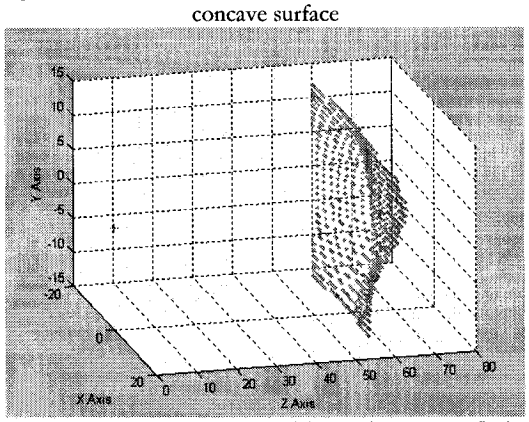

Figure 4-37 – Extrinsic stereovision to laser range finder reconstruction of concave surface

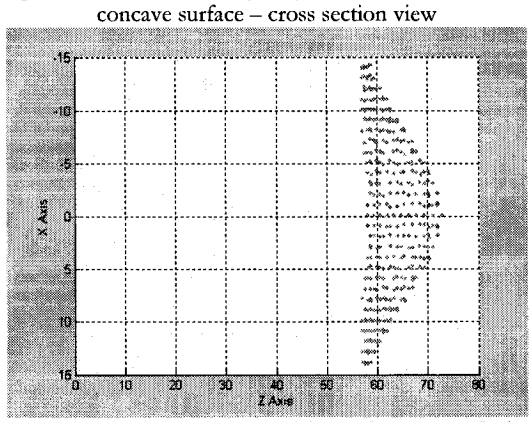

Figure 4-38 – Extrinsic stereovision to laser range finder reconstruction of concave surface – cross section view

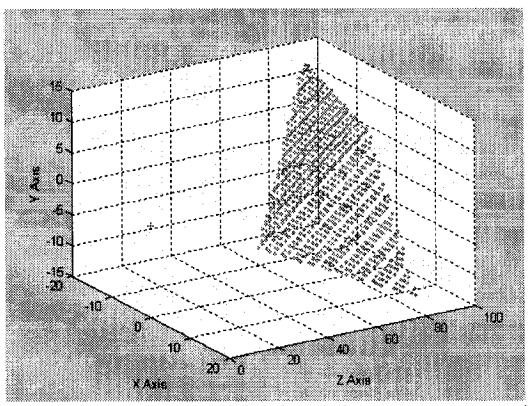

Figure 4-39 – Structured lighting (left) reconstruction of intersecting planar surfaces with edge facing inwards

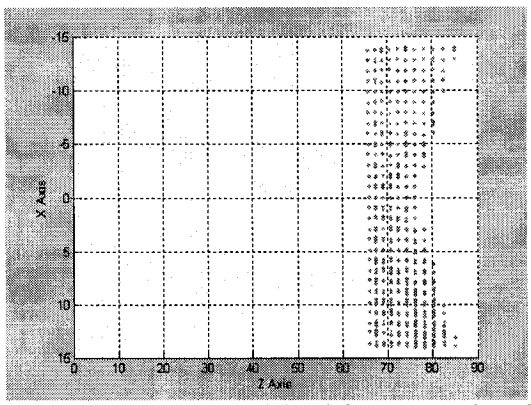

Figure 4-40 – Structured lighting (left) reconstruction of intersecting planar surfaces with edge facing inwards – cross section view

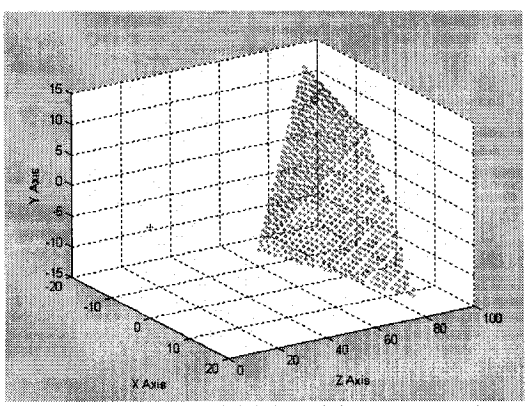

Figure 4-41 – Structured lighting (right) reconstruction of intersecting planar surfaces with edge facing inwards

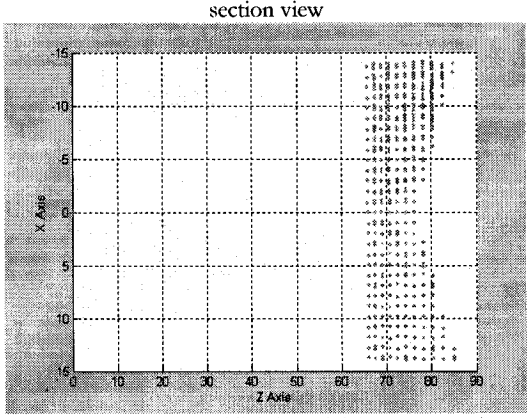

Figure 4-42 – Structured lighting (right) reconstruction of intersecting planar surfaces with edge facing inwards – cross section view

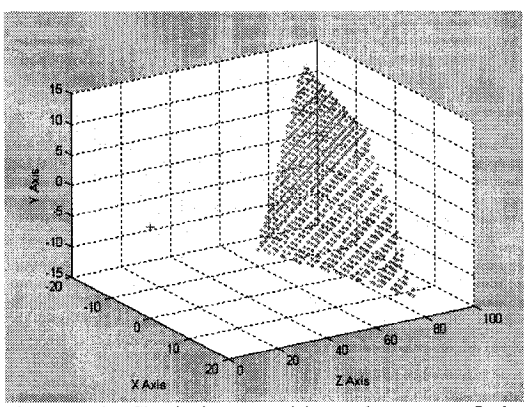

Figure 4-43 – Extrinsic stereovision to laser range finder reconstruction of intersecting planar surfaces with edge facing inwards

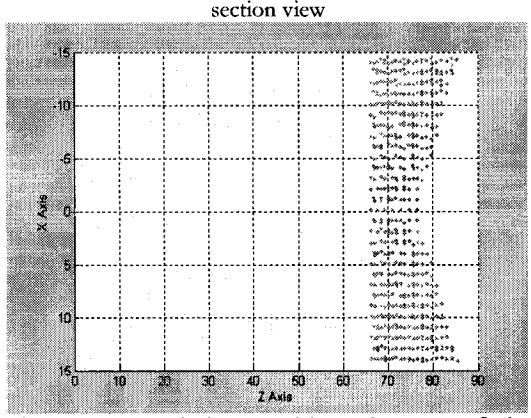

Figure 4-44 – Extrinsic stereovision to laser range finder reconstruction of intersecting planar surfaces with edge facing inwards – cross section view

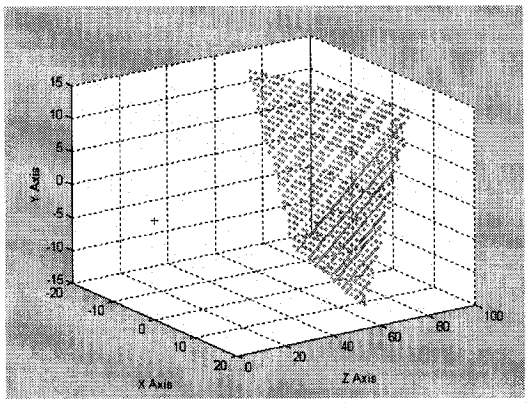

Figure 4-45 – Structured lighting (left) reconstruction of intersecting planar surfaces with edge facing outwards

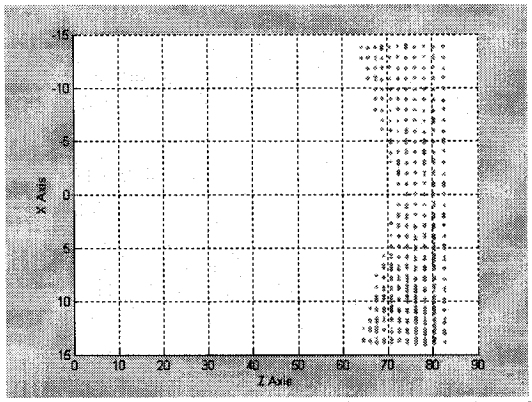

Figure 4-46 – Structured lighting (left) reconstruction of intersecting planar surfaces with edge facing outwards – cross section view

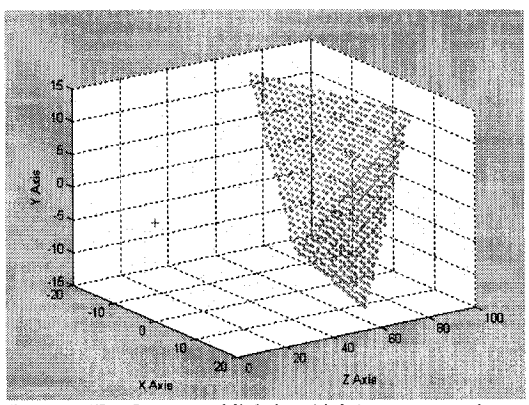

Figure 4-47 – Structured lighting (right) reconstruction of intersecting planar surfaces with edge facing outwards

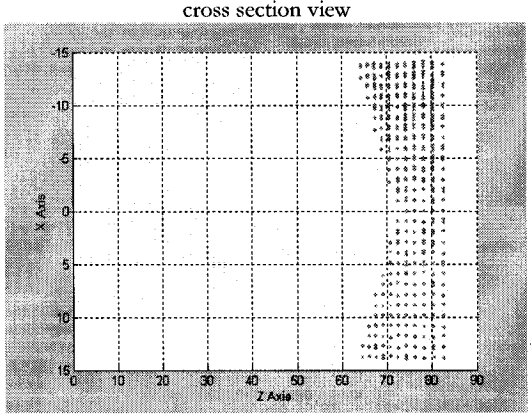

Figure 4-48 – Structured lighting (right) reconstruction of intersecting planar surfaces with edge facing outwards – cross section view

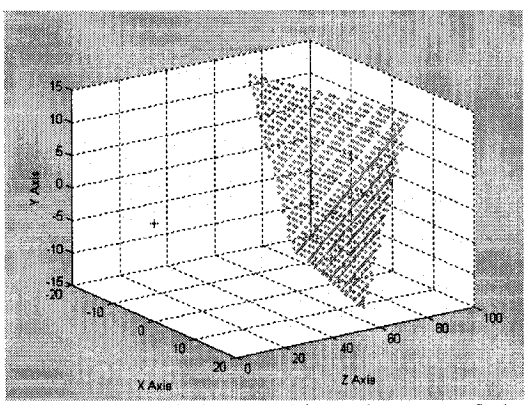

Figure 4-49 – Extrinsic stereovision to laser range finder reconstruction of intersecting planar surfaces with edge facing outwards

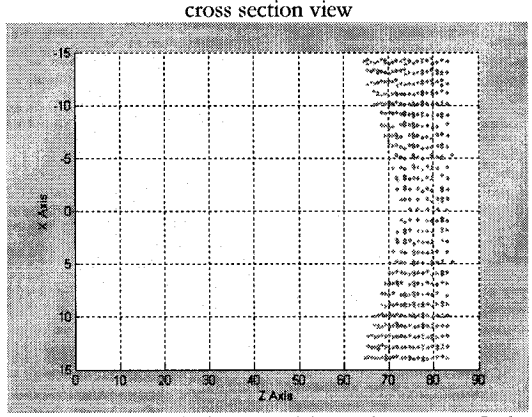

Figure 4-50 – Extrinsic stereovision to laser range finder reconstruction of intersecting planar surfaces with edge facing outwards – cross section view

The reconstruction of the synthetic surfaces proves one crucial factor: that the theoretical framework for multi-modal perception is indeed usable. Moreover, results prove that the dataset from each subsystem indeed provides range sensing from the laser range finder frame of reference regardless of its original subsystem frame of reference.

Referring to the first synthetic model of the planar surface shown in Figure 4-21 to Figure 4-26, it is quite apparent that the reconstruction increasingly fails to accurately report the depth of objects further away from the origin point of reference. However, this problem is due to two main factors: the image pixel resolution and the physical distance between the subsystem and the multi-modal origin. As objects are placed further away from the multi-modal system, the structured light detects the light pattern at the same pixel locations. Therefore, if the same distant object was moved slightly forward or slightly behind its original depth would have no significant impact on the structured light pattern detected in the image and thus would be interpreted as no difference in depth. By improving the resolution of sampling by either increasing the pixel resolution of the camera system and/or increasing the physical distance between the camera system and the structured light projector, hence the origin of the multi-modal system, it is possible to reduce the error. However the trade-off lies in the subsystem's ability to perceive objects placed closer to the origin.

The second and third synthetic models shown in Figure 4-27 to Figure 4-38, enforce the analysis of the first synthetic model but also demonstrate how stereovision and structured lighting subsystems interpret contoured surfaces. Both subsystems are oblivious to the contour and its gradient as long as there are feature points that can be extracted. Although, the stereovision subsystem provides equivalent results to structured lighting, it is important to note that the correlation of feature points along the contour are always correct and thus, regardless of pixel intensity or any other factors of importance to stereovision disparity algorithms, the simulated subsystem will accurately report the correct disparity.

The fourth and fifth synthetic models shown in Figure 4-39 to Figure 4-50 provide an analysis of how the multi-modal system perceives intersecting planar surfaces that create an edge in the scene or a corner. With both synthetic models, structured lighting subsystem reconstructions of the feature point of the two intersecting planes are not quite similar to the actual synthetic model in Figure 4-15 to Figure 4-18. Once again, the phenomenon observed in the reconstruction of the first synthetic model and in reconstruction of the *a priori* calibration points are visible. Instead of a "flat" planar surface, the reconstructed results present striated planar surfaces where the reconstructed feature points are grouped together as highlighted in Figure 4-51. Stereovision reconstruction analysis also exhibits the striated planar surface, which is partially due to the short baseline between the two cameras and the short distance from the stereovision frame of reference to the laser range finder frame of reference and the resolution of the camera.

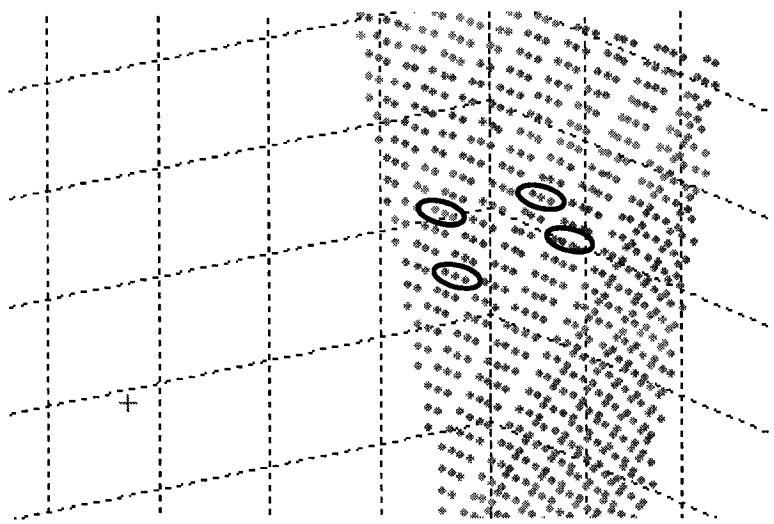

Figure 4-51 – Striated planar surface from reconstruction

The right structured lighting system reconstruction of these synthetic models apparently does not exhibit this phenomenon as visibly as its left counterpart, due to an increased distance from the right camera optical centre to the structured light projector centre. Overall, all three

subsystems can roughly detect where the two intersecting planes intersect regardless whether or not the edge faces towards or behind.

One point to mention is that just by increasing the distance between the projector and camera system in structured lighting or increasing the baseline distance within the stereovision system will not completely reduce or eliminate the inaccuracies described above. By increasing these distances a limitation will be introduced where each modality will be unable to share a field of view that is closer to the system.

4.4 Error Analysis

Simulations demonstrated that both structured lighting and stereovision subsystems are dependent upon the precision of the cameras used. If the resolution of the image is coarse, the accuracy of the extracted features from the image suffers drastically. As objects are placed further away, there is no distinguishable difference in the image if the object is slightly moved closer or further away from the camera. The structured light pattern that is projected onto the scene would be observed at the same pixel location in the image reference plane. To determine how drastic these errors are the first synthetic model is used to compare between the actual referenced points to the experimentally determined points. To calculate the error, a simple equation, eq. (67), that relates the difference between the experimental results to the actual points is used. The following graphs depict the increasing errors from structured lighting and stereovision subsystems as the planar surface is distant from the multi-modal system.

$$PercentageError = \frac{\left|D_{theo} - D_{exp}\right|}{D_{theo}} \cdot 100\% \tag{67}$$

where:

D_{theo} = Theoretical distance from *a priori* feature to multi-modal centre

D_{exp} = Experimental sampled distance from *a priori* feature to multi-modal centre

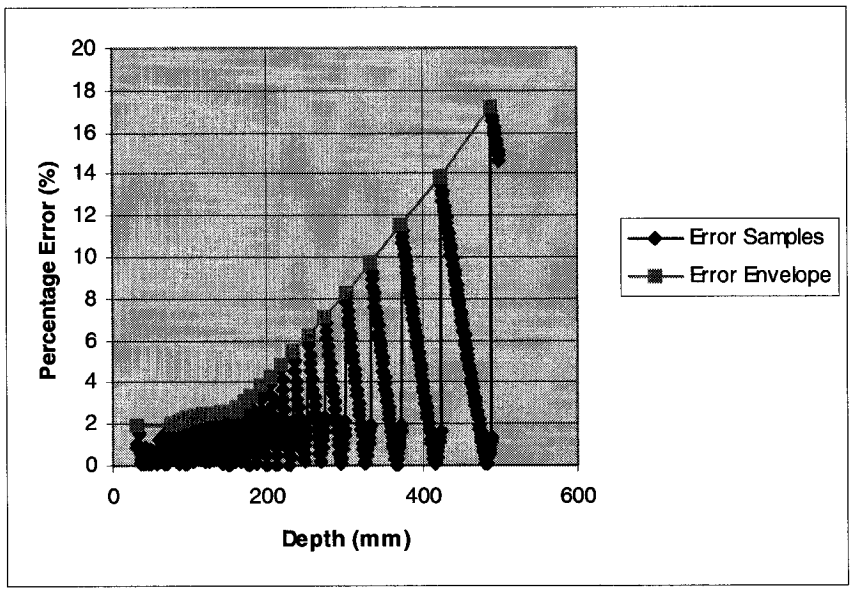

Figure 4-52 – Structured lighting (left) percentage error and envelope

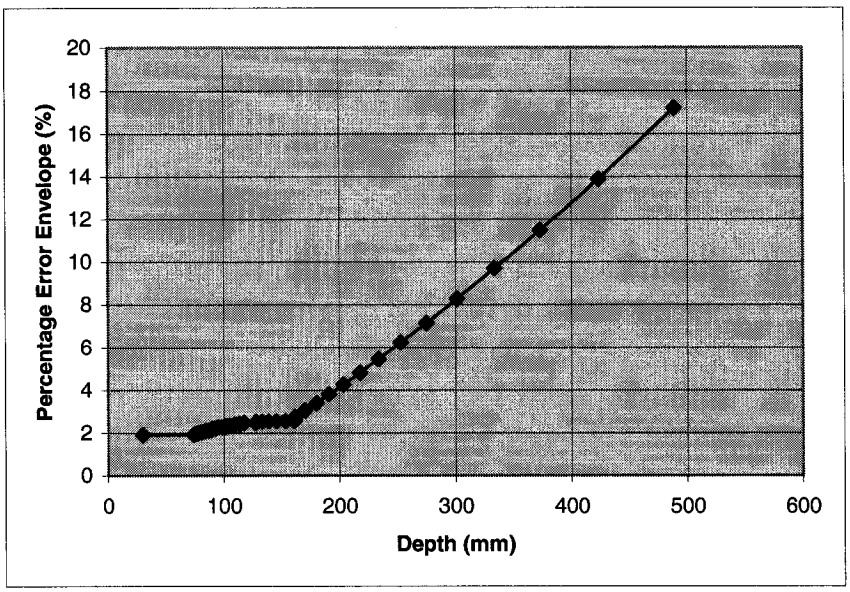

Figure 4-53 – Structured lighting (left) percentage error envelope

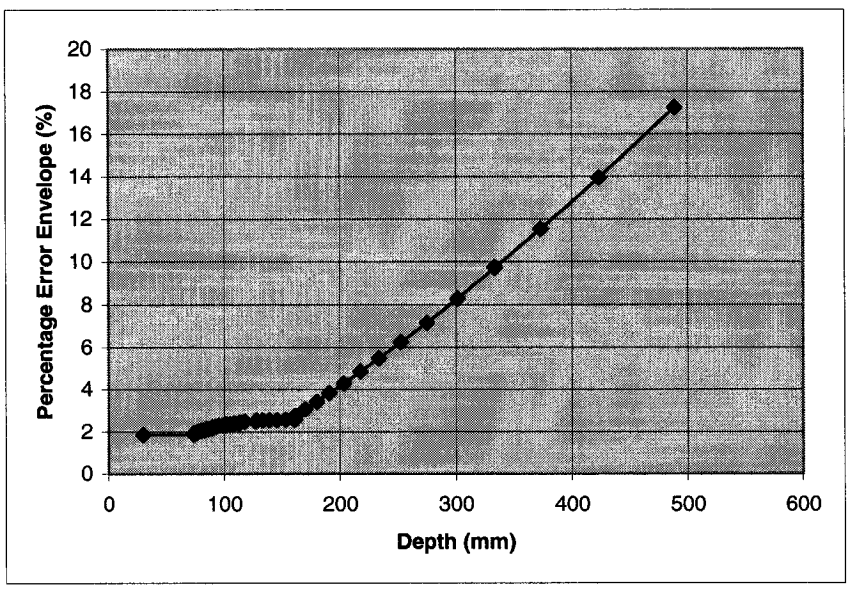

Figure 4-54 – Structured lighting (right) percentage error envelope

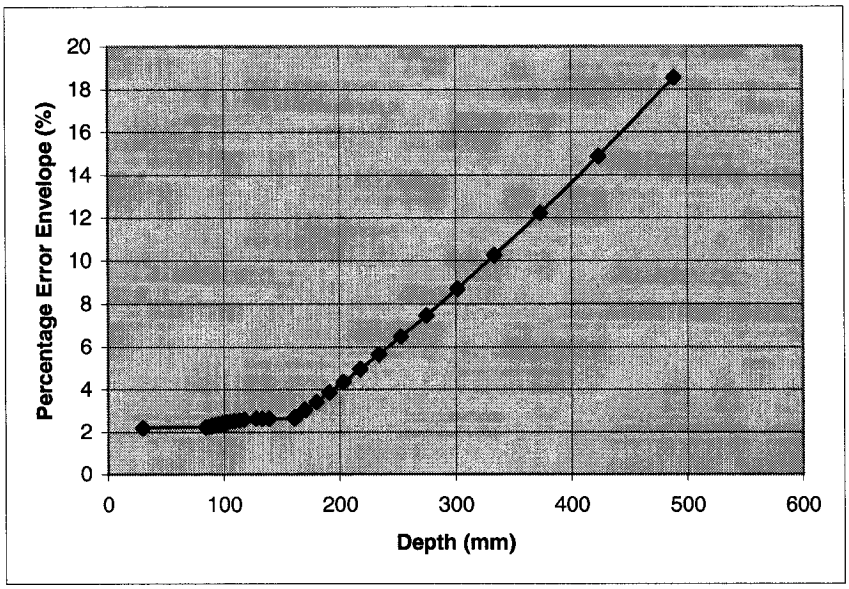

Figure 4-55 – Stereovision to laser range finder percentage error envelope

The percentage error from structured lighting describes the amount of error between the independent subsystems. Figure 4-52 shows the actual errors calculated at a one-millimetre granularity on the distance to the object. As observed, large deviations occur as the distance of the *a priori* object is moved further away. The characteristic of this deviation is based on the resolution of sampling of the point. For feature points detected further away from the camera system, it is difficult to precisely determine from the image sampling how far away the feature is. A common analogy to this phenomenon is to ask oneself to visually estimate the distance to an object and then move this object a unit away from its current position before re-estimating the distance. If the object was close, then it is quite easy to determine that the object moved exactly a unit distance. However, if the object was already far away and then moved by the same unit distance, it is difficult to determine that the object in fact actually moved. The same characteristics apply to the camera perspective projection. When a far object is detected, it is difficult to determine exactly how far the object is. If the object were placed a bit closer or a bit further from its position and sampled again, the camera would not

be able to detect any changes because the variation in the perspective projection would be within one pixel resolution. Thus the reported distance of the object would remain fixed based upon the discretization of the detected feature in the image. To better illustrate the trend of this error, an envelope of the maximum error can be built as shown in Figure 4-53 to Figure 4-55. Although the left subsystem is placed closer to the origin of the multi-modal system, its error is slightly higher than that of the right subsystem whose distance to the origin is slightly larger. Analysis of error envelope for stereovision reveals that its characteristics are similar to that of structured lighting but with a slightly higher error most likely due to the camera resolution.

The goal of this simulation is to provide a model for the real proposed system discussed in the next chapter. Assuming that the operation of the real system follows closely the simulation model, the results made available by the prototype multi-modal system should be expected to be similar to those determined by simulation.

Chapter 5

PROTOTYPE DEVELOPMENT AND EXPERIMENTAL RESULTS

In the previous chapters, a multi-modal range sensing strategy combining three modalities, its calibration, and three-dimensional reconstruction algorithms have been defined and applied in a simulated environment with controlled synthetic data. To test the proposed multi-modal range sensing strategy in a real environment, a number of extensions have been achieved. This chapter will discuss these additions as well as the construction and the results obtained with our prototype of the multi-modal sensor, named VIVA M^2S-SSL (**M**ulti-**M**odal **S**ystem, **S**tereovision, **S**tructured lighting, and **L**aser range finder).

5.1 Operation in Real World

In the simulation experiments presented in Chapter 4, all the feature points were assumed to be visible by all range sensing subsystems. Of course, this is not going to be the case for real image data for the following reasons:

i. Features can be occluded within a subsystem detection, such as in stereovision where some feature points detected in one camera are not visible in the other;

ii. Features extracted from one subsystem could be occluded to another subsystem;

iii. Introduction of noise from equipment or environment lighting hampers feature detection;

iv. Disparity algorithms may produce sparse disparity maps that ineffectively provide sufficient information of objects within the scene, or produce highly dense disparity maps that cause under-smoothing at object boundaries and over-

smoothing in smooth regions, resulting in incorrect disparity assignment and reconstruction.

Along with the introduction of these four factors, it should be noted that existing disparity algorithms remain fairly inaccurate. Besides stereovision's dependency on an illuminated environment, it is also dependent on the difference of textures and patterns on the objects. The chameleon effect where a given foreground blends with its surroundings and background is impervious to detection to the stereovision system as well as the human visual system. In addition, structured light subsystems are highly dependent on the ability to extract the structured light pattern. In a highly illuminated environment, it happens that this subsystem is unable to filter out the structured light pattern without the use of specific optical filters, a solution that was not considered here to keep the approach as general as possible.

Other real image processing factors and calibration steps that have not been considered in the theoretical approach and simulations are discussed in the following subsections.

5.1.1 Structured Light Filtering - Red Line Segment Filter

One of two image processing factors that must be considered is the ability to extract the structured lighting feature the active triangulation laser range finder system produces. This is essential for the structured lighting system calibration, where expected patterns must be detected to accurately determine the end points of the illuminated line segment made on the calibration target, as discussed in Section 3.5.3. In our simulation analysis, described in Chapter 4, it takes for granted that the illuminated pattern follows an ideal Labertian model such that each modality can uniquely differentiate each feature point along the pattern with no concerns of noise or misdetection. However, in real image processing, active filters and thresholds must be applied to isolate the structured pattern from the acquired images. As seen in Figure 5-2 the Servo-Robot Inc. Jupiter laser range finder forms a distinguishable red line stripe. Instinctively, the most suitable filter is to eliminate all non-red pixels in the RGB

image. However, this filter all by itself is insufficient to remove unwanted or unrelated red features within the scene. To further isolate the structured light pattern, an intensity filter is applied by converting the original RGB image to YCrCb colour space representation. The luma component (Y), also known as luminance, can be used to threshold the image that isolates the pixels with the highest amount of light that passes through them. Once again, if the approach was to be taken such that only a luminance filter was applied without the use of a red colour filter, other intensities may be detected as well, such as a well-lit background.

As shown in Figure 5-1, the process of isolating the structured light features requires the two filters and by either successively combining the filters or by individually applying the filters and merging only common pixels, the resulting image is that of red and high luminance pixels.

The last step of this filtering process is the isolation of the highest red and luminance pixel per vertical scan-line. The reasoning for having this filter is to select one single pixel in each vertical scan line that will represent the structured light feature detected in the scene. Although this step is not a mandatory requirement, its benefits arise during the structured lighting system calibration procedure. When selecting the endpoint of the structured light segment that intersects the calibration target, only one pixel within the image can be selected. As the structured light laser pattern disperses onto the scene, the camera filtering process detects the entire dispersed projection and thus a single striped line may occupy more than one pixel per vertical scan-line. The end result of this filtering process on an example image is shown in Figure 5-2 and Figure 5-3.

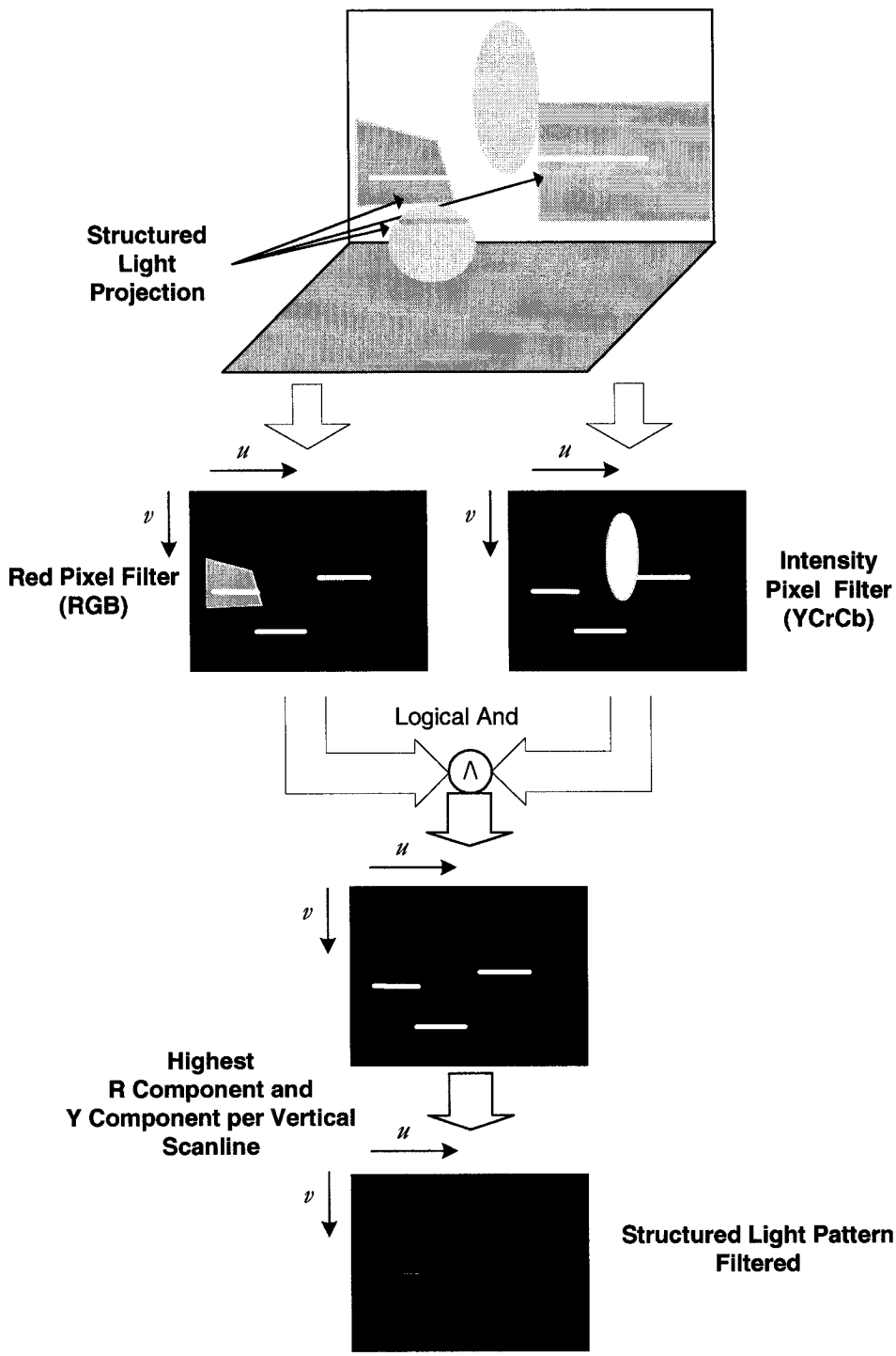

Figure 5-1 – Structured Lighting Pattern Filtering

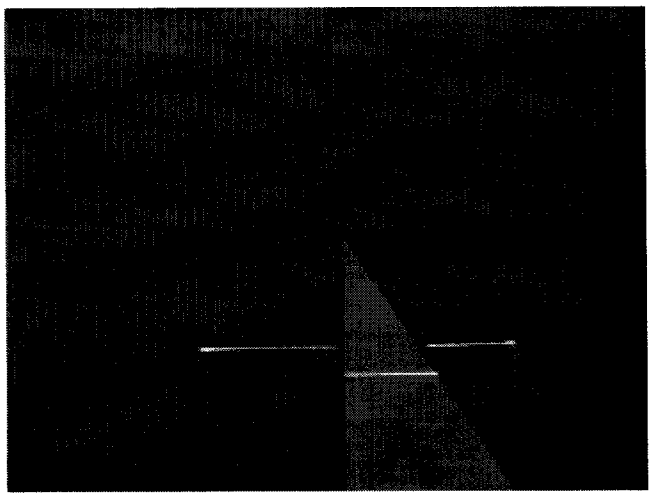

Figure 5-2 – Structured lighting system image of stripe pattern across the calibration target

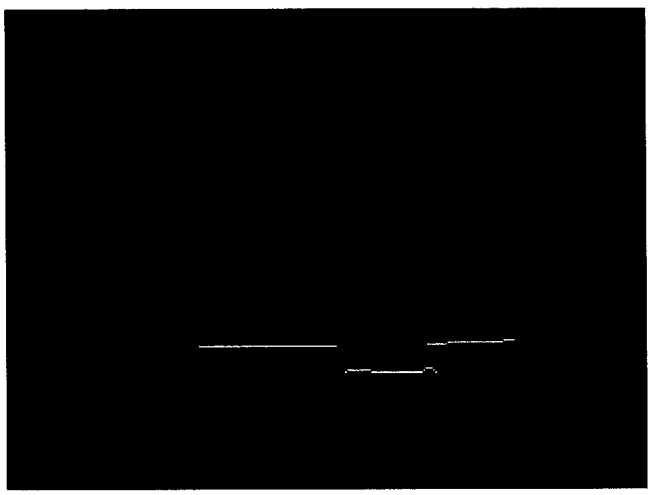

Figure 5-3 – Stripe pattern filtered using red line segment filter

5.1.2 Foreground and Background Structured Light Pattern Detection

Once the red line segment detector is able to extract the structured light pattern from the scene as described in Section 5.1.1, the automated calibration processes will then use the end-points from the structured light pattern that intersects the calibration target. To determine these endpoints, a process must be put into place that isolates this part of the structured light pattern.

Fortunately, the foreground structured light line segment is easily determined with the use of the guidelines for validating calibration samples as proposed in Section 3.5.2 and 3.5.3. Since a valid calibration sample requires three segments: background – foreground – background, the middle segment must always be the foreground segment. Once this rule has been established, an iterative process is initiated that identifies the number of segments and the pixels that belong to that segment. Sweeping horizontally across the image per vertical scan-line, the first redline filtered pixel is identified as part of the first line segment in the image, thus a background segment. Any other redline pixels that are within a three-pixel radius are also determined as part of the segment. Any other redline pixels that are beyond the boundary of the three-pixel radius are identified as part of another segment. The entire filtered image is then processed following these same steps, shown in Figure 5-4, thus resulting in an image where the number of segments are determined and where pixels within the image belong to each segment. If the number of segments detected does not coincide with three expected segments, then the calibration-sampled image is marked as invalid and another sample is taken. If after five trials the process is unable to extract the desired pattern, it is then determined that the position of the multi-modal system in relation to the calibration target cannot produce the calibration sample and another calibration sample is taken at a different position and so forth. Figure 5-5 shows the end result of this process with the endpoints of the foreground pattern highlighted. These endpoints are then used to calculate the calibration matrices for structured lighting and extrinsic stereovision.

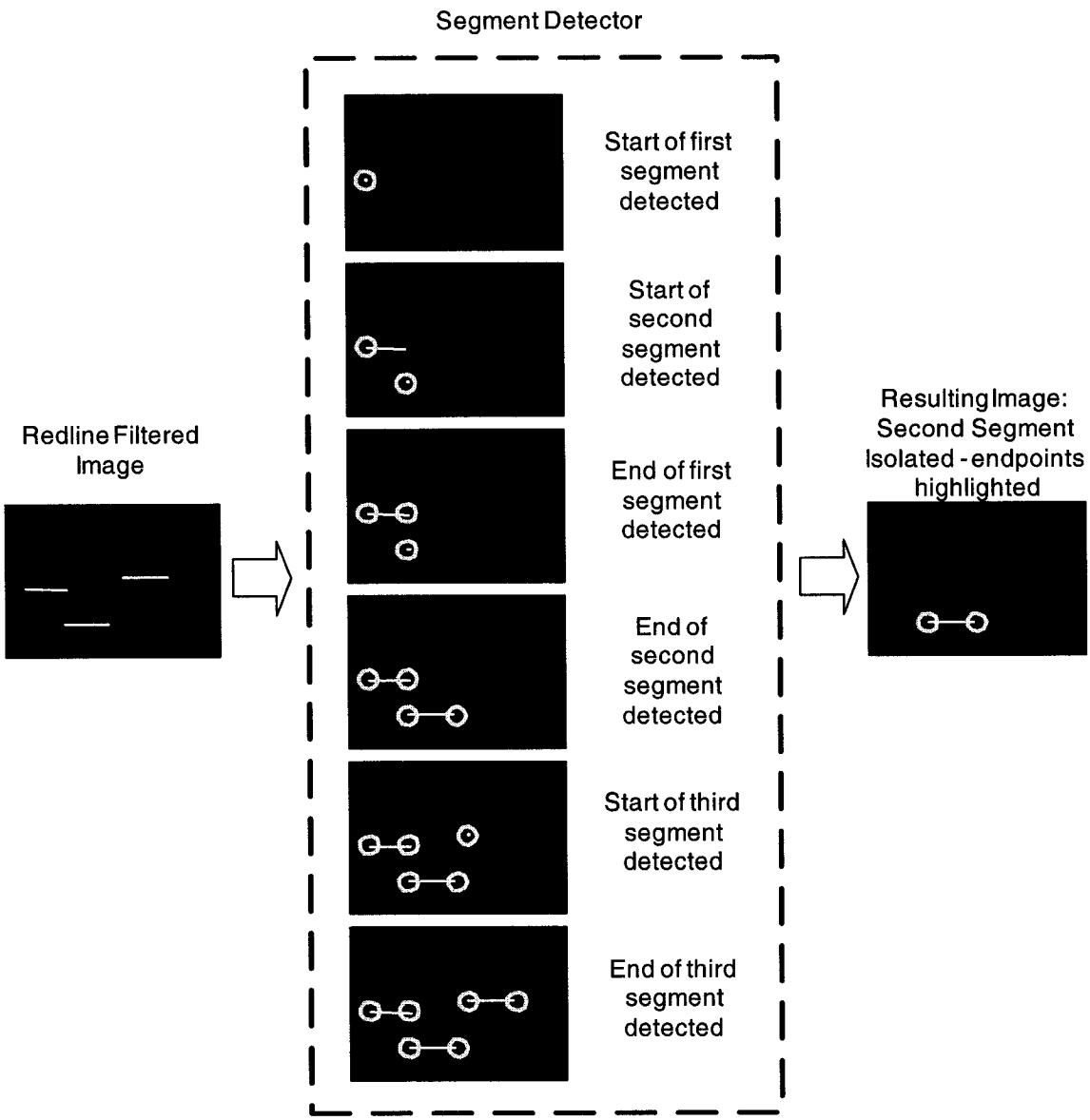

Figure 5-4 – Segment detection system

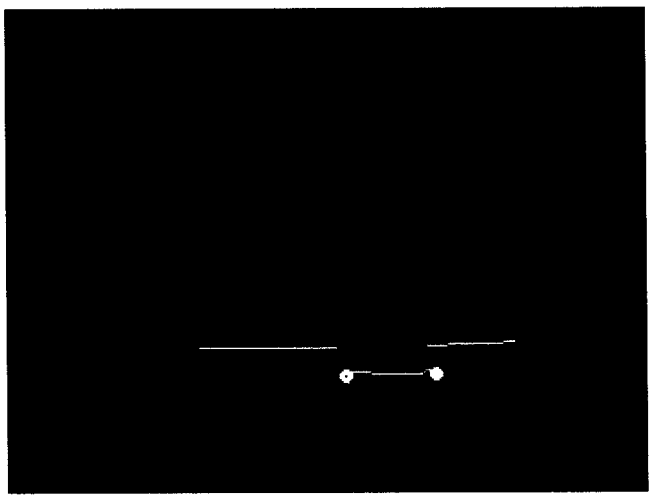

Figure 5-5 – Foreground structured light pattern isolated - endpoints highlighted

5.1.3 Automated Calibration Sampling Procedure

As briefly mentioned in Section 3.4.2, the CRS-F3 robotic system cradles the multi-modal system at its end-effector and can place the system in various positions within the workspace or within a designated calibration area. Auto calibration process of the multi-modal system is then properly controlled. A series of positions are programmed to the CRS-F3 robotic system, which in turn define the path positions where the multi-modal system samples calibration points. An example of such a sampling path consists of positioning the laser range finder at a constant height from the base of the robot and moving to various latitude and longitude positions, as shown in Figure 5-6, to produce a grid-like calibration path.

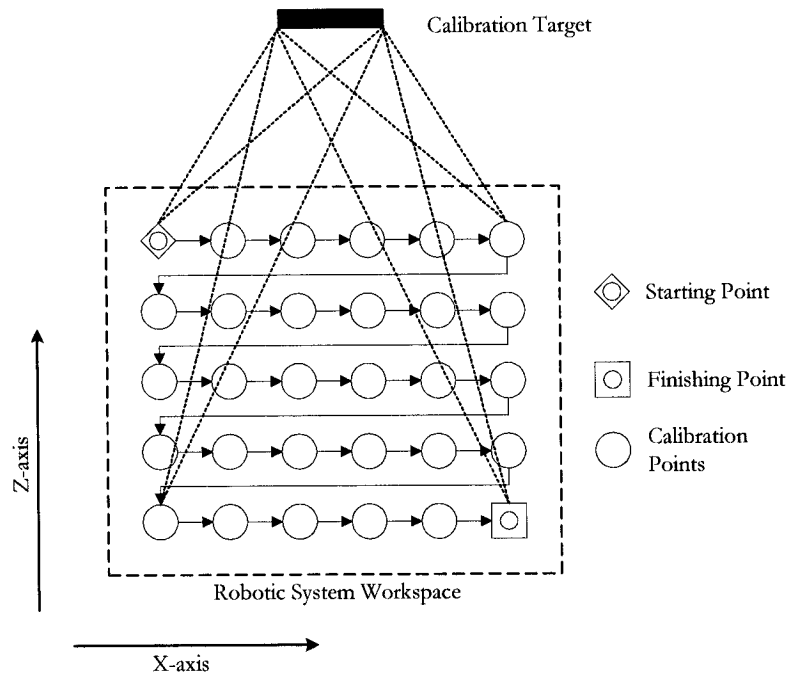

Figure 5-6 – Auto calibration path and sampling positions

Although the positioning of the robotic end-effector is not crucial during the multi-modal system calibration, its precision positioning allows itself to validate calibration results. As the system samples key features from the calibration target, a relative validation of each calibration sampled point is performed using the knowledge of the initial calibration position of the multi-modal system to the target and its displacement from the next sampling grid position as shown in Figure 5-7. At each new calibration position the distance to the same feature point is denoted as $\vec{u}' = \vec{u} - \Delta\vec{d}$ where $\vec{u}$ is the determined distance of the feature point from the last sampled position and $\Delta\vec{d}$ is the distance from the last calibration position and the new calibration position provided by the robot controller.

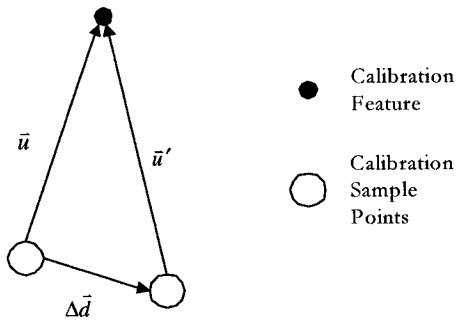

Figure 5-7 – Feature relationship between end-effector calibration sample positions

During the calibration process, if the relationships between each calibration sample produce diverging results such that the feature is logically much further or closer than expected, the calibration process is easily determined as faulty and the automated calibration procedure is re-initiated again to acquire new samples.

5.2 Experimental Results

To validate the operation of our prototype of the multi-modal range sensor, sequences of images were taken by the sensor on various objects following intra- and inter-calibration of the system using the proposed approach. A series of horizontal scan lines was taken in a non-textured environment after M^2S-SSL was subjected to 20 calibrations positions following a calibration path similar to that as shown on Figure 5-6. The proposed M^2S-SSL automated calibration procedure takes approximately 10 minutes to perform using 20 different calibration path points. This provides a total of 40 structured light sampling points for each subsystem and 40 sampling points for stereovision and laser range finder inter-subsystem calibration. Results from structured light and extrinsic stereovision to laser range finder calibration can be found in Table B.2 of Appendix B and intrinsic and extrinsic camera parameters are found in Table A.2 of Appendix A.

The following figures from Figure 5-8 to Figure 5-37 present experimental results for scenes of various complexities and objects with different reflectance characteristics, namely: two sets of indented planar surfaces, a rectangular vase whose rim is slightly tilted, a chair made of sponge, a tractor toy, and a wooden house frame respectively. The results present the distribution of three-dimensional points collected on these objects after calibration of the prototype of the multi-modal sensor. Each dataset from structured lighting systems (left and right), the laser range finder and the stereovision subsystem have been transformed such that the views are seen from the same perspective. Scans were performed at a granularity of 5 mm between successive scan lines of the laser range finder.

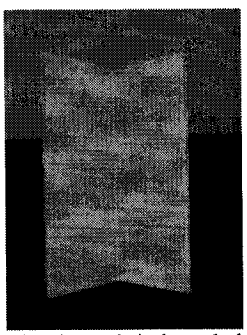

Figure 5-8 – Two inwards indented planar surfaces

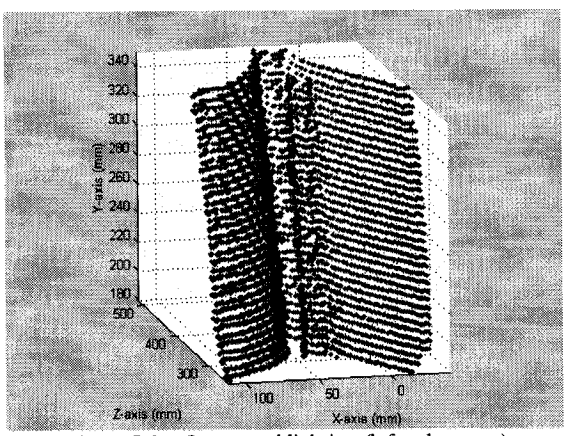

Figure 5-9 – Structured lighting (left subsystem) reconstruction of inwards indented planar surfaces

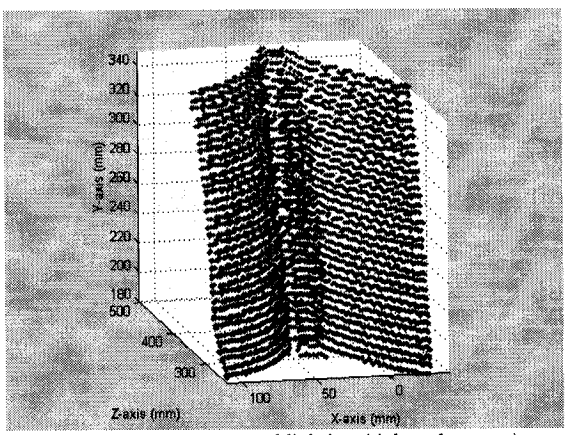

Figure 5-10 – Structured lighting (right subsystem) reconstruction of inwards indented planar surfaces

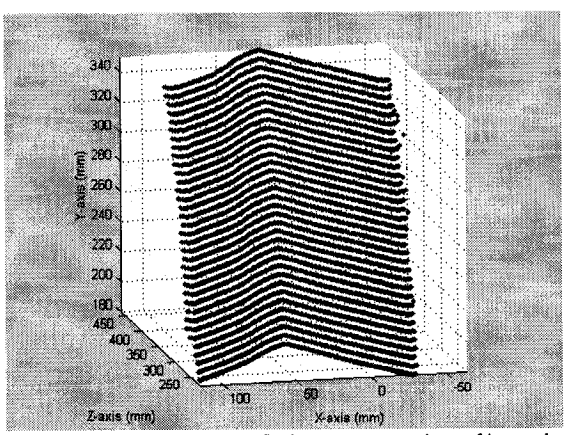

Figure 5-11 – Laser range finder reconstruction of inwards indented planar surfaces

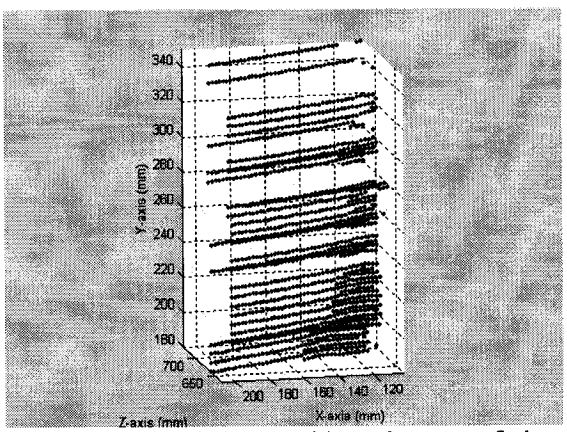

Figure 5-12 – Extrinsic stereovision to laser range finder reconstruction of inwards indented planar surfaces

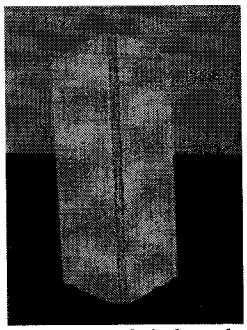

Figure 5-13 – Two outwards indented planar surfaces

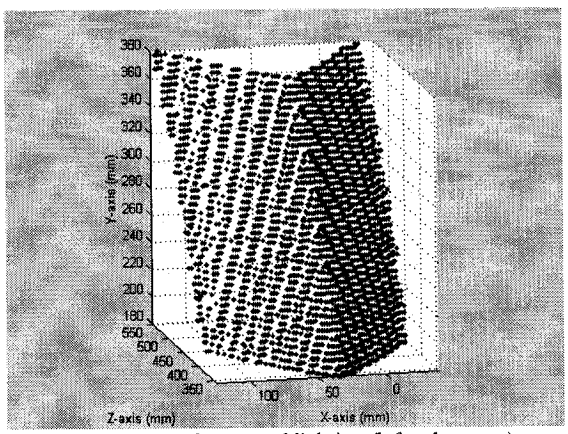

Figure 5-14 – Structured lighting (left subsystem) reconstruction of outwards indented planar surfaces

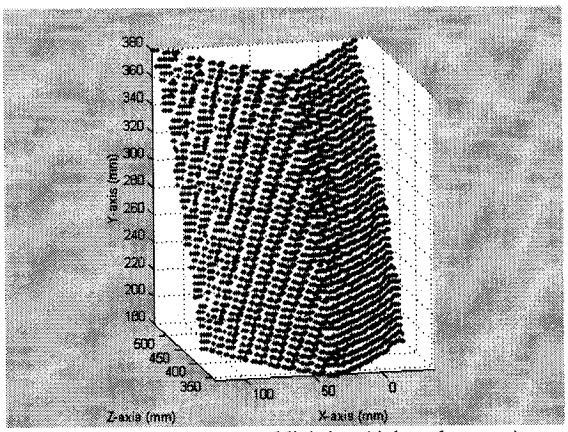

Figure 5-15 – Structured lighting (right subsystem) reconstruction of outwards indented planar surfaces

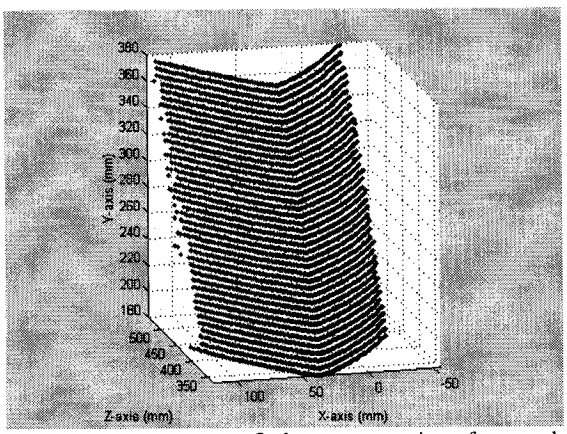

Figure 5-16 – Laser range finder reconstruction of outwards indented planar surfaces

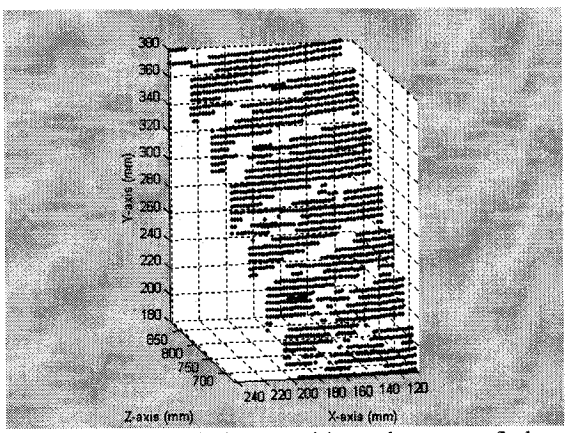

Figure 5-17 – Extrinsic stereovision to laser range finder reconstruction of outwards indented planar surfaces

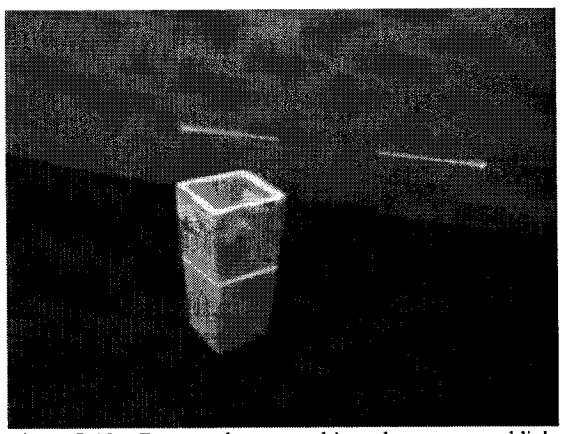

Figure 5-18 – Rectangular vase subjected to structured light

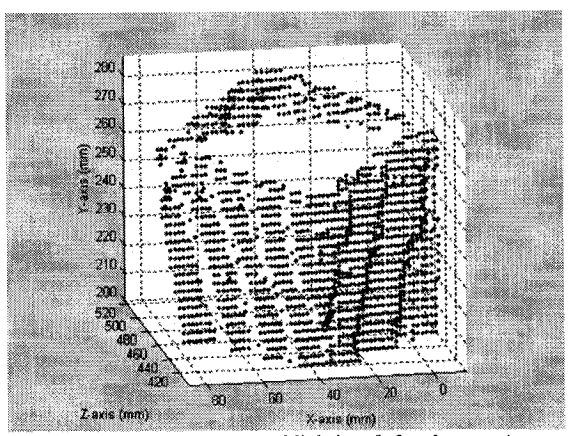

Figure 5-19 – Structured lighting (left subsystem) reconstruction of rectangular vase

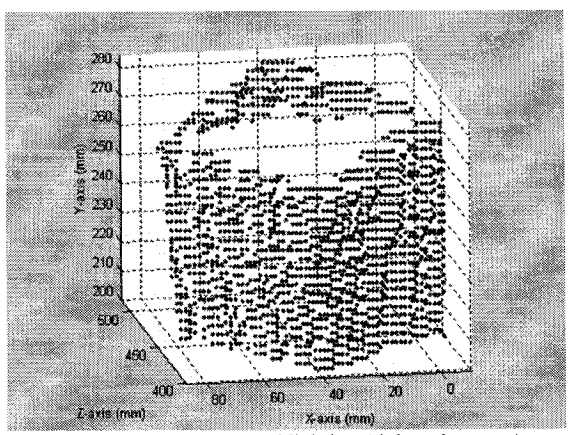

Figure 5-20 – Structured lighting (right subsystem) reconstruction of rectangular vase

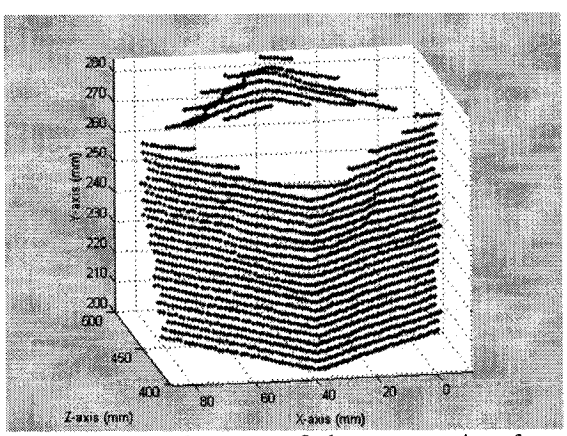

Figure 5-21 – Laser range finder reconstruction of rectangular vase

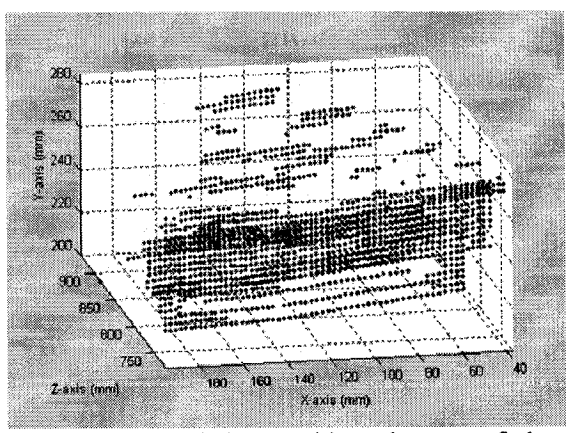

Figure 5-22 – Extrinsic stereovision to laser range finder reconstruction of rectangular vase

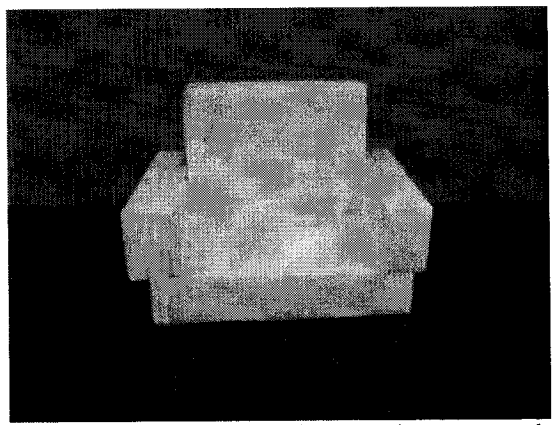

Figure 5-23 – Sponge chair covered with white paper in non-textured environment

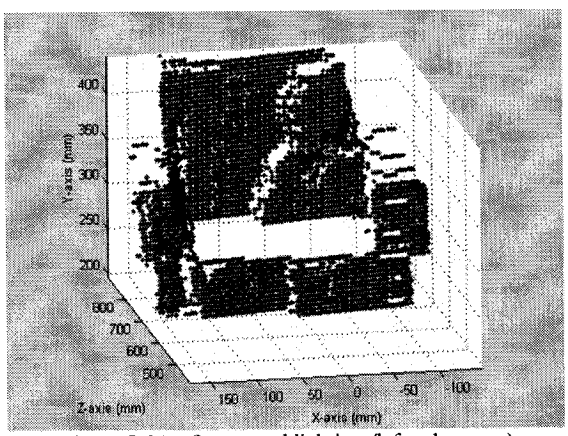

Figure 5-24 – Structured lighting (left subsystem) reconstruction of sponge chair

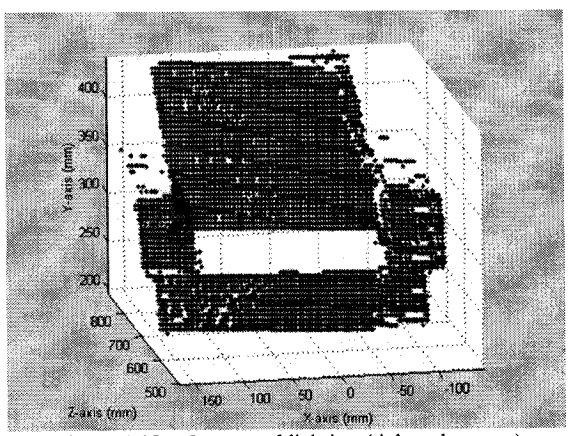

Figure 5-25 – Structured lighting (right subsystem) reconstruction of sponge chair

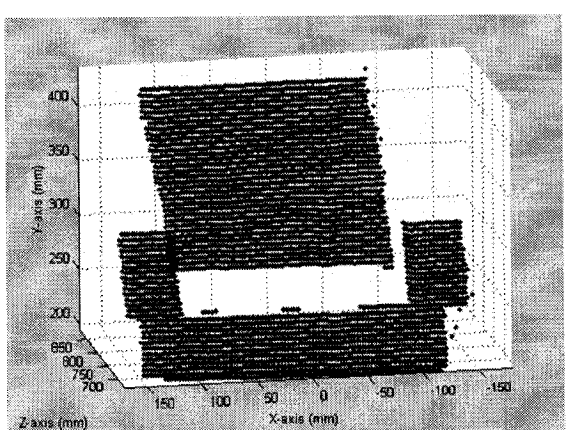

Figure 5-26 – Laser range finder reconstruction of sponge chair

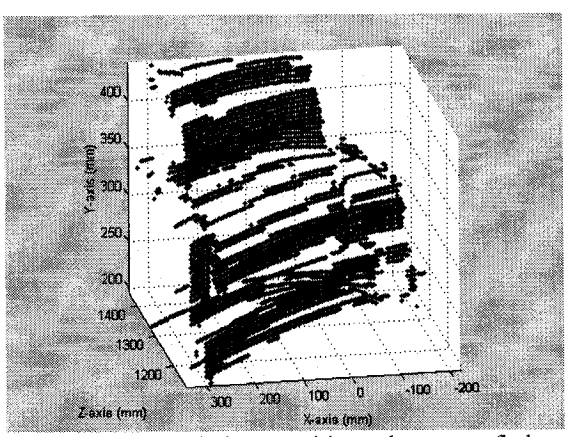

Figure 5-27 – Extrinsic stereovision to laser range finder reconstruction of sponge chair

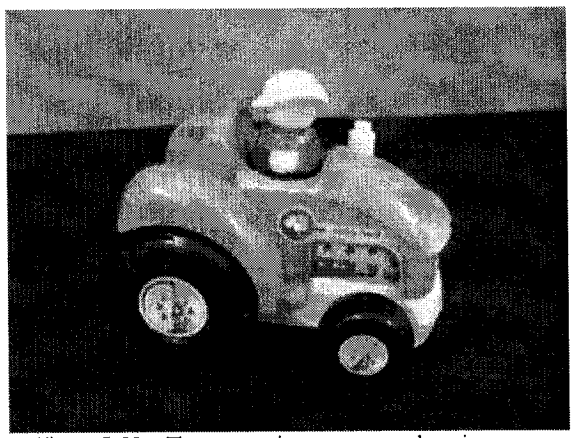

Figure 5-28 – Toy tractor in non-textured environment

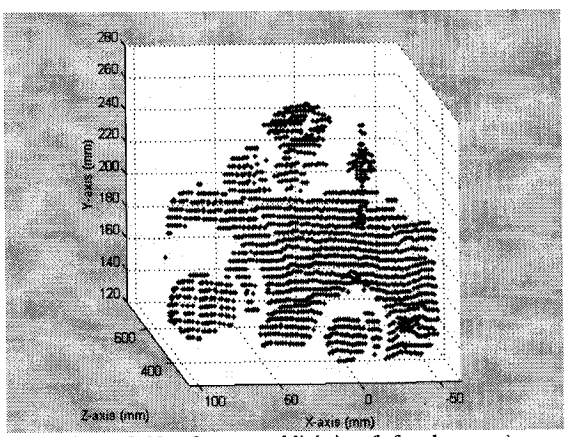

Figure 5-29 – Structured lighting (left subsystem) reconstruction of toy tractor

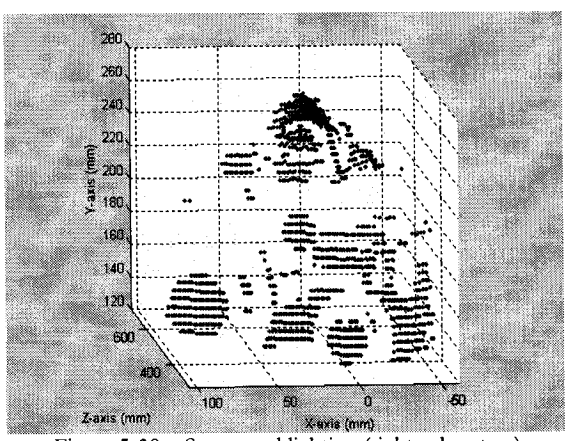

Figure 5-30 – Structured lighting (right subsystem) reconstruction of toy tractor

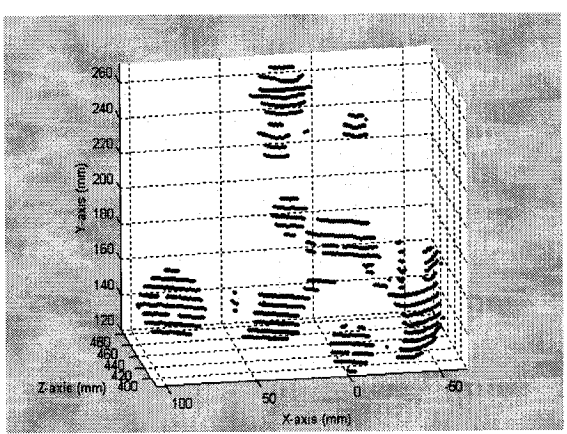

Figure 5-31 – Laser range finder reconstruction of toy tractor

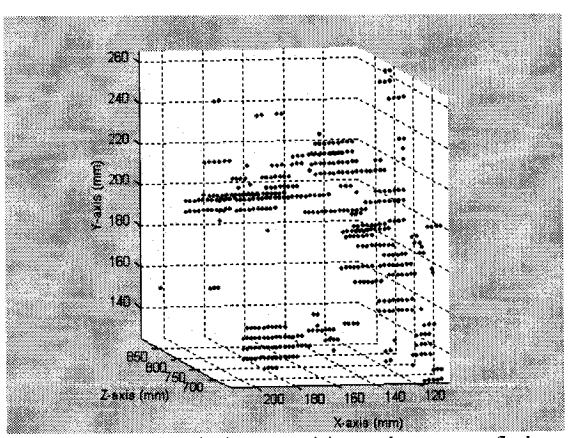

Figure 5-32 – Extrinsic stereovision to laser range finder reconstruction of toy tractor

Figure 5-33 – Wooden house frame in non-textured environment subjected to structured light

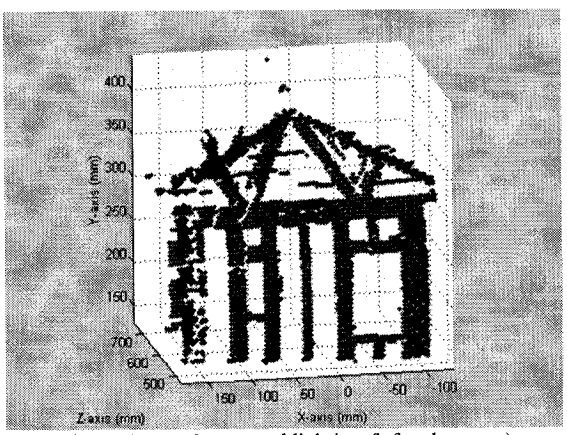

Figure 5-34 – Structured lighting (left subsystem) reconstruction of wooden house frame

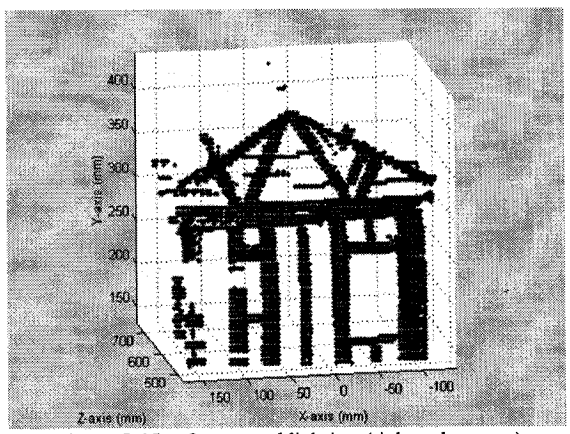

Figure 5-35 – Structured lighting (right subsystem) reconstruction of wooden house frame

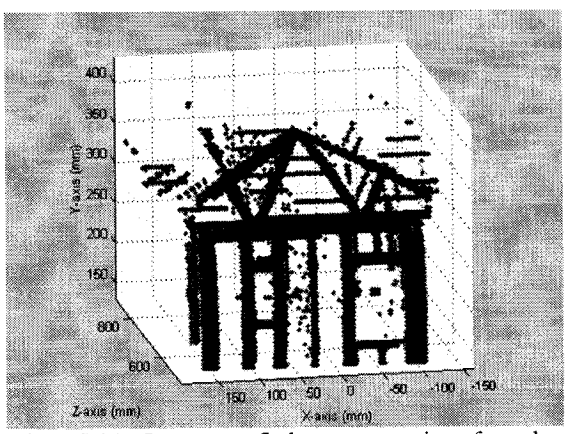

Figure 5-36 – Laser range finder reconstruction of wooden house frame

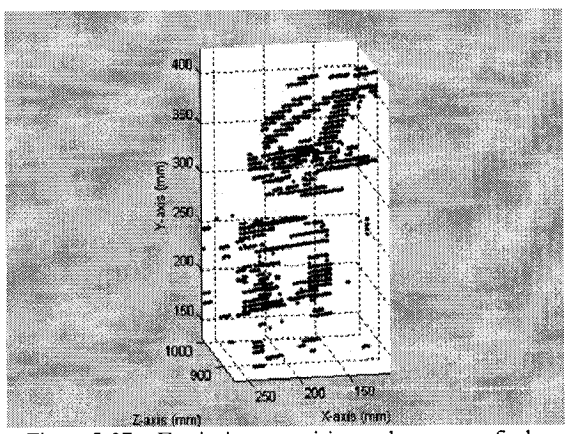

Figure 5-37 – Extrinsic stereovision to laser range finder reconstruction of wooden house frame

From visual inspection of the datasets, the difference in precision and accuracy provided by each system is noticeable. For example, the scans resulting from the stereovision system identify the vase but do not clearly distinguish that the vase is positioned such that its corner edge is the closest to the scanner. In the example of the wooden house, the disparity algorithm detects only the left side of the house.

The poor quality of the stereovision reconstruction is due to the weaknesses of the disparity algorithm for stereoscopic vision systems proposed by Birchfield *et al.* [38] as discussed in Section 2.2.5 that has been used in our implementation. This disparity algorithm focuses on accurately determining the location of discontinuities based upon threshold intensities and is smoothed out in regions where intensities are consistent. Thus if a complex surface object with little or no pixel dissimilarity is scanned, the algorithm fails to distinguish contour features of the object. An example of this limitation can be seen in Figure 5-22 where the edge contour of the vase is detected as a flat surface. In some scanning scenarios, objects that have little pixel dissimilarity to the background are unable to be detected and are treated as a background object. One feature that is lacking in the Birchfield *et al.* algorithm is the precision of the disparity value. As seen in most of the point cloud reconstruction results from stereovision, most of the objects are detected as being much further away from the system. This would imply that the disparity value is much lower than expected (recall: increasing distance of the object results in decreasing disparity value).

On the other hand, the results obtained after calibration from both structured lighting subsystems, consistently provide datasets similar to each other and are visually comparable to the high accuracy measurements of the laser range finder. These results demonstrate the validity and the accuracy that can be achieved with the proposed calibration scheme. Such an example can be seen from the results presented in Figure 5-29 on the tractor toy where the left structured lighting system found more features in the rear wheel area than the laser range finder. This phenomenon mainly results from the fact that this part of the object has a relatively high reflectivity and an orientation that made the laser beam diverge from the

sensor. In comparison between left and right structured lighting system, the left structured lighting system benefited from the positioning of the ambient lighting that allowed the feature structured light pattern to be easily detected.

The results of the reconstructed tractor exemplify a scenario where the laser range finder would not provide the best set of results among the other modalities. What this particular result proves is that a multi-modal range sensor is far more beneficial where one modality in the system could complement lacking data from other modalities.

Experimentation conducted during the development of the multi-modal sensor provided an opportunity to observe the sensitivity of the various modes of acquisition. For example, stereovision was revealed to be highly dependent upon an illuminated environment, which ensures that the scan target is detected and distinguished from other foreground and background objects or was able to distinguish varying intensities in the field of view. Thus a foreground object that blends to its background, similarly to a chameleon effect, becomes extremely difficult for the stereovision system to distinguish any significant intensity alterations. Opposite to stereovision, structured lighting can easily determine foreground to background objects regardless of variant intensity. The structured light pattern, the laser stripe, becomes a distinguishable active feature on the scene. With the use of thresholds, filters, and isolation algorithms of the stripe pattern from the rest of the scene, the structured light system is capable of object/scene reconstruction with limited dependency on lighting from its surroundings. Thus, a dimmed light setting was found to be the most suitable environment for this range sensor to perform. These two contrasting environmental conditions provide a supplementary challenge if simultaneous calibration is desired. This new condition becomes an important factor in validating the entire multi-modal calibration process. The question arises as to how is it possible to accommodate diverse lighting conditions simultaneously for two separate subsystems. Unfortunately, there is no clear-cut solution for this dilemma, but an attempt to rectify this problem is endeavoured.

For the purpose of development, the proposed multi-modal range sensor system was placed in an accommodating environment where each subsystem can fully operate and can correlate features between each other. Lighting is adjusted such that stereovision can easily differentiate textured objects and where the structured lighting and laser range finder system accurately extracts a structured light pattern without over-saturated lighting. To establish what environment lighting is sufficient, a trial-and-error method has been developed that consists of placing the triangular calibration target directly in front of the multi-modal system. A simple test procedure requiring user intervention is needed to determine whether the laser range finder can detect a single foreground object, while both the striped structured light segments can be extracted by the structured light subsystems, and the stereovision can also compute disparity from the calibration target.

This procedure starts off with placing the multi-modal calibration target directly within the shared field of view of each subsystem in the multi-modal system. The structured light pattern is projected on the scene and the laser range finder and structured light system sample the scene and attempt to extract the active features. Equivalent to calibration sampling of structured lighting calibration and extrinsic stereovision to laser range finder calibration the expected "background – foreground – background" pattern is isolated. If the active triangulation and stereovision systems cannot determine that there are three segments of the projected pattern then the user must intervene and alter the environmental lighting. For our purposes the proposed multi-modal system was placed in a room with a variable light dimmer where it was possible to brighten or attenuate the surrounding light.

As discussed in our simulation attempts, to determine how effectively the multi-modal system accurately reconstructs feature points within its field of view an *a priori* known object is placed at incremental distances away from the multi-modal system and its edges act as feature points to each subsystem. As the length of the *a priori* known object has already been measured, it is then possible to compare the actual length of the object to the reconstructed length of the object, following eq. (68). Figure 5-38 depicts the envelope surmounting errors, similar to

those performed in simulation in Section 4.4, from structured lighting and stereovision subsystems as the distance between the object and the multi-modal system varies.

$$PercentageError = \frac{\left| L_{actual} - L_{experimental} \right|}{L_{actual}} \cdot 100\% \tag{68}$$

where:

$L_{experimental}$ = Length of object from subsystem three-dimensional reconstruction

L_{actual} = Actual length across the *a priori* known object

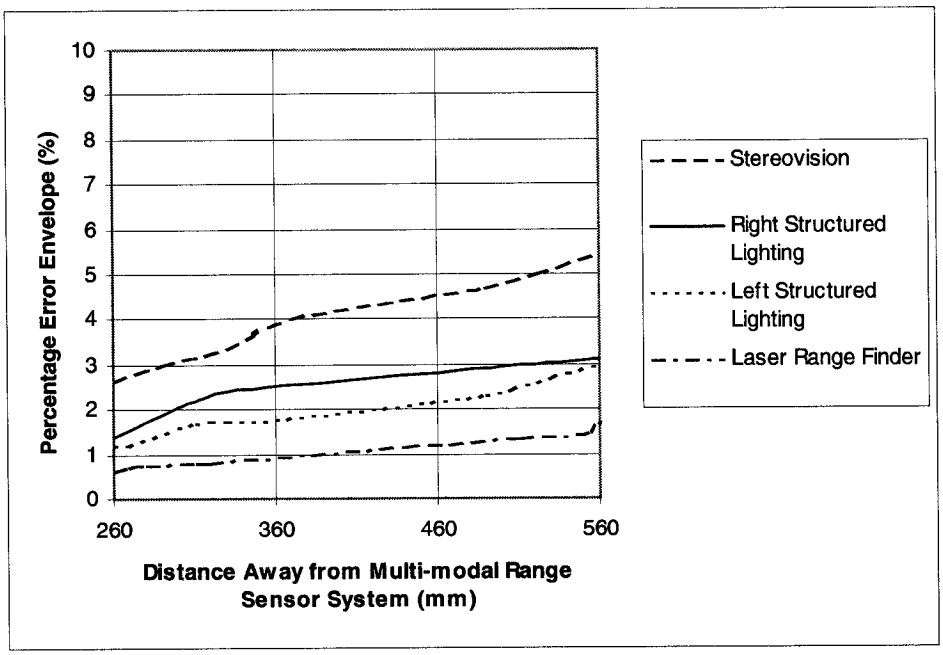

Figure 5-38 – Multi-modal subsystem percentage error envelope

As expected, the laser range finder fared the best between the structured lighting and stereovision systems. What is interesting from these results is that the left structured lighting system scored slightly lower error results than that of the right structured lighting system.

From the relative position of the left and right cameras to the centre of the laser light projector, the left camera was placed further away than the right camera. As mentioned in Section 4.3, this supplementary distance away from the structured light projector improved the resolution of detecting illuminated features. Stereovision, as expected, reported a slightly higher error than all other subsystems, but followed the common incrementing error trend as the other subsystems, where the error increased as the object became increasingly distant to the multi-modal system and maintained a relative low error curve within the shared field of view.

5.3 Performance

Once the calibration phase is completed, the scanning procedure can be performed from any viewpoint given that the sensor assembly preserves the registration between sensing modalities. When using a legacy robotic sensing system that only relied on the laser range finder, the acquisition time of 64 scan lines from three different viewing areas took approximately three hours. Using the integrated robotic system including the multi-modal sensor to scan the same 64 scan lines from three different viewing areas, the acquisition time is drastically reduced to 30 minutes, which corresponds to about 9 seconds per scan line of 512 range points for the laser range finder. This reduction is due to the advantage of fully programmed sensor displacements and the precision of the integrated computer controlled solution. Moreover, four complementary streams of range data are made available simultaneously from a single scanning procedure performed within the same amount of time. Experimentation demonstrated that the main limitation on the scanning speed comes from the control loop of the robot manipulator currently used to move the multi-modal sensor and not from the sensor itself.

Chapter 6

CONCLUSION

6.1 Summary

In this thesis the notion of the improvement of a single modality to achieve more reliable and versatile range sensing was replaced by the concept of complementary systems using multiple modalities over a shared field of view. Several popular and classical modalities were discussed as well as their calibration and reconstruction methods and from this base a multi-modal system was constructed from three operational range sensors: stereovision, structured lighting and active triangulation laser range finder. The proposed system required a new calibration scheme that would enable the ability to reconstruct a scene from a common reference point. Classical calibration approaches such as those found for stereovision and structured lighting have been used and customized for the purposes of the proposed system and additional processes have been integrated to facilitate automation.

Prior to any actual development with real equipment, a simulation of each modality was generated and the proposed calibration process was put to rigorous testing. This step fortified the theoretical calibration process and presented results that later reflected in real system results. A prototype was then constructed and the same calibration process was used to calibrate the resulting multi-modal system. Additional processes required beyond the simulated environment were added to facilitate the identification and isolation of the structured light pattern.

With the additional processes and contributions added to the development of the multi-modal range sensor, the proposed system was proven to provide series of complementary data. To provide a qualitative analysis of the system, a testing mechanism was used to determine the accuracy of calibration and reconstruction by each modality within the system. Results from this test were found to be as expected and similar to the testing results obtained

from simulation. Experimentation put in evidence that strict requirements must be applied to ensure that calibration is performed in suitable conditions. Using the defined multi-modal calibration process with the proposed multi-modal range sensor, applications that were traditionally dependent upon single mode range sensors can be easily handled by multi-modal range sensors that provide supportive results, and improve the accuracy of the entire system and the robustness to environment conditions.

What makes the problem of multi-modal calibration and reconstruction difficult is that the problems from each modality are reflected in the multi-modal system. With each system's dependency and nominal environmental working conditions, all of these issues exist and have to be considered in the real world application. This thesis has demonstrated how each subsystem in the proposed multi-modal system cannot only provide complementary data during scene reconstruction, but how it can share complementary features during calibration. This automated calibration approach goes beyond the constraints of calibration of a single modal system where calibration reference points must be either determined manually or predicted.

Consequently, the new calibration approach has made the acquisition process of the integrated robotic solution more flexible and less arduous. It provides a significant reduction of the acquisition time over the operation of traditional integrated systems while the necessity of human intervention is no longer required and repeatability is inherently improved. The proposed fully automated calibration process of a multi-modal sensing system increases the accuracy of the fusion of each scan within a consistent three-dimensional dataset.

6.2 Contributions

This thesis has contributed an in-depth study of classical range sensing techniques and their theoretical tenets. The introduction of proposed calibration technique and multi-modal

sensing device will open future opportunities to the development of other innovative multi-modal systems. These contributions are outlined as follows and will be discussed in detail.

- Exploration and development of the concept of heterogeneous multi-modal sensing by the integration of sensing technologies typically used independently;

- Original automated intra- and inter-subsystem calibration procedure, free of human intervention and independent from *a priori* knowledge on calibration objects that fully takes advantage of the complementarities between data collected from each modality;

- Integration, implementation and extensive testing of a functional automated prototype of a multi-modal range sensor that has been used as an experimental test-bed for validating the proposed multi-modal calibration process;

- Evaluation and demonstration of the strengths and constraints inherited from the respective sensing modalities.

This thesis presented the fundamental tenets of range sensing techniques and their reconstructive abilities from many reliable sources, opening the floor with the concept of single modality systems. Chapter 2 analyzed the theoretical basis of classical stereovision and structured lighting elaborately describing their geometries and approaches of calibration and reconstruction. This work is crucial to understanding the limitations of each modality and the requirements necessary to build a working multi-modal system. It is important to understand that the classical approach of improving a single modality is far from "thinking outside of the box".

This led to the analysis of existing multi-modal systems that have appeared in the literature. A comparative study, in Chapter 3, of three multi-modal systems was presented to evaluate their physical infrastructure, their ability to acquire features, and their ability to present meaningful

reconstructive results. A new approach for multi-modal calibration that existing multi-modal systems have not considered was presented in hopes to reduce post-processing reconstruction. The steps of a successful multi-modal system construction were outlined as well as an extended improvement of classical techniques that suit the multi-modal environment.

The premise of the work presented in this thesis has also contributed two conference papers at the IEEE International Instrumentation and Measurement Technology Conference (IMTC) in May 2005. The first conference paper, entitled "Calibration of a multi-modal 3D scanner" [66] discusses the implementation of multi-modal calibration on the VIVA M^2S-SSL prototype. The second paper, entitled "An integrated robotic multi-modal range sensing system" [7] introduces the system integration of the multi-modal sensor with the robotic arm interface and the firmware and software design that operate the systems together.

To establish that the stated steps are legitimate processes, Chapter 4 presented a simulated model using classical range sensing techniques introduced in Chapter 2. Using the proposed steps as a multi-modal construction guideline, the simulated multi-modal system was then used to reconstruct synthetic models and its performance has been evaluated. Its simulated success warranted the real creation of the proposed multi-modal system following the same multi-modal construction guideline. Once again, the system was evaluated and reconstruction results have been presented to indicate that a multi-modal system with effortless calibration procedures can provide improved results to that of a single modal system or a multi-modal system that utilizes data fitting.

6.3 Future Work

Immediate work under investigation is to provide a complete registration solution and scanning system utilizing the full potential of the CRS Robotic system. This would enable real-time data acquisition and modelling within the entire robotic workspace in any reachable

pose by the F3 manipulator. In addition, the use of other stereo disparity algorithms in conjunction with the current Birchfield *et al.* algorithm is being considered to improve performance of the stereovision modality.

It was determined during calibration of the proposed system that the contrasting environment required for structured lighting and extrinsic stereovision to laser range finder calibration proved to be quite troublesome. Further investigation must be performed to eliminate the need of user intervention for fine-tuning the environmental lighting for both systems to operate in tandem. Additional image processing may be required to either brighten or darken the images such that the appropriate features of the structured light can be detected and the calibration target is easily discerned from the rest of the environment.

Further work under investigation is the ability to amalgamate the various datasets produced by the multi-modal system into a single dataset, which advertently provides the statistically optimal representation of the scanned scene. Other improvements considered would be to provide additional information of colouring, texturing, and other essential attributes otherwise not detected by single mode range sensors. Such an enhancement to the multi-modal range sensing system will lead to a more sophisticated model of the captured environment. Finally, the data are to be used to create an occupancy model, in order to serve in higher-level robotic applications such as path planning and task completion.

Appendix A

CAMERA PARAMETERS FOR SIMULATIONS AND EXPERIMENTS

Table A.1 and Table A.2 list the intrinsic and extrinsic parameters used by the simulated multi-modal environment and by the Sony XC-999 camera system and VRex CAM-3000C. Third party camera calibration software has been used to calibrate the real multi-modal equipment and its intrinsic parameters have been propagated into the Matlab multi-modal simulated environment as the stereovision model. The parameters are defined as follow:

f_l: Focal length (mm)

(x_O, y_O): Principal point (pixels)

$(f_l / a_x, f_l / a_y)$: Effective pixel sizes (mm)

R: Camera Rotation Matrix

T: Camera Translation

	Left Camera	**Right Camera**
(x_O, y_O):	(156.70, 122.82)	(151.70, 120.74)
$(f_l / a_x, f_l / a_y)$:	(322.08, 320.78)	(321.98, 320.83)
R:	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$
T:	$\begin{bmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \end{bmatrix}$	$\begin{bmatrix} 26.0000 \\ 0.0000 \\ 0.0000 \end{bmatrix}$

Table A.1 – Parameters of synthetic camera system used in the simulation experiments

	Left Camera	Right Camera
(x_O, y_O):	(156.70, 122.82)	(151.70, 120.74)
$(f_l / a_x, f_l / a_y)$:	(322.08, 320.78)	(321.98, 320.83)
R:	$\begin{bmatrix} 1.0000 & 0.0000 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 \\ 0.0000 & 0.0000 & 1.0000 \end{bmatrix}$	$\begin{bmatrix} 0.9986 & 0.0527 & 0.0020 \\ -0.0527 & 0.9986 & -0.0106 \\ -0.0026 & 0.0105 & 0.9999 \end{bmatrix}$
T:	$\begin{bmatrix} 0.0000 \\ 0.0000 \\ 0.0000 \end{bmatrix}$	$\begin{bmatrix} 53.6758 \\ 0.7521 \\ -5.5077 \end{bmatrix}$

Table A.2 – Parameters of synthetic camera system used in the real image experiments

Appendix B

SIMULATED AND REAL CALIBRATION RESULTS FOR STRUCTURED LIGHTING AND EXTRINSIC STEREO TO LASER RANGE FINDER CALIBRATION

Here lists the calibration matrices determined by multi-modal calibration for intrinsic structured lighting and extrinsic stereovision to active triangulation laser range finder calibration. Table B.1 refers to results determined in the multi-modal Matlab simulated environment and Table B.2 refers to results determined using real equipment and calibration targets. The parameters are defined as follow:

	Calibration Matrix
Structured lighting (left camera) calibration matrix as defined in eq. (39)	$\begin{bmatrix} -0.0791 & 0.0785 & 3.2469 \\ 0.0000 & 0.0000 & 0.0000 \\ -0.0002 & 0.0089 & -25.9716 \\ 0.0000 & -0.0082 & 1.0000 \end{bmatrix}$
Structured lighting (right camera) calibration matrix as defined in eq. (39)	$\begin{bmatrix} -0.0790 & -0.1270 & 28.0691 \\ 0.0000 & 0.0000 & 0.0000 \\ -0.0009 & 0.0075 & -25.6417 \\ 0.0000 & -0.0082 & 1.0000 \end{bmatrix}$
Extrinsic stereovision to laser range finder calibration matrix: as defined in eq. (65)	$\begin{bmatrix} 9.6405 & -9.6021 & -0.1538 \\ -0.0036 & -1.1604 & 9.4554 \\ 0.0000 & 1.0000 & 0.0000 \end{bmatrix}$

Table B.1 – Multi-modal calibration matrices determined in simulation experiments

	Calibration Matrix
Structured lighting (left camera) calibration matrix as defined in eq. (39)	$\begin{bmatrix} -0.7156 & -0.0824 & 125.6671 \\ 0.0000 & 0.0000 & 0.0000 \\ 0.0218 & 0.2642 & -264.4483 \\ -0.0002 & -0.0074 & 1.0000 \end{bmatrix}$
Structured lighting (right camera) calibration matrix as defined in eq. (39)	$\begin{bmatrix} -0.7312 & -0.5198 & 179.5097 \\ 0.0000 & 0.0000 & 0.0000 \\ -0.0643 & 0.2801 & -267.0991 \\ 0.0001 & -0.0078 & 1.0000 \end{bmatrix}$
Extrinsic stereovision to laser range finder calibration matrix: as defined in eq. (65)	$\begin{bmatrix} 0.1271 & 1.2234 & -0.0002 \\ -0.9243 & 0.1096 & 1.3602 \\ 0.0118 & -0.0002 & -0.0001 \end{bmatrix}$

Table B.2 – Multi-modal calibration matrices determined in real image experiments

BIBLIOGRAPHY

[1] National Aeronautical and Space Administration – Jet Propulsion Laboratory, *NASA Facts – Mars Pathfinder*, California Institute of Technology, Pasadena California, Sep. 10, 2004. http://www.jpl.nasa.gov/news/fact_sheets/mpf.pdf.

[2] R.O. Duda, D. Nitzan, and P.Barret, "Use of range and reflectance data to find planar surface regions", in *IEEE Trans. on Pattern Analysis Machine Intelligence*, vol. PAMI-1, no. 3, pp. 259-271, 1979.

[3] R.A. Jarvis, "A laser time-of-flight range scanner for robotic vision", in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. PAMI-5, no. 5, pp. 505-512, 1983.

[4] K.S. Fu, R.C. Gonzalez, and C.S.G. Lee, *Robotics: Control, Sensing, Vision, and Intelligence*, New York, NY: McGraw-Hill, 1987.

[5] U.R. Dhond and J.K. Aggarwal, "Structure from stereo – a review", in *IEEE Trans. on Systems, Man and Cybernetics*, vol. 19, issue 6, pp. 1489-1510, Nov.-Dec. 1989.

[6] S.D. Cochran and G. Medioni, "3-D surface description from binocular stereo", in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 981-994, 1992.

[7] P. Curtis, C.S. Yang, and P. Payeur, "An integrated robotic multi-modal range sensing system", in *Proc. of the IEEE International Instrumentation and Measurement Technology Conference*, Ottawa, ON, pp. 1991-1996, May 2005.

[8] M. Hebert, "Active and passive range sensing for robotics", in *Proc. of the 2000 IEEE International Conference on Robotics & Automation*, San Francisco, California, vol. 1, pp. 102 – 110, 24-28 Apr. 2000.

[9] E. Trucco and A. Verri, Introduction Techniques for 3-D Computer Vision, Prentice Hall, 1998.

[10] A. Bovik, Handbook of Image and Video Processing, Academic Press, 2000.

[11] H. Malm, and A. Heyden, "Stereo head calibration from a planar object", in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001*, vol. 2, pp. II-657 - II-662, 2001.

[12] Z. Zhang, "Camera calibration with one-dimensional objects", in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, issue 7, pp. 892-899, Jul. 2004.

[13] J-S. Lee and Y-H. Lee, "CCD camera calibrations and projection error analysis", in *Proc. of the 4th Korea-Russia International Symposium on Science and Technology. KORUS 2000*, vol. 2, pp. 50-55, 27 Jun.-1 Jul. 2000.

[14] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses", in *IEEE Journal of Robotics and Automation*, vol. 3, issue 4, pp. 323-344, Aug. 1987.

[15] T. Kawakami, M. Endo, and T. Iwasaki, "Adaptive multifrequency modulation method for an advanced laser range finder", in *IEEE Trans. on Instrumentation and Measurement*, vol. 43, issue 6, pp. 857-860, Dec. 1994.

[16] C. Archibald, E. Petriu, and A. Harb, "Robot skills development using a laser range finder", in *IEEE Trans. on Instrumentation and Measurement*, vol. 43, issue 2, pp. 265-271, Apr. 1994.

[17] Y.D. Chen and J. Ni, "Dynamic calibration and compensation of a 3D laser radar scanning system", in *IEEE Trans. on Robotics and Automation*, vol. 9, issue 3, pp. 318-323, Jun. 1993.

[18] P. Besesty, P. Labeye, and P. Thony, "Compact FMCW advanced laser rangefinder", in *Summaries of Papers Presented at the Conference on Lasers and Electro-Optics*, pp. 552-553, 23-28 May 1999.

[19] G. Bazin and B. Journet, "A new laser range-finder based on FMCW-like method", in *Proc. of the IEEE Instrumentation and Measurement Technology Conference. IMTC-96. 'Quality Measurements: The Indispensable Bridge between Theory and Reality'*, vol. 1, pp. 90-93, 1996.

[20] D. Dupuy and M. Lescure, "Improvement of the FMCW laser range-finder by an APD working as an optoelectronic mixer", in *IEEE Trans. on Instrumentation and Measurement*, vol. 51, issue 5, pp. 1010-1014, Oct. 2002.

[21] M.D. McNeill, L. Williams, and C. HuaMeng, "Design of a time-of-flight range-finder", in *29th Annual Frontiers in Education Conference. FIE '99*, vol. 3, pp. 13D6/17-13D6/22, 10-13 Nov. 1999.

[22] D. Stoppa, L. Viarani, A. Simoni, L. Gonzo, and M. Malfatti, "A new architecture for TOF-based range-finding sensor", in *Proc. of IEEE Sensors*, vol. 1, pp. 481-484, Oct. 2004.

[23] G. Godin, J.-A. Beraldin, J. Taylor, L. Cournoyer, M. Rioux, S. El-Hakim, R. Baribeau, F. Blais, P. Boulanger, J. Domey, and M. Picard, "Active optical 3D imaging for heritage applications", in *IEEE Computer Graphics and Applications*, vol. 22, issue 5, pp. 24-35, Sep.-Oct. 2002.

[24] M. Rioux, F. Blais, J. Beraldin, and P. Boulanger, "Range imaging sensors development at NRC Laboratories", in *Proc. of the Workshop on Interpretation of 3D Scenes*, pp. 154-160, 27-29 Nov. 1989.

[25] J.-A. Beraldin, M. Rioux, F. Blais, G. Godin, and R. Baribeau, "Model-based calibration of a range camera", in *Proc. of the 11th IAPR Conference on Pattern Recognition Vol.1. Conference A: Computer Vision and Applications*, pp. 163-67, 30 Aug. -3 Sep. 1992.

[26] H. Khali, Y. Savaria, J.L. Houle, J.-A. Beraldin, F. Blais, and M. Rioux, "A VLSI chip for 3-D camera calibration", in *Canadian Conference on Electrical and Computer Engineering*, vol. 1, pp. 120-123, 5-8 Sep. 1995.

[27] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method", in *Int. J. of Computer Vision*, vol. 9, issue 2, pp. 137-154, Nov. 1992.

[28] I. Shimshoni, R. Basri, and E. Rivlin, "A geometric interpretation of weak-perspective motion", in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 21, issue 3, pp. 252-257, Mar. 1999.

[29] S. Christy and R. Horaud, "A quasi linear reconstruction method from multiple perspective views", in *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 'Human Robot Interaction and Cooperative Robots'*, vol.2, pp. 374-380, 5-9 Aug. 1995.

[30] C.J. Poelman and T. Kanade, "A paraperspective factorization method for shape and motion recovery", in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, issue 3, pp. 206-218, Mar. 1997.

[31] R. Horaud, S. Christy, F. Dornaika, and B. Lamiroy, "Object pose: links between paraperspecitve and perspective", in *Proc. Of the Fifth International Conference on Computer Vision*, pp. 426-433, 20-23 Jun. 1995.

[32] K. Mi-Hyun, and K-H. Sohn, "Edge-preserving directional regularization technique for disparity estimation of stereoscopic images", in *IEEE Trans. on Consumer Electronics*, vol. 45, issue: 3, pp. 804-811, Aug. 1999.

[33] C.L. Pagliari and T.J. Dennis, "Disparity estimation using edge-oriented classification in the DCT domain," in *IEEE Electronics Letters*, vol. 34, issue 12, pp.1214-1216, Jun. 1998.

[34] J. Karathanasis, D. Kalivas, and J. Vlontzos, "Disparity estimation using block matching and dynamic programming", in *IEEE International Conference on Electronics, Circuits, and Systems. ICECS '96.*, vol. 2, pp.728-731, 13-16 Oct. 1996.

[35] C-W. Lin; E-Y. Fei, and Y-C. Chen; "Hierarchical disparity estimation using spatial correlation", in *IEEE International Conference on Consumer Electronics. ICCE. 1998 Digest of Technical Papers*, pp. 130-131, 2-4 Jun. 1998.

[36] A. El Zaart, D. Ziou, and F. Dubeau, "Phase-based disparity estimation: a spatial approach", in *IEEE International Conference on Image Processing*, vol. 3, pp. 244-247, 26-29 Oct. 1997.

[37] T-Y. Chen, A.C. Bovik, and B.J. Super, "Multiscale stereopsis via Gabor filter phase response", *in IEEE International Conference on Systems, Man, and Cybernetics. 'Humans, Information and Technology'*, vol. 1, pp.55-60, 2-5 Oct. 1994.

[38] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo", in *Sixth International Conference on Computer Vision, 1998*, pp. 1073-1080, 4-7 Jan 1998.

[39] P. Saint-Marc, J. Jezouin, and G. Medioni, "A versatile PC-based range finding system", in *IEEE Trans. on Robotics and Automation*, vol. 7, issue 2, pp. 250-256, Apr. 1991.

[40] J. Maver, and R. Bajcsy, "Occlusions as a guide for planning the next view", in *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 15, issue 5, pp. 417-433, May 1993.

[41] D. Chen, W. Gao, and X. Chen, "A new approach of recovering 3-D shape from structure-lighting", in *3rd International Conference on Signal Processing. ICSP' 96*, vol. 2, pp. 839-842, 14-18 Oct. 1996.

[42] P. Lavoie, D. Ionescu, and E.M. Petriu, "3-D object model recovery from 2-D images using structured light", in *Proc. of the IEEE International Instrumentation and Measurement Technology Conference 1996. 'Quality Measurements: The Indispensable Bridge between Theory and Reality' IMTC-96*, vol. 1, pp. 377-382, 1996.

[43] R.B. Fisher, A.P. Ashbrook, C. Robertson, and N. Werghi, "A low-cost range finder using a visually located, structured light", in *Proc. of the Second International Conference on 3-D Digital Imaging and Modeling*, pp. 24-33, 4-8 Oct. 1999.

[44] E. Mouaddib, J.Batlle, and J. Salvi, "Recent progress in structured light in order to solve the correspondence problem in stereovision", in *Proc. of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 130-136, 20-25 Apr. 1997.

[45] T-Z. Shen and C-H. Menq, "Digital projector calibration for 3-D active vision systems", in *Journal of Manufacturing Science and Engineering Trans. of the American Society of Mechanical Engineers*, vol. 124, issue 1, pp. 126-134, Feb. 2002.

[46] E. Trucco and B. Fisher, "Acquisition of consistent range data using local Calibration", in *Proc. of the IEEE International Conference on Robotics and Automation*, vol. 4, pp. 3410-3415, 8-13 May 1994.

[47] C. Chen and A. Kak, "Modeling and calibration of a structured light scanner for 3-D robot vision", in *Proc. of the IEEE International Conference on Robotics and Automation*, vol. 4, pp. 804-815, Mar. 1987.

[48] G. Agin, "Calibration and use of a light stripe range sensor mounted on the hand of a robot", in *Proc. of the IEEE International Conference on Robotics and Automation*, vol. 2, pp. 680-685, Mar. 1985.

[49] D.Q. Huynh, "Calibration of a structured light system: a projective approach", in *Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 225-230, 17-19 Jun. 1997.

[50] I.D. Reid, "Projective calibration of a laser-stripe range finder", in *Image and Vision Computing*, vol. 14, issue 9, pp. 659-666, Oct. 1996.

[51] H.S.M. Coxeter, Projective geometry, Springer-Verlag, New York, Inc, New York, NY, 2003.

[52] F. Ayres, Schaum's Outline of Theory and Problems of Projective Geometry, Schaum Publishing Co. New York, NY, 1967.

[53] D. Laurent, M. El Mustapha, P. Claude, and V. Pascal, "A mobile robot localization based on a multisensor cooperation approach", in *Proc. of the IEEE IECON 22nd International Conference on Industrial Electronics, Control, and Instrumentation*, vol 1. pp. 155-160, 5-10 Aug. 1996.

[54] G. Dudek, P. Freedman, and I.M. Rekleitis, "Just-in-time sensing: efficiently combining sonar and laser range data from exploring unknown worlds", in *Proc. of the IEEE International Conference on Robotics and Automation*, vol. 1, pp. 667-672, 22-28 Apr. 1996.

[55] A. Elfes, "Using occupancy grids for mobile robot perception and navigation", in *IEEE Computer*, vol. 22, issue 6, pp. 46-58, Jun. 1989.

[56] J. Miura, Y. Negishi, and Y. Shirai, "Mobile robot map generation by integrating omnidirectional stereo and laser range finder", in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, vol. 1, pp. 250-255, 2002.

[57] A. Elfes, "Sonar-based real world mapping and navigation", in *International Journal of Robotics and Automation*, vol. 3, issue 3, pp. 249-265, Jun. 1987.

[58] VRex Stereoscopic 3D, *VR-MUX 2 Multiplexer*, Elmsford New York, Aug. 7, 2001. http://www.vrex.com/download/mux2.pdf.

[59] P. Curtis and P. Payeur, "An integrated robotic laser range sensing system for automatic mapping of wide workspaces", in *Canadian Conference on Electrical and Computer Engineering*, vol. 2, pp. 1135-1138, May 2004.

[60] Servo-Robot Inc., *Jupiter 3-D Laser Vision Camera Installation and Operation Manual*, St-Bruno, QC, Canada, 1996.

[61] Matrox Electronic Systems Ltd., *Matrox Image Library Version 6.1 User Guide*, Dorval, QC, March 1, 2000.

[62] W.M. Keck, Virtual Factory Lab, *Virtual Robot System version 2,* Technical report, School of Industrial and Systems Engineering, Georgia Institute of Technology, pp. 25-39, 2002.

[63] O.D. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, Cambridge, MA: MIT Press, 1993.

[64] R. Pless and Q. Zhang, "Extrinsic calibration of a camera and laser range finder", in *Proc. of the IEEE/RSJ Internal Conference on Intelligent Robots and Systems*, vol. 3, pp. 2301-2306, 28 Sep.-2 Oct. 2004.

[65] Intel Corporation, *Open Source Computer Vision Library – Reference Manual*, Santa Clara, CA, 2001.

[66] C.S. Yang, P. Curtis, and P. Payeur, "Calibration of a multi-modal 3D scanner", in *Proc. of the IEEE International Instrumentation and Measurement Technology Conference*, Ottawa, ON, pp. 865-870, May 2005.