

Torso-mounted Visual Servoing of an Outdoor Robotic Manipulator

by

Nima Ziraknejad

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
Master of Applied Science

in

THE FACULTY OF GRADUATE STUDIES
(Department of Electrical and Computer Engineering)

The University of British Columbia

April 2007

© Nima Ziraknejad, 2007

Abstract

In the majority of vision applications, sensor calibration is a prerequisite to proper use of the sensor for both measurement and control. The objective in camera calibration is to estimate a set of parameters to construct a mapping between the 3D position of a target point and its 2D image coordinates. This thesis presents an autonomous stereo camera calibration technique based on a pinhole camera model with applications in industrial outdoor visual servoing systems. An efficient iterative least-squares parameter estimation algorithm is used to estimate the camera model parameters. The obtained stereo camera model is used to estimate the pose of the target object during the robot servoing process. The heavy-duty stereo camera rig is installed on the torso of an outdoor 3DOF robotic manipulator.

The stereo camera calibration is entirely an autonomous process as the robot moves the calibration tool within its workspace and the stereo camera model is produced after the data collection process. The stereo cameras are treated as a single unit and a single transformation is obtained for the stereo camera pair in the system. The calibration process is fast, efficient and no human interaction is required during the process.

The developed vision system is capable of detecting the 3D positions of the objects within the robot workspace with an average error of 1.93cm (maximum error of 7.6cm) for a region equal and smaller than the region in which the camera calibration is performed. The errors are comprised of errors in camera calibration, robot positioning and image processing. The maximum error represents an error of 2.1% of the total robot workspace reach. The visual servoing tasks are being performed with an update rate of 29ms.

The complete visual servo system, used for pipe handling, has been implemented on a combined platform with two main processing units and a fully integrated 3DOF torso-mounted robotic manipulator. The processing units include an industrial PC104 embedded system with 900MHz Pentium III CPU for vision related tasks and an industrial Programmable Logic Control (PLC) unit for low level control of the robot.

Table of Contents

| | |
|--|-------------|
| ABSTRACT | ii |
| TABLE OF CONTENTS | iii |
| LIST OF FIGURES | vi |
| LIST OF TABLES | viii |
| ACKNOWLEDGEMENTS | ix |
| CHAPTER 1. OVERVIEW AND BACKGROUND | 1 |
| 1.1 INTRODUCTION | 1 |
| 1.2 PROBLEM STATEMENT AND MACHINE VISION SOLUTIONS | 2 |
| 1.2.1 Oil Well Service Industry | 2 |
| 1.2.2 Position-Based Visual Servoing | 3 |
| 1.2.3 Stereo Cameras and Calibration Process | 3 |
| 1.2.4 3D Position Sensing..... | 4 |
| 1.3 REVIEW OF PREVIOUS WORK AND PROPOSED SOLUTIONS | 5 |
| 1.3.1 Three-dimensional Position Sensing Systems | 5 |
| 1.3.1.1 Laser Range Scanning (LRS) | 5 |
| 1.3.1.2 Laser Scanners vs. Cameras | 6 |
| 1.3.2 Camera-Based Range Sensing | 7 |
| 1.3.2.1 Available Sensors | 7 |
| 1.3.2.2 Selection of Single Camera and Stereo Vision..... | 10 |
| 1.3.3 Visual Servoing and Image Guided Robots..... | 11 |
| 1.3.3.1 Definition of Visual Servoing..... | 11 |
| 1.3.3.2 Camera Configuration in Visual Servoing | 11 |
| 1.3.3.3 Servoing Architecture in Visual Servoing..... | 12 |
| 1.3.3.4 Comparing Direct Visual Servo and Dynamic Look-and-Move..... | 13 |
| 1.3.3.5 Comparing Position-based and Image-based Approaches | 13 |
| 1.3.4 Camera Calibration | 15 |
| 1.3.5 Selection of Stereo Camera Locations..... | 17 |
| 1.3.6 Selected Servoing Architecture..... | 19 |
| 1.3.6.1 Position-based vs. Image-based Approach..... | 19 |
| 1.3.6.2 Dynamic Look-and-move or Direct Visual Servo?..... | 20 |
| 1.4 CONTRIBUTIONS OF THESIS | 21 |
| CHAPTER 2. SYSTEM DESIGN | 23 |
| 2.1 OVERVIEW OF SYSTEM | 23 |
| 2.1.1 Introduction..... | 23 |
| 2.1.2 System Block Diagram | 23 |
| 2.2 STEREO CAMERAS AND TORSO-MOUNTED CAMERA LOCATION | 26 |
| 2.2.1 Custom-designed Stereo Cameras | 26 |
| 2.2.2 Torso-mounted Stereo Cameras..... | 28 |
| 2.3 POSITION-BASED LOOK-AND-MOVE VISUAL SERVOING | 29 |

| | | |
|--|---|-----------|
| 2.4 | COORDINATE SYSTEM ASSIGNMENT | 31 |
| 2.5 | HARDWARE INTERRUPT DRIVEN DATA COMMUNICATION..... | 34 |
| CHAPTER 3. STEREO CAMERA CALIBRATION..... | | 35 |
| 3.1 | THE FINITE CAMERA PROJECTION MODEL | 35 |
| 3.1.1 | Camera Extrinsic and Intrinsic Parameters..... | 35 |
| 3.2 | ESTIMATION OF CAMERA MODEL USING DLT METHOD | 36 |
| 3.2.1 | Introduction..... | 36 |
| 3.2.2 | Implementation of the DLT Method..... | 36 |
| 3.2.3 | Estimation of Camera Model Parameters | 38 |
| 3.3 | LEAST-SQUARES SOLUTION TO OBTAIN CAMERA MODEL PARAMETERS..... | 40 |
| 3.3.1 | Parameter Estimation by Minimizing LS Loss Function..... | 40 |
| 3.3.2 | Associating the LS Components with Camera Parameters..... | 41 |
| 3.3.3 | An Iterative LS Approach to Find the Camera Parameters | 43 |
| 3.4 | A REDUCED STEREO CAMERA MODEL CALIBRATION | 44 |
| 3.4.1 | Introduction..... | 44 |
| 3.4.2 | System Parameterization..... | 44 |
| 3.5 | DIFFERENT USES OF THE CAMERA MODELS | 46 |
| 3.6 | DATA COLLECTION TECHNIQUES FOR CAMERA CALIBRATION | 47 |
| 3.6.1 | Camera Calibration Using a Cartesian Robot..... | 47 |
| 3.6.1.1 | Camera Calibration Setup..... | 48 |
| 3.6.1.2 | The Specification of the Cartesian Camera Calibration Robot | 50 |
| 3.6.2 | In situ camera Calibration..... | 52 |
| 3.6.2.1 | System Parameters for In situ Camera Calibration | 54 |
| 3.6.2.2 | Specification of the In Situ Camera Calibration Setup..... | 55 |
| 3.7 | LENS DISTORTION CORRECTION..... | 57 |
| 3.7.1 | Introduction..... | 57 |
| 3.7.2 | Lens Distortion Extraction..... | 58 |
| 3.7.2.1 | Distortion Coefficients Calculation | 58 |
| 3.7.2.2 | Obtained Distortion Model Visualization..... | 59 |
| CHAPTER 4. 3D POSITION SENSING USING STEREO VISION..... | | 60 |
| 4.1 | INTRODUCTION..... | 60 |
| 4.2 | STEREO VISION AND EPIPOLAR GEOMETRY..... | 61 |
| 4.2.1 | Stereo Vision Principles..... | 61 |
| 4.2.2 | Facts Perturbing Parallel Stereo Vision Characteristics | 65 |
| 4.3 | 3D RECONSTRUCTION AND POSITION ESTIMATION | 66 |
| 4.3.1 | Two Methods to Estimate the Object Position in 3D | 66 |
| 4.3.1.1 | First method, matrix equations | 66 |
| 4.3.1.2 | Second method, direct transformation..... | 67 |
| 4.4 | LINE AND PLANE INTERSECTION IN 3D..... | 68 |
| CHAPTER 5. EXPERIMENTAL RESULTS..... | | 70 |
| 5.1 | CAMERA CALIBRATION EXPERIMENTAL RESULTS | 70 |
| 5.1.1 | Data Collection using the Cartesian Robot..... | 71 |
| 5.1.1.1 | Validate Camera Model using DLT and SDLT Methods..... | 73 |
| 5.1.2 | Data Collection using the In situ Calibration Technique..... | 75 |

| | | |
|-------------------------|---|------------|
| 5.1.2.1 | Validate Camera Model using DLT and SDLT Methods..... | 76 |
| 5.2 | EXPERIMENTAL RESULTS FOR 3D POSITION ESTIMATION | 78 |
| 5.2.1 | 3D Positioning without Distortion Correction..... | 80 |
| 5.2.1.1 | 3D Positioning for Area Larger than Calibration Area | 80 |
| 5.2.1.2 | 3D Positioning within Calibration Area | 81 |
| 5.2.2 | 3D positioning with Distortion Correction | 82 |
| 5.2.2.1 | 3D Positioning for Area Larger than Calibration Area | 82 |
| 5.2.2.2 | 3D Positioning within Calibration Area | 83 |
| 5.2.2.3 | 3D Positioning for Area Smaller than Calibration Area..... | 84 |
| 5.2.3 | Comparing the Accuracy Measurement Results Using DLT Method | 85 |
| 5.2.4 | Comparing the Accuracy Measurement Results using SDLT Method..... | 86 |
| 5.2.5 | Effects of Lens Distortion Correction..... | 87 |
| 5.3 | SENSITIVITY ANALYSIS OF THE 3D POSITION SENSING..... | 88 |
| 5.3.1 | Introduction..... | 88 |
| 5.3.2 | Horizontal Pixel Error..... | 89 |
| 5.3.3 | Vertical Pixel Error | 91 |
| 5.4 | JOINT ANGLE INDEPENDENT VISUAL SERVOING..... | 93 |
| CHAPTER 6. | CONCLUSIONS AND FUTURE WORK | 96 |
| 6.1 | SYSTEM EVALUATION..... | 96 |
| 6.1.1 | 3D Position Sensing Accuracy..... | 96 |
| 6.1.2 | Lens Distortion Correction Effects | 97 |
| 6.1.3 | Real-world System Test..... | 98 |
| 6.1.4 | Vision System Errors | 98 |
| 6.1.5 | Mechanical Inaccuracies..... | 98 |
| 6.1.6 | Comparing with Affine (Weak Perspective) Stereo Systems | 99 |
| 6.2 | FUTURE REFINEMENTS | 99 |
| 6.2.1 | An Automatic Accuracy Test for Obtained Camera Model | 99 |
| 6.2.2 | Accurate Camera Calibration Platform..... | 99 |
| REFERENCES..... | | 100 |
| APPENDIX A | | 107 |

List of Figures

| | |
|---|----|
| Figure 1-1. The TDH robot installed on top deck of a mobile rig in Red Deer, Alberta owned by Nabors Canada Ltd. | 2 |
| Figure 1-2. 3D scanner consisting of a 2D scanner and servo driver | 5 |
| Figure 1-3. A conventional structured-light scanning system | 8 |
| Figure 1-4. Biprism-based vision system..... | 9 |
| Figure 1-5. The possible camera locations on the TDH robot structure | 17 |
| Figure 2-1. The proposed visual Servoing block diagram..... | 24 |
| Figure 2-2. The stereo camera based visual servoing block diagram | 25 |
| Figure 2-3. The designed torso-mounted stereo camera rig (Courtesy of Motion Metrics Int'l) | 27 |
| Figure 2-4. The real-world stereo images of the proposed visual servoing system..... | 27 |
| Figure 2-5. The stereo cameras rig installed on the TDH robot torso | 28 |
| Figure 2-6. The Position-based (pipe cross section) servoing architecture | 29 |
| Figure 2-7. Position-based look-and-move visual servoing block diagram | 30 |
| Figure 2-8. DH frame assignment for TDH robot | 31 |
| Figure 2-9. Frame assignment for the TDH arm and the stereo cameras | 31 |
| Figure 2-10. The hardware interrupt driven threads | 34 |
| Figure 3-1. Schematic diagram of data collection using Cartesian robot | 48 |
| Figure 3-2. The Cartesian calibration robot..... | 51 |
| Figure 3-3. Schematic diagram of data collection for in situ camera calibration | 54 |
| Figure 3-4. The in situ camera calibration setup..... | 56 |
| Figure 3-5. The left and right image views of the in situ calibration tool | 56 |
| Figure 3-6. The radial distortion visualization..... | 59 |
| Figure 4-1. The epipolar geometry schematic | 61 |
| Figure 4-2. The stereo camera system with parallel image planes | 63 |
| Figure 4-3. Stereo vision with parallel and coplanar image planes (The image origins have been shown on the image corners to avoid confusion) | 64 |
| Figure 4-4. The cross section schematic of the pipe with gripper plane..... | 68 |
| Figure 5-1. The calibration cube generated by the Cartesian robot..... | 71 |
| Figure 5-2. The image trajectory of the LED during the calibration process using Cartesian robot. | 72 |
| Figure 5-3. The absolute error between est and act 3D Cartesian coordinates..... | 73 |
| Figure 5-4. The absolute error between est and act 3D Cartesian coordinates..... | 74 |
| Figure 5-5. The calibration volume generated by in situ calibration | 75 |
| Figure 5-6. The absolute error between est and act 3D Cartesian coordinates..... | 76 |
| Figure 5-7. The absolute error between est and act 3D Cartesian coordinates..... | 76 |
| Figure 5-8. The 3D positioning validation areas | 79 |
| Figure 5-9. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R3 without lens distortion correction has been selected..... | 80 |
| Figure 5-10. The histogram of the MAE between the Cartesian coordinates (x_0 , y_0 , z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions..... | 80 |

| | |
|---|----|
| Figure 5-11. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R2 without lens distortion correction has been selected..... | 81 |
| Figure 5-12. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions..... | 81 |
| Figure 5-13. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R3 with lens distortion correction has been selected. | 82 |
| Figure 5-14. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions..... | 82 |
| Figure 5-15. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R2 with lens distortion correction has been selected. | 83 |
| Figure 5-16. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions..... | 83 |
| Figure 5-17. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R1 with lens distortion correction has been selected. | 84 |
| Figure 5-18. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions..... | 84 |
| Figure 5-19. The image coordinates of the target object coincided with the radial distortion map | 87 |
| Figure 5-20. The region of interest in the left and right image planes..... | 89 |
| Figure 5-21. The error on the Cartesian coordinates of the target point occurred by the misplacement of the horizontal pixel in the right image plane. | 90 |
| Figure 5-22. The region of interest in the left and right image planes..... | 91 |
| Figure 5-23. The error on the Cartesian coordinates of the target point occurred by the misplacement of the vertical pixel in the right image plane. | 91 |
| Figure 5-24. The 3D position estimation of a stationary object during the angular motion of the robot base angle (θ_1 , third subplot from top)..... | 94 |

List of Tables

| | |
|---|----|
| Table 1-1. Camera configuration table of comparison | 18 |
| Table 2-1. DH table for the TDH robot | 32 |
| Table 3-1. Camera calibration and 3D estimation procedures for the DLT method | 46 |
| Table 3-2. Camera calibration and 3D estimation procedures for SDLT method | 46 |
| Table 5-1. Comparing the Mean Absolute Error (MAE) and Maximum Error (MAXE) between the actual and estimated Cartesian coordinates obtained using the classic and direct method | 74 |
| Table 5-2. Comparing the Mean Absolute Error (MAE) and Maximum Error (MAXE) between the actual and estimated Cartesian coordinates obtained using the classic and direct method | 77 |
| Table 5-3. The accuracy measurement comparison in three different AMA^\dagger using DLT. | 85 |
| Table 5-4. The worst case error between the actual and estimated 3D positions using DLT | 85 |
| Table 5-5. The accuracy measurement comparison in three different AMA using SDLT. | 86 |
| Table 5-6. The worst case error between the actual and estimated 3D positions using SDLT..... | 86 |
| Table 5-7. The errors between the actual and estimated stationary coordinates | 95 |
| Table 6-1. Maximum error in DLT and SDLT | 96 |
| Table 6-2. Average error in DLT and SDLT | 96 |
| Table 6-3. Effects of RLDC in SDLT method..... | 97 |
| Table 6-4. Effects of RLDC in DLT method..... | 97 |
| Table 6-5. Max error with RLDC (DLT)..... | 97 |
| Table 6-6. Max error with RLDC (SDLT)..... | 97 |

Acknowledgements

My sincerest gratitude is extended to my supervisors Dr. Peter Lawrence and Dr. Shahram Tafazoli for their dedicated supervision and support and for giving me this unique opportunity to explore my academic and professional interests. Their attention and insightful ideas have been a constant source of inspiration. It has been a privilege and great honor to work under such knowledgeable and understanding supervisors.

I am deeply grateful and indebted to my dearest parents, my brother and my sister for their patience and encouragement during challenging periods of my life. Their motivation and unconditional support have always encouraged me to pursue my interests and explore new horizons. I would like to express my warmest gratitude to my sister for her kind support in the early period of my residence in Canada.

I would like to express my foremost appreciation to Motion Metrics International Corp. and its hard-working engineers and scientists. This work would have not been possible without the leading-edge industrial contributions and financial support from Motion Metrics. I am also very thankful to Roboweld Inc. and its dedicated engineers and technicians for their help in manufacturing of the mechanical components of the system and for their support in running the laboratory experiments and the field trials in Nabors Canada facility in Alberta.

I would also like to express my heartfelt appreciation to all of my friends and colleagues who stood by my side and provided their valuable and generous assistance. Their consistent support and encouragement have been my greatest treasure and for which nothing can substitute.

Last, but definitely not least, I am very grateful to Precarn Inc. for their financial and management support of the project. The project reported here is a proof of the effectiveness of the Precarn model for successful innovation through collaboration.

Nima Ziraknejad

April 2007

Chapter 1. Overview and Background

1.1 Introduction

Outdoor machine vision applications have recently gained a lot of attention in the automation industry. Contriving visual servoing on outdoor manipulators provides remote sensing in different applications such as object placement, welding, component handling and so on. Visual servoing opens a vast area for vision guided robotic systems to provide closed-loop control using cameras as the main sensors of the system. Vision is a useful robotic sensor since it mimics the human sense of vision and allows non-contact measurement of the environment [1]. Heavy-duty vision sensors have been manufactured by high-tech industries to be employed in extreme conditions and environments. Examples of such conditions are extreme weather, extreme humidity, temperature between -40°C and $+50^{\circ}\text{C}$, wind velocity to 60 km/h, shock and vibration, dust and so on. Cameras as the most delicate parts of such visual aided robotic systems have evolved to meet these requirements.

A pipe handling robotic manipulator with applications in the oil industry is selected to accommodate the stereo camera rig as a torso-mounted visual servoing system. The robotic manipulator designed and fabricated by Roboweld Inc. will be referred to as TDH¹ in this thesis. A *position-based look-and-move* visual servoing method has been explained in detail to provide a comprehensive understanding of the performed work.

A heavy-duty stereo camera rig was designed and fabricated by Motion Metrics International Corp. to meet the required environmental effects. In this thesis work, a novel direct calibration process for the stereo cameras is introduced and explained in detail.

The developed visual servoing system has been successfully integrated with the TDH robotic manipulator over a period of one year. The system passed several testing procedures in its real-world environment (on top of the mast of the mobile rig). The shortcomings and problem areas were documented to be addressed in future work.

¹ TDH is the abbreviation for Top Deck Handler. This is referred to the robot handling the required tripping actions on top deck of an oil well.

1.2 Problem Statement and Machine Vision Solutions

1.2.1 Oil Well Service Industry

The oil well service industry provides the maintenance of the oil wells around the world. One of the most dangerous jobs in this area is assigned to a human operator to handle the pipes tripped out/in from/to the well. The typical height of the structure where the operator stands is 18 meters. The “trip out” action refers to the action of extracting the pipes from the well and the “trip in” action refers to the task of inserting the pipes back into the well. The servicing operation is performed on the pipes after they are extracted (tripped out) from the well and stacked in the fingers.

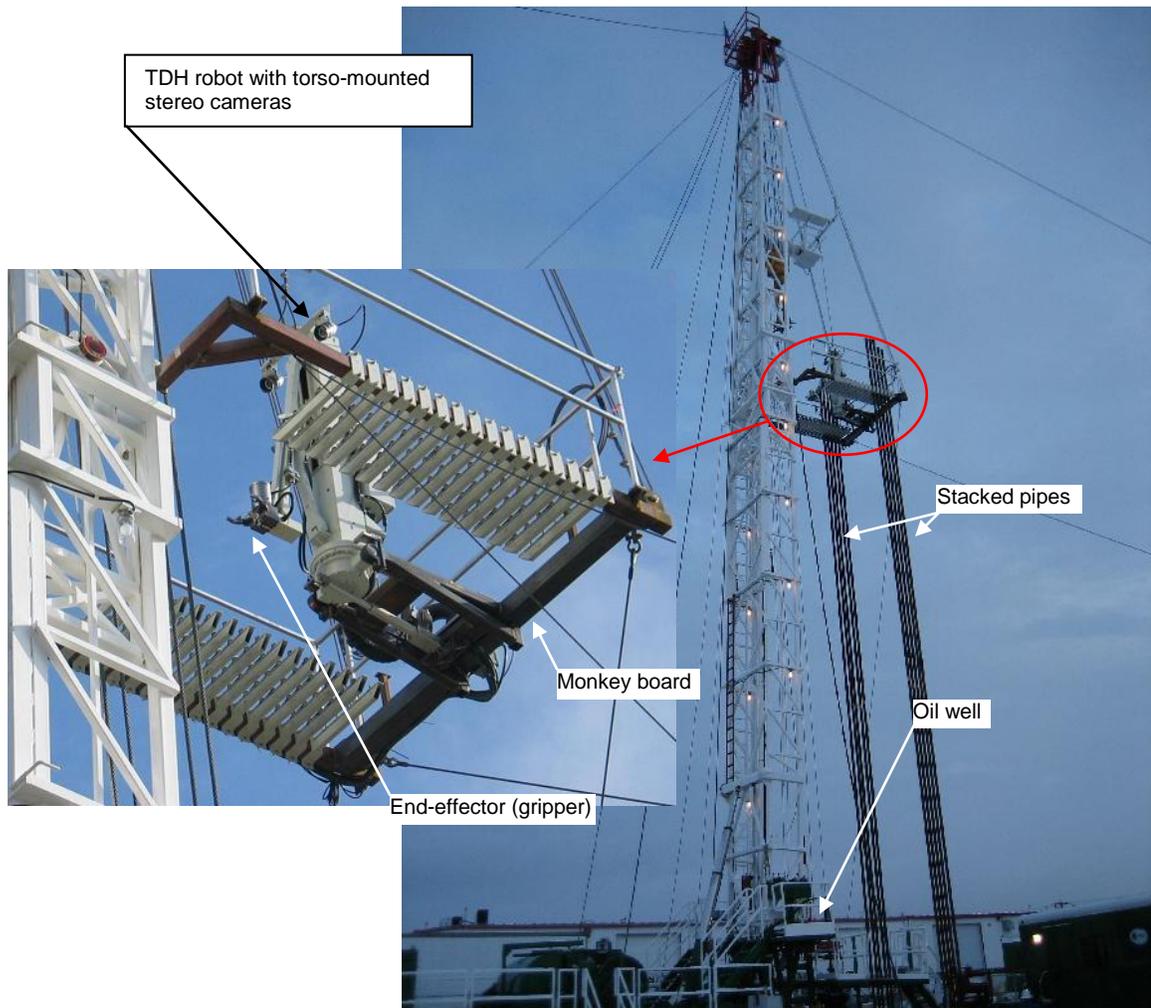


Figure 1-1. The TDH robot installed on top deck of a mobile rig in Red Deer, Alberta owned by Nabors Canada Ltd.

The employed vision guided robotic manipulator is intended to perform the tripping in and out actions. The goal is to replace the human operator with the vision guided robotic system. Figure 1-1 shows a typical mobile well service rig (a tall mast located on top of the well). The TDH robot is installed on the so-called monkey board structure.

1.2.2 Position-Based Visual Servoing

The robotic manipulator should be able to capture the pipe in real-time during trip out action. A torso-mounted *position-based look-and-move* visual servoing technique has been employed on the TDH robotic manipulator to accurately control the robot end-effector. The vision system provides the desired joint commands to the motion controller unit discussed in Section 2.1.2. The motion controller (not part of the vision system) then guides the robot end-effector to the desired position to capture the pipe and store in the predetermined location on the monkey board, during the trip out action.

1.2.3 Stereo Cameras and Calibration Process

Stereo vision is intended to be employed to perform the 3D position estimation. Two heavy-duty monochrome analog cameras have been placed on a bracket mounted on top of the TDH robot torso. The two cameras have been carefully aligned in parallel with each other and fastened tightly to the mounting bracket. Figure 2-5 shows the stereo camera rig installed on the TDH torso. The stereo camera system provides the required image coordinates to estimate the 3D position of the pipe cross section with the gripper motion plane. Different calibration methods have been performed and investigated in this thesis and the method satisfying the application requirements has been selected. The camera calibration process provides a transformation model to estimate the 3D position of the target object using its 2D footprint (image coordinates) in the acquired real-time images.

A calibration technique was employed that meets the industrial requirements of the application. The examples of such requirements are: easy to be performed, to be fast enough to impose the shortest possible downtime in the operation cycle.

The main environmental requirements for the cameras:

- 1- Extended temperature from $-40^{\circ}C$ to $85^{\circ}C$
- 2- Wider baseline than commercially available stereo cameras due to the close distance of the robot reachable space to the cameras. This provides noticeable disparity in the image pairs.
- 3- Heavy-duty to handle the high amount of shock and vibration
- 4- Waterproof
- 5- Easy to be installed tightly on the robot

1.2.4 3D Position Sensing

An accurate 3D position sensing algorithm is required to estimate the 3D position of the target object in the robot workspace. The process should be performed in real-time to meet the application requirements. The 2D coordinates of the target object should be detected in both images captured from the stereo cameras. In this application the pipe appear with lower intensity and different width compared to the other objects in the image. A 3D reconstruction algorithm is needed to provide the position of enough points on the pipe. The equation of the line associated with the pipe should be obtained. As the TDH robot workspace is planar, crossing the line with the gripper motion plane provides the desired grabbing position to successfully capture the pipe in trip out action.

1.3 Review of Previous Work and Proposed Solutions

1.3.1 Three-dimensional Position Sensing Systems

3D position sensing can be performed using different methods and sensory equipment. In this section the laser scanners and cameras as vision sensors are compared. It will then be shown why a camera as a vision sensor has been selected in the proposed visual servoing application.

1.3.1.1 Laser Range Scanning (LRS)

A review of Laser Range Scanning (LRS) and 3D laser scanner-based positioning systems is provided in this section. The review focuses on the shortcomings mentioned in the literature for laser-based positioning systems. It will be described how the mentioned drawbacks could affect the accuracy and requirements of the proposed visual servoing system.



Figure 1-2. 3D scanner consisting of a 2D scanner and servo driver

3D Laser range scanning method has been widely used in navigation and mobile robotics. Object detection and collision avoidance are the main goals in systems using the laser scanning methods. As there is no fast 3D scanner available at present, a commercial 2D Laser scanner that works on the time-of-flight measurement principle and an extra servo drive to reach the third dimension can be deployed [2]. Fast data collection with this type of sensor is still a challenging problem. In this case accurate synchronization of

the laser measurement and the scanning device is important. The majority of commercially available 3D LADAR (Laser Radar) devices are too expensive [3]. A 2D Laser scanner provides a vertical scan line and in order to obtain the 3rd dimension, mirrors or servo motors can be used. Implementation of a 3D Laser Radar system using 6 mirrors to form a hexagon-shaped reflector is presented in [4]. Wulf et al. has introduced a 3D Laser scanner consisting of a 2D laser range sensor (SICK) and a servo drive (Powercube) in [3].

Scanning rate and 3D reconstruction of the objects in the scene can be a challenge. By using a real-time operating system to generate accurate measurement timestamps and data correlation upon this timestamps, it is possible to get undistorted point clouds within a short scanning time of 4 seconds [2]. Nuchter et al. [5] introduces an object localization technique employing 3D laser range scans using a cascade of classifiers. The learning process employed in their research has to be executed once for training and the classification in the fastest mode requires 300ms. The whole process for the accurate object localization in 3D takes 1.4s.

1.3.1.2 Laser Scanners vs. Cameras

The main two drawbacks of position sensing systems based on laser scanners are the electro mechanical requirements and the processing speed. Servo motor-based actuators and mirrors are required to provide the 3rd dimension in the scanning process. The outlined visual servoing problem requires a fast processing speed to satisfy the real-time requirements of live tracking. Camera based 3D position sensing has been selected in this research based on the information below:

- Employing cameras in a visual servoing application can help provide a visual perception of the environment to the operator using the system. This can be extremely advantageous when the visual servoing or 3D position sensing is performed in inaccessible and remote locations.
- Employing a stereo camera system leads to obtaining a 3D mapping of the scene without a need for using additional components (actuators and mirrors). In scanners, accurate servoing of the laser beam requires a delicate and sophisticated

mechanical structure that may get damaged during extensive shock and vibration. In fact, using fixed-lens camera-based systems does not require any moving parts in the sensory equipment.

- Image acquisition systems are available for collecting the stereo images with a sampling rate of typical 30fps. The image update can be significantly increased by employing high speed digital image acquisition sensors (e.g. Firewire). Such systems lead frame rates of up to 90fps for images with a resolution of 640x480.
- High-tech camera manufacture companies provide heavy-duty rugged cameras that can perform well in extreme weather conditions and low temperature. These cameras have been widely used in the mining and underwater industries and their improvement is an ongoing process.
- Using the color characteristics and histogram information in color and monochrome image acquisition respectively, provide useful information to localize the objects in 2D images.

1.3.2 Camera-Based Range Sensing

1.3.2.1 Available Sensors

As mentioned in the previous section, a camera-based method has been selected over the 3D laser scanning method. In this section a review of single and stereo camera sensing systems is provided and then it will be explained why the stereo camera system has been selected over single camera case. Before comparing the single and stereo camera platforms, it is important to consider what we lose in the single camera case. Depth is the most important variable that cannot be extracted from the 2D image produced by a single camera alone. However, various augmented systems have been employed in various ways to recover the depth as the 3rd required dimension. Structured-light and biprism-based systems employ a single camera and provide the required inputs to a 3D reconstruction engine.

- **Structured-light (single camera)**

Structured-Lighting was introduced in 1970s as a means of recovering 3D shape of objects. It uses the same principle of triangulation that is used in stereo vision but avoids the difficulty in matching stereo points as explained in [6].

Structured-light scanners are widely used for various applications in robotics and computer vision. They are especially effective in 3D object bin picking and 3D object modeling applications because of the accuracy and reliability of the range data yielded [7]. Figure 1-3 shows a typical structured-light scanning system.

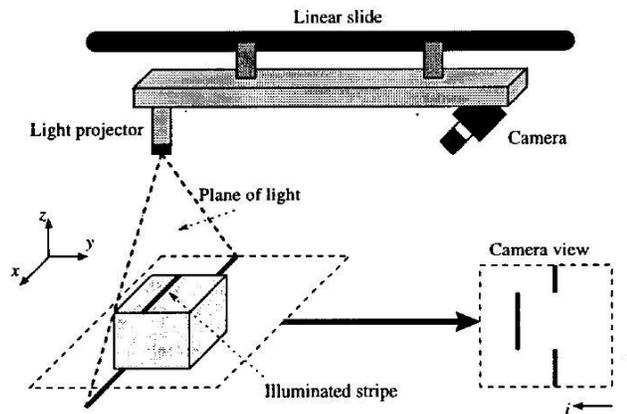


Figure 1-3. A conventional structured-light scanning system

Two important conditions that should be always considered in structured-light sensing systems:

- 1- There should be one or more light sources to illuminate the object of interest.
- 2- Occlusion which can be the result of the object shape and its orientation should not affect the light source illumination and the camera observation.

To address the second condition researchers have introduced structured-light techniques using dual beam and multiple cameras [7]-[9].

- **Biprism-based vision system (single camera)**

Lee et al. in [40] proposed a practical stereo camera that uses only one camera and a biprism in front of the camera. Biprism is an optical device for obtaining interference fringes. The equivalent of a stereo pair of images is formed

as the left and right halves of a single charge coupled device (CCD) image using a biprism.

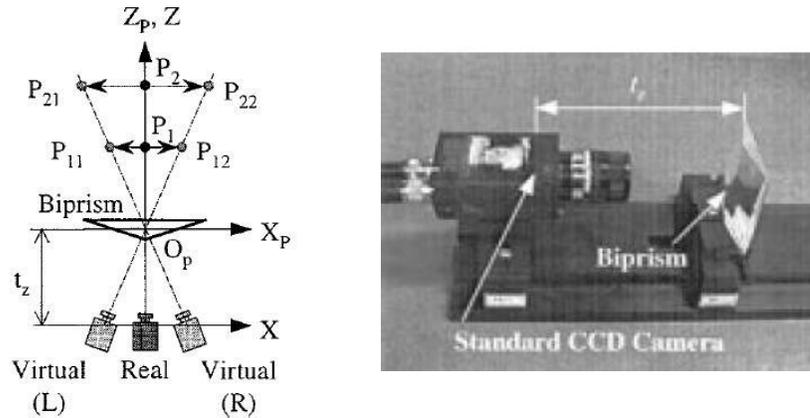


Figure 1-4. Biprism-based vision system

- **Stereo cameras and binocular vision (two cameras)**

Stereo vision systems provide a binocular vision of the target object in 3D scene. Stereo matching is generally used to extract the projection of the points on the target object in the left and right image plane. A 3D reconstruction algorithm uses the obtained stereo points to estimate the 3D position of the target object.

A stereo camera system consists of two cameras with a certain baseline (the distance between the two CCD planes of the cameras) and orientation related to a base frame. Stereo vision has been widely used in 3D sensing since it mimics the human visionary system. In stereo vision systems, 2D images of an object appear with a certain disparity in the left and right images. This disparity is a valuable source of information to reconstruct the 3D position of the object using triangulation method.

Stereo camera systems must deal with more input data but can also offer accurate 3D information [36]. Twenty one years ago, Kuno et al. [37] introduced a robotic system using stereo vision to perform object grasping. Silva et al. [38] designed a stereo vision system for real-time inspection and 3D reconstruction. Yamamoto et al. [39] introduced an arm-pointing gesture system using stereo cameras. Stereo vision has been widely used in visual servoing applications where the *eye-to-hand* [17]-[20] approach is considered for sensor configuration. The

main problem of employing stereo cameras for *eye-in-hand* applications is the size of the stereo camera rig that may interfere with the objects in the workspace of the robot.

1.3.2.2 Selection of Single Camera and Stereo Vision

Non-stereo-camera based vision systems require additional devices and components in order to obtain the depth information lost in the 2D image plane. Here are the most important reasons why a stereo camera has been preferred over the other platforms:

- 1- Components such as the light source, laser beam, biprism may not meet the required industrial standards for outdoor environment.
- 2- The baseline between a stereo camera pair can be adjusted to give accurate depth reading over the range required (unlike biprism).
- 3- A high laser intensity is required to overcome sunlight and this may be unsafe. This can be a disadvantage for systems employing structured-light or laser scanners.

1.3.3 Visual Servoing and Image Guided Robots

Efforts in the area of visual servoing are too numerous to exhaustively list here. Instead some notable efforts that led to the proposed visual servoing system are provided.

1.3.3.1 Definition of Visual Servoing

Hutchinson et al. in [1] defines visual servoing as when machine vision provides closed-loop position control for a robot end-effector. Shirai and Inoue in [10] refer visual servoing to when visual feedback loop can be used to correct the position of a robot to increase the task accuracy. The term *visual servoing* was introduced in 1977 by Hill and Park and prior to this introduction, the term *visual feedback* was generally used.

Visual servoing can be employed using different techniques. There are two preliminary concepts in visual servoing that categorize the techniques to be used in the visual servo system design.

- 1- Camera configuration refers to where to install the cameras
- 2- Servoing architecture refers to how to control the robot using vision

1.3.3.2 Camera Configuration in Visual Servoing

Camera configuration (location of the camera on the robotic work cell) in visual servo systems can be categorized as follows:

- 1- Mounted on the end-effector. This is called *eye-in-hand*.
- 2- Fixed in the workspace. This is referred to *eye-to-hand*.

Employing the cameras in the *eye-in-hand* configuration leads to installing the cameras on the robot end-effector. In this scenario, there is known relationship between the pose of the camera and the pose of the end-effector. The image of the object is dependent on the pose of the end-effector. The *eye-in-hand* configuration is the most popular method of visual servoing and has been extensively described in [11]-[16].

In the *eye-to-hand* approach the image of the target is independent of the robot motion. In this case the cameras are related to the base coordinate system. Therefore knowledge of the robot joint angles does not directly lead to the camera location in the *eye-to-hand* approach. In systems in which the camera location is unknown, a calibration process should be employed to obtain the relationship between the cameras and the base

coordinate system. The *eye-to-hand* method is described in [16]. Muis et al. [18] and Flandin et al. [19] have introduced a method that incorporates both the *eye-in-hand* and the *eye-to-hand* techniques to perform the desired visual servoing task.

1.3.3.3 Servoing Architecture in Visual Servoing

The discussion in this section is based on the tutorial by Hutchinson et al. [1]. Visual feedback can be provided either using the image-based perception of the environment or the 3D position of the desired object calculated via a reconstruction algorithm. As such four categories have been identified as follows:

1- **Dynamic position-based look-and-move** ([22], [30], [34]):

When the control architecture is hierarchical and uses the vision system to calculate the set of inputs to the joint level controller, making use of inner joint feedback loops to stabilize the robot, it is referred to as a dynamic *look-and-move* structure. Using the camera model, the position of the target object is estimated. The setpoint inputs are the desired joint angles calculated based on the estimated target position.

2- **Dynamic image-based look-and-move** ([28], [29], [35]):

In image-based *look-and-move* architecture the control values are computed on the basis of image features directly. A feature space control law is required to calculate the desired joint angles.

3- **Position-based visual servo (PBVS)** ([21], [23]) **and image-based visual servo (IBVS)** ([23][24]-[26], [31]-[32]):

The visual servo structure eliminates the robot controller by replacing it with a visual controller. The visual controller is used to compute the joint inputs directly, thus using vision alone to control the mechanism. Both *position-based* and *image-based* methods can be employed in this architecture. Since the term *visual servoing* has come to be accepted as a generic description for any type of visual control of a robotic system, the term *direct visual servo* is used for PBVS and IBVS.

1.3.3.4 Comparing Direct Visual Servo and Dynamic Look-and-Move

Dynamic look-and-move requires a very accurate calibration process [27] and could be preferred to direct visual servo because of the following reasons:

- 1- The relatively low sampling rates of the vision systems make the direct control of a robot end-effector with complex, non-linear dynamics an extremely challenging problem.
- 2- Many robots already have an interface for accepting Cartesian velocity and incremental position commands.
- 3- The *look-and-move* strategy separates the kinematics singularities of the mechanism from the vision control. The robot is then considered as an idealized Cartesian motion device.

Because of the above reasons and the available advanced technology, people prefer to employ the dynamic *position-based* or *image-based look-and-move* for their visual servoing tasks. However there are applications in which *position-based* and *image-based* (PBVS and IBVS) visual servo have been employed. For example Weiss et al. [32] introduced an IBVS system that integrates visual servoing with robot movements. They have emphasized the design of such a “visual servo control system” in order to achieve stable and predictable system dynamics. Mahony et al. [31] designed an IBVS control of a serial robotic system using linear image features.

1.3.3.5 Comparing Position-based and Image-based Approaches

The main drawback of the *image-based* approaches is that the desired feature (the image of the robot end-effector) should be taught to the vision system. This is called *teach by showing* [32]. The advantage of the *image-based* approach over the *position-based* approach is that it does not require precise calibration and modeling as the feedback loop is closed in the image space [20]. Image based methods are mostly used in *eye-to-hand* applications [24]. In *image-based* servoing when the control error is calculated directly from image plane measurements, the main drawback would be the unpredictability of trajectories in Cartesian space particularly for the large initial pose error [22]. However, *position-based* approach is sensitive to camera calibration and the chosen pose estimation algorithm [22]. Another issue in *image-based* servoing is that

near singular configurations of the robot, very large changes in the joint space are required for very small Cartesian space changes.

The selected architecture in this work is dynamic *position-based look-and-move* and Section 1.3.6 explains the reasons why this architecture was preferred.

1.3.4 Camera Calibration

Camera calibration has been a long standing research issue in computer and robot vision. This survey is mainly focused on stereo camera calibration since this thesis employs stereo cameras.

According to Tsai [41], camera calibration in the context of three dimensional machine vision is the process of determining the internal camera geometric and optical characteristics (intrinsic parameters) and/or the position and orientation of the camera frame relative to a certain world coordinate system (extrinsic parameters). Camera calibration can be roughly classified into two categories: Photogrammetric calibration and self-calibration [42]. The self-calibration technique is usually employed for moving cameras e.g., when the camera is moved by an *eye-in-hand* manipulator. There are reports of self-calibration technique in [43]-[46].

The closest research to the calibration process employed in this thesis is provided in [47] and [70]. Sid-Ahmed et al. in [47] performed a calibration process for the stereo cameras individually using the DLT (Direct Linear Transformation) method. Hence, a transformation must be obtained for each camera in the stereo rig. The introduced stereo camera calibration method provides a reduced transformation to relate the 3D calibration points to their corresponding image coordinates in both left and right image planes. The reduced transformation is obtained by adding the equations with equal vertical image displacement and representing them as one equation. This assumes that the image coordinates in the stereo images (with parallel and coplanar images) share the same vertical displacement.

Studying the literature has revealed that raw data collection (collecting the 3D calibration and image points) is an essential step in all calibration procedures. Most of the proposed camera calibration techniques use a simple cardboard sheet containing equal black and white squares with known dimensions and pose relative to an arbitrary base frame. However there are techniques using other calibration tools to perform the calibration. Marita et al. in [48] calibrated the cameras using square shape objects with printed cross signs on them. Zhang [49] has used a bar with three equally distanced beads to calibrate the camera. Mallem et al. [50] introduced a calibration system with a data

collection technique using a LED mounted on the end-effector of a robot. The 3D position of the LED is known using the robot encoders. This technique is the closest research to the calibration method proposed for this thesis.

1.3.5 Selection of Stereo Camera Locations

The *eye-to-hand* and *eye-in-hand* camera configurations were discussed in Section 1.3.3.2. These configurations dictate whether the camera(s) is mounted on a location fixed in the workspace or a moving location on the robot, respectively. The figure below shows the possible camera locations on the employed TDH robot structure:

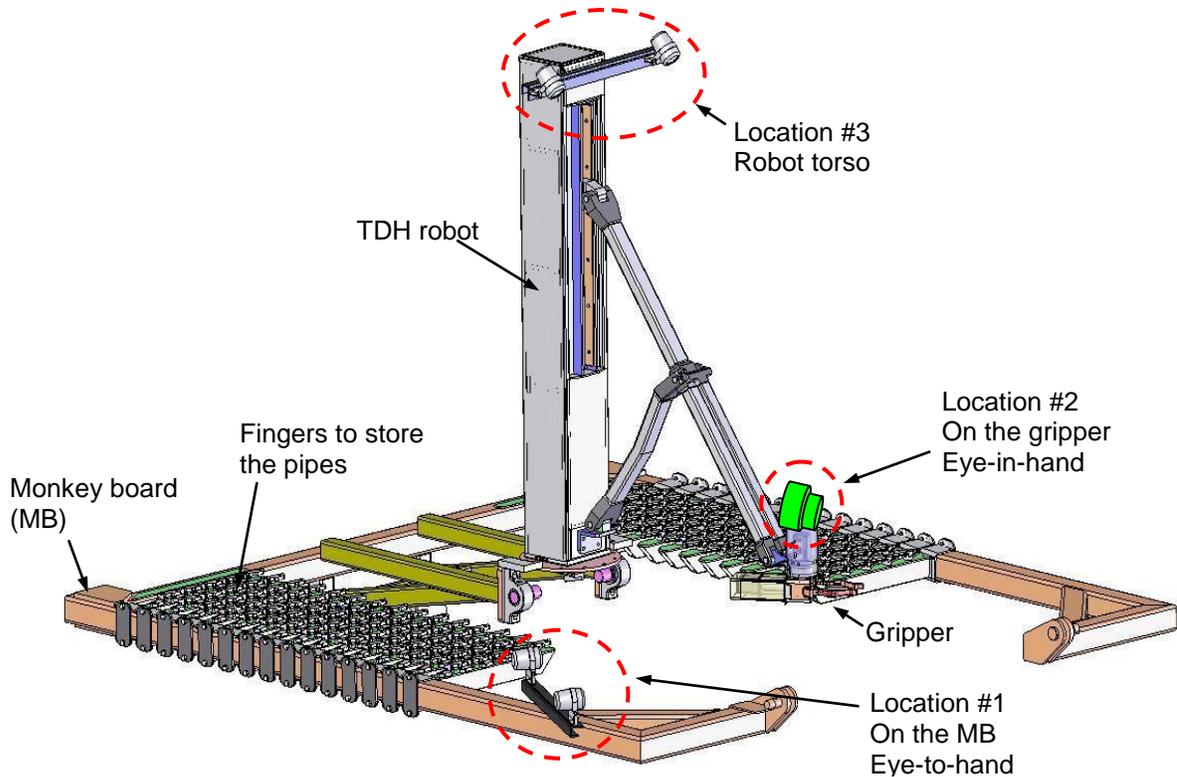


Figure 1-5. The possible camera locations on the TDH robot structure

Location 1 and 2 in Figure 1-5 show the concept of the *eye-to-hand* and *eye-in-hand* configurations respectively. The stereo cameras employing the *eye-to-hand* configuration can be installed on the monkey board structure. The orientation of the stereo camera rig is fixed to the robot base frame in this configuration. The stereo cameras employing the *eye-in-hand* configuration can be installed on the robot end-effector (gripper). Since most of the *eye-in-hand* visual servoing systems employ one single camera on the robot gripper, the stereo camera rig has not been considered in this configuration. A third possible stereo camera rig position is on the top of the robot torso as depicted in Figure 1-5 and it is referred to as *torso-mounted* in this thesis.

The angular displacement of the TDH robot base (robot torso) is the single DOF for the stereo camera rig. The TDH robot with such torso-mounted camera configuration acts similar to human body as the eyes are located on the human head. The pros and cons of the mentioned camera configurations have been listed in Table 1-1. The table is prepared solely for the proposed visual servoing system.

Table 1-1. Camera configuration table of comparison

| Camera configuration | Pros | Cons |
|-----------------------------|--|---|
| Eye-in-hand | <ol style="list-style-type: none"> 1- <i>Image-based</i> solution is easier to implement [11]-[16]. | <ol style="list-style-type: none"> 1- The camera rig interferes with the fingers on the Monkey board. 2- The camera may be hit by other pipes during the trip actions. 3- Camera may be facing in the wrong direction. |
| Eye-to-hand | <ol style="list-style-type: none"> 1- The image does not change with the robot motion [17], [18]. 2- No need for robot motion compensation. | <ol style="list-style-type: none"> 1- The stationary scene does not satisfy the visual monitoring requirements for the operators. 2- Hard to install on the monkey board structure due to the space constraints. 3- Great demands on calibration accuracy 4- Asymmetric 3D position sensing within the workspace. |
| Torso-mounted | <ol style="list-style-type: none"> 1- The tripping actions, pipe manipulation and the stored pipes can be monitored by operators. 2- The most convenient installation position w.r.t mechanical constraints. 3- No rain or sun effects as the cameras are tilted down. 4- Symmetric 3D position sensing for the pipes. | <ol style="list-style-type: none"> 1- The cameras move with the robot base angle. 2- The robot base angle should be compensated during the servoing tasks. |

The *torso-mounted* camera configuration was selected for three main reasons:

- 1- Provide a dynamic view of the robot workspace during tripping, pipe handling, and stacking phases.
- 2- In this configuration the cameras are installed well above the robot workspace and can be tilted down. This significantly reduces the effects of sun beam and rain in the images and on the front glasses of the cameras respectively.
- 3- The stereo rig in this location is protected against the possible impacts caused by other objects in the workspace.
- 4- Due to the symmetry of the workspace as viewed from the camera pair perspective, a symmetric 3D position sensing accuracy is achieved.

1.3.6 Selected Servoing Architecture

Now that the “torso-mounted” camera configuration has been selected, the servoing architecture should be selected out of the four major architectures discussed previously in Section 1.3.3.3. Dynamic *position-based look-and-move* visual servoing has been chosen to be used in the proposed visual servoing system. In the following, a quick comparison is outlined between *position-based* vs. *image-based* methods, as well as dynamic *look-and-move* vs. direct visual servoing.

1.3.6.1 Position-based vs. Image-based Approach

As discussed in Section 1.2, the target object to be grabbed by the robot end-effector in the trip out action is a pipe located in the reachable workspace of the robot. The *position-based* approach was selected for two main reasons:

- 1- The TDH robot is a planar manipulator. The objects placed in its workspace can be grabbed if their cross section positions with the robot gripper motion plane are determined. Considering the pipe as a line in 3D space, intersecting this line with the robot planar workspace provides the desired cross section. The equation of the line associated with the pipe should be obtained for this purpose. The obtained cross section is in fact the desired position for the robot end-effector (gripper). The complete algorithm has been provided in Section 4.4. Unlike an *image-based*

system which requires the features of the target object and the end-effector to be taught to it, a well-calibrated *position-based* system can perform this process by less computation and memory allocation.

- 2- The robot and the desired visual servoing tasks take place in a remote structure with no human operator access. The real-time simulation of the robot as well as the instantaneous 3D position of the pipe can be graphically depicted on the Graphical User Interface (GUI) in front of the operator monitoring the tasks. The 3D position of the pipe cross section with the robot planar workspace can be derived from a reliable *position-based* system. In an *image-based* system the robot end-effector is guided to the target object using the observed features and the system may not necessary produce any 3D perception of the environment.

1.3.6.2 Dynamic Look-and-move or Direct Visual Servo?

The look-and-move architecture is selected for the proposed visual servoing system. The TDH robotic manipulator is equipped with a powerful PLC-based closed-loop position controller and as such, it only makes sense to take advantage of this rather than build a separate closed-loop controller. The available motion controller unit requires the joint-space commands as setpoint inputs. The control scheme inside the motion controller unit performs the coordinated motion control of the robot and guides the robot to the desired joint angles and consequently the grabbing position. The look-and-move approach is essentially the most appropriate method here to provide the desired joint angles by solving the inverse kinematics of the robotic arm for the destination position (estimated by the 3D positioning system).

A definite advantage of the dynamic *look-and-move* approach is its generality regardless of the robot configuration. The stereo camera pair is used as a remote 3D sensing device for the pipe that appears within the scene and its dynamic maneuver generates the commands to be followed by the robotic manipulator. Due to the fairly simple structure of the TDH robot, direct visual servoing may as well produce satisfactory results. However, such approach is not considered in this thesis.

1.4 Contributions of Thesis

The contributions of the proposed visual servoing system fall into three categories. The contributions have been specifically made for the purpose of the unique visual aided robotic system and its applications in the discussed industry.

- **Outdoor Torso-mounted Visual Servoing using Stereo Vision**

One of the main features of the proposed visual servoing system is its outdoor functionality. In this thesis the outdoor functionality is referred to as a stereo camera based visual servoing robotic system tolerating the extreme weather conditions discussed in Section 1.1. The other main feature of the proposed visual servoing system is the fact that the cameras are installed on the robot torso and turn with the robot base angle. The contribution in this section is the selection of torso mounting cameras to make it compatible with human operator as well as robot control.

Mobile robots are the main robotic systems employing camera(s) for object detection and trajectory planning out of doors. The mobile robot in [72] can be an example. To the best of the author's knowledge there is no report of such outdoor torso-mounted visual servoing robotic system to be used in the oil well service industry or similar industries. However there is at least one report of an indoor robotic system with torso-mounted cameras (see [22]).

- ***In situ* Stereo Camera Autonomous Calibration**

In a majority of applications, sensor calibration is a prerequisite to proper use of the sensor for both measurement and control. Such a calibration method should be effective and not to be repeated frequently during the system long term performance. The *in situ* or in place stereo camera calibration has four main advantages not only satisfying the application requirement but to be considered in similar industrial visual servoing applications:

- i. The system does not require an extra manipulator to move the calibration tool within the workspace of the robot. The *in situ* calibration tool can be simply installed on the robot end-effector. Please refer to 3.6.2 for further detail.
- ii. The possibility of camera position misalignment due to shock and vibration is high during the pipe handling process. Additionally, there may be a need to replace a faulty camera with a new one. The stereo cameras can be recalibrated on site in case of such problems. However, one needs to first manually install the *in situ* calibration tool on the robot gripper. The robot manufacture designed the gripper structure for fast and easy accommodation of the *in situ* calibration tool.
- iii. The *in situ* stereo camera calibration can be started by pressing a button on the user interface of the system. The TDH robot moves the *in situ* calibration tool within the robot workspace automatically and after a data collection period of approximately 4min, the required camera parameters are estimated.
- iv. Most importantly, the *in situ* calibration takes into account any inaccuracies in camera installation and the TDH robot joint angle readings. It is well-known that the calibration of a sensor should be performed at similar conditions to its use.

Chapter 2. System Design

2.1 Overview of System

2.1.1 Introduction

The proposed visual servoing system employs a novel platform containing industrial PC104 and PLC architectures. The description of the hardware components has been provided in Appendix A. The system block diagram containing the main hardware components is provided in Section 2.1.2. The developed visual servoing system employs the torso-mounted camera configuration as the stereo cameras are mounted on the TDH robot structure. Sections 2.2.1 and 2.2.2 discuss the designed stereo cameras and the selected camera location on the real TDH robot structure. A *position-based look-and-move* architecture was selected to be the visual servoing architecture. Section 2.3 provides the details of the architecture implementation. The kinematics of the TDH robotic arm and its relation to the camera frames requires a well defined coordinate system assignment. Section 2.4 provides the assignment of the robot and camera coordinate systems. This section also discusses the employed notation for different points and coordinate systems.

2.1.2 System Block Diagram

The block diagram depicted in Figure 2-1 shows the components employed in the proposed visual serving system. The system contains the following main components:

- 1- TDH manipulator
- 2- Stereo cameras
- 3- PC104 embedded system
- 4- PLC and motion controller units
- 5- Servo motor drives
- 6- LCD touch screen monitor

The main focus of this thesis has been placed on the camera configuration, stereo camera calibration and the 3D position sensing algorithms. These algorithms together with the robot joint-level control (not discussed here) construct the desired *position-based look-and-move* visual servoing system.

As shown in the system block diagram below, the data channel between the PC104 and the PLC/motion controller unit is in charge of the data exchange between these two computing devices. A *position-based look-and-move* servoing process running on the PC104 embedded vision system is directly communicating with PLC/motion controller via this channel. Section 2.5 discusses this communication.

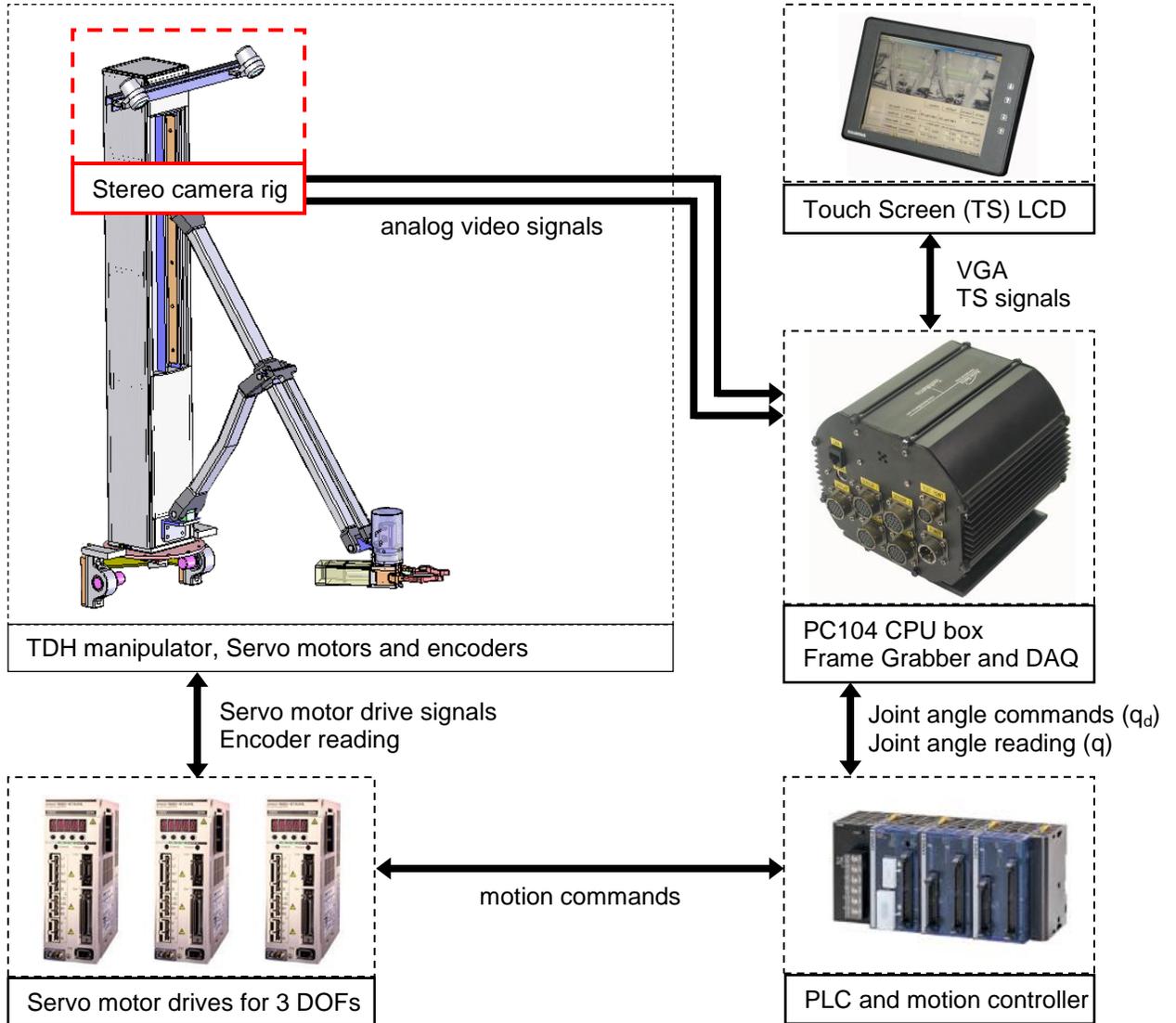


Figure 2-1. The proposed visual Servoing block diagram

(Pictures are included with permission from Motion Metrics Int'l Corp and Roboweld Inc.)

The stereo camera is the only visual sensor employed in the proposed visual servoing system. The sub-block associated to this sensor in Figure 2-1 falls into four categories indicating the organization of this thesis.

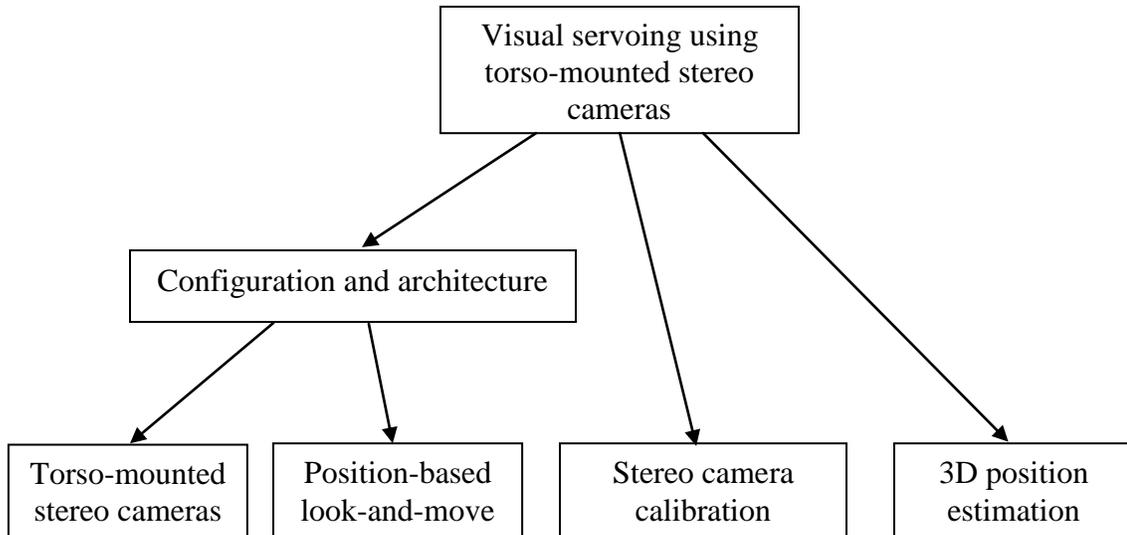


Figure 2-2. The stereo camera based visual servoing block diagram

The implementation of the torso-mounted camera configuration and the *position-based look-and-move* architecture is explained in this chapter. The stereo camera calibration methods and the proposed novel stereo camera calibration are discussed in Chapter 3. The 3D position sensing algorithm and the proposed method are discussed in Chapter 4.

2.2 Stereo Cameras and Torso-mounted Camera Location

2.2.1 Custom-designed Stereo Cameras

For the purpose of providing a heavy-duty and rugged stereo camera platform, the stereo camera mount shown in Figure 2-3 has been designed and fabricated to accommodate the cameras.

There are two DOFs for each of the cameras in the fabricated stereo camera rig. The DOF on the bottom of the camera stand provides the pan angle and the DOF at the joint connecting the camera to its stand provides the tilt angle. The adjustable pan and tilt angles were determined by mounting the stereo rig on the robot torso to obtain the required viewing angles. Once the angles were determined, specialized mechanical fasteners were used to tightly fasten the cameras to the steel mounting bracket. Two cubic steel pieces have been welded on the mounting bracket around the camera stand. These small pieces will help the camera fasteners to tolerate more vibrations during the robot operation. Figure 2-3 shows the employed fasteners.

The relatively long baseline (21in \approx 53cm) between the camera centers has been selected to provide a noticeable disparity between the image coordinates of the object of interest located in the robot workspace, about 1 to 3m from the camera planes.

The pan and tilt angles as well as the camera positions on the mounting bracket have been adjusted to construct a stereo vision sensor such that:

1. The CCDs are coplanar
2. The right CCD is horizontally shifted parallel to the left CCD
3. Stereo matching points in the left and right images have the same vertical displacement from the images origins. Consequently, the disparity between the stereo image points is in their horizontal displacement only. The specification of such geometry has been explained in detail in Section 4.2.

Figure 2-4 shows the stereo images captured from the left and right cameras on the stereo rig. The stereo coordinates of a point on the pipe as well as some other random



Figure 2-3. The designed torso-mounted stereo camera rig (Courtesy of Motion Metrics Int'l)

stereo points have been highlighted to show the observed disparity in the images. The FOV of the stereo cameras has been adjusted to satisfy the visual requirements of the servoing system. Such requirements are the appearance of the end-effector in both images and to the ability to see the fingers accommodating the tripped out pipes. The employed coordinate system and the notation will be discussed in Section 2.4 in detail.

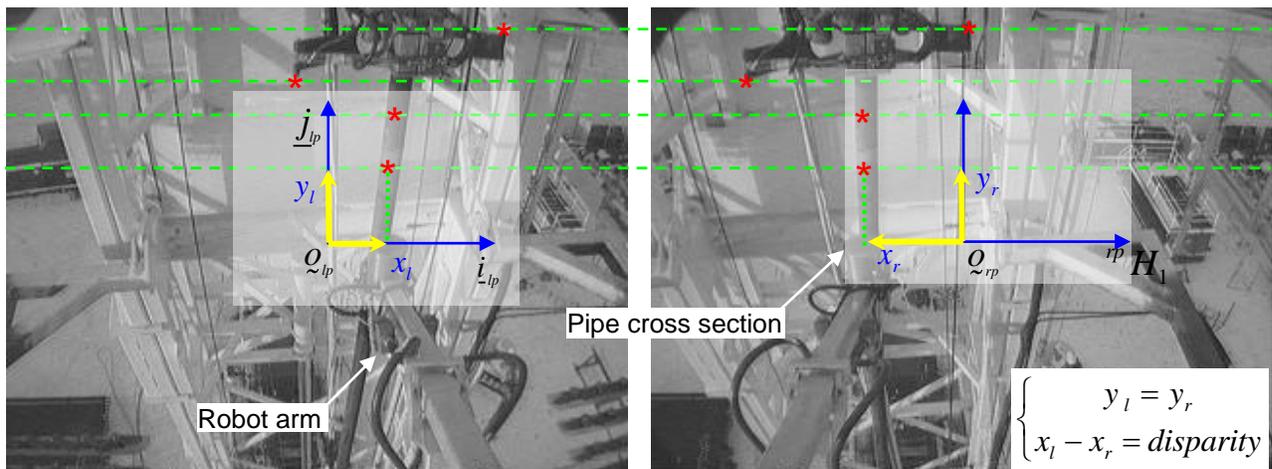


Figure 2-4. The real-world stereo images of the proposed visual servoing system Installed on a mobile service rig in Red Deer, Alberta, owned and operated by Nabors Canada Ltd.

2.2.2 Torso-mounted Stereo Cameras

As discussed in Section 1.3.5, the torso-mounted camera location was selected for the proposed visual servoing system. Figure 2-5 shows the stereo cameras mounted on the robot torso at 190cm height above the monkey board structure surface. The TDH robot is a 3 DOF manipulator with two revolute joints and one linear actuator (to provide the radial motion). As shown in Figure 2-5, the motion of the stereo camera rig installed on the robot torso is independent of the last two DOFs (d_2 and θ_3). The robot base angle θ_1 is the only DOF providing an angular motion for the stereo camera rig. The estimated 3D position of the object of interest is local to the robot torso. Section 4.1 discusses how the robot base angle θ_1 must be compensated during the servoing task. It discusses the required homogenous transformation to transfer the estimated 3D position to the base (world) coordinate system fixed to the monkey board $\{o_0, \underline{C}_0\}$.

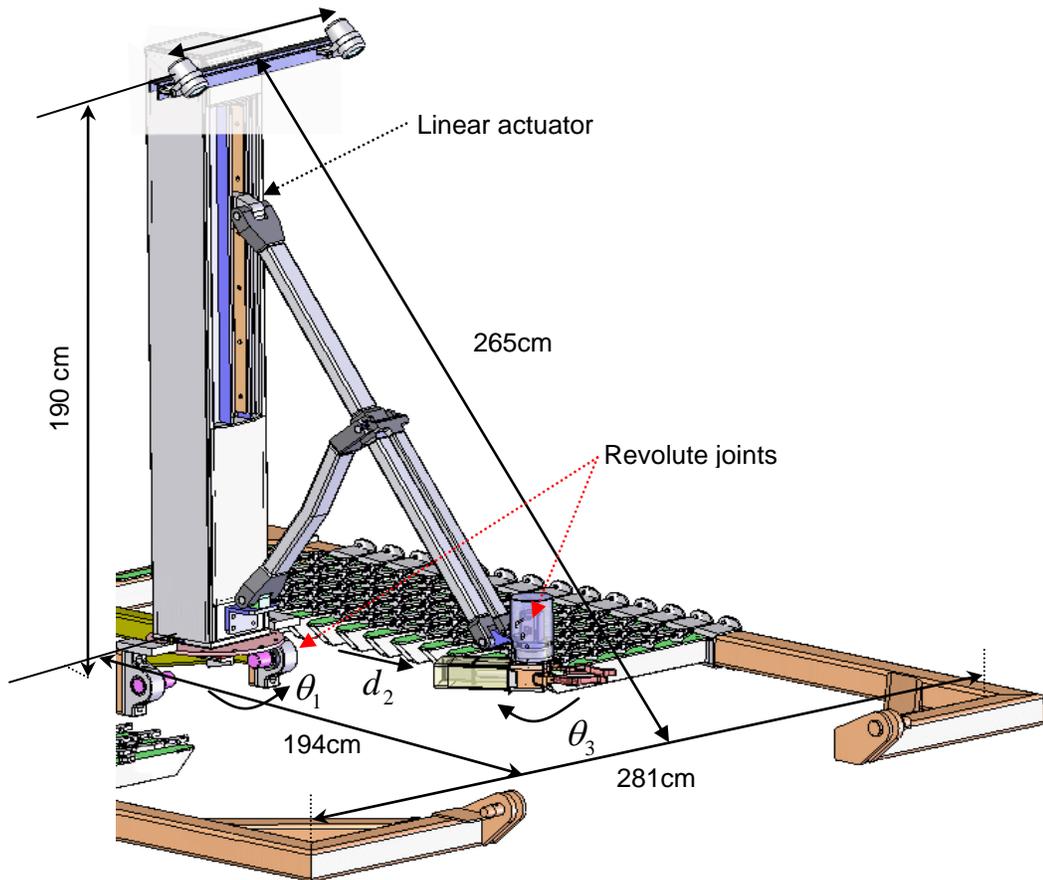


Figure 2-5. The stereo cameras rig installed on the TDH robot torso

2.3 Position-Based Look-and-Move Visual Servoing

The *position-based look-and-move* servoing architecture was selected in Section 1.3.6. The concept of the selected architecture can be explained more elaborately using Figure 2-6. An appropriate usage of the developed 3D position sensing system discussed in Chapter 4 is to estimate the cross section of the pipe and the planar workspace of the robot for trip out action. The TDH robot end-effector (gripper) moves on a plane horizontal to the monkey board structure. This plane contains the robot base origin and is coplanar with the plane formed by the $\underline{\dot{l}}_0$ and $\underline{\dot{j}}_0$ axes of the base frame (refer to Section 2.4 for more details). The estimated cross section of the pipe is in fact the desired position for the robot end-effector. This explains the *position-based* functionality of the servoing architecture.

Using the estimated desired end-effector position and the inverse kinematics of the TDH robotic manipulator allows one to calculate the robot desired joint angles. As the inverse kinematics of the TDH robot is a straightforward process, it has not been discussed here. The calculated joint angle commands are sent to the motion controller unit to be used in the closed-loop joint-level control of the robot.

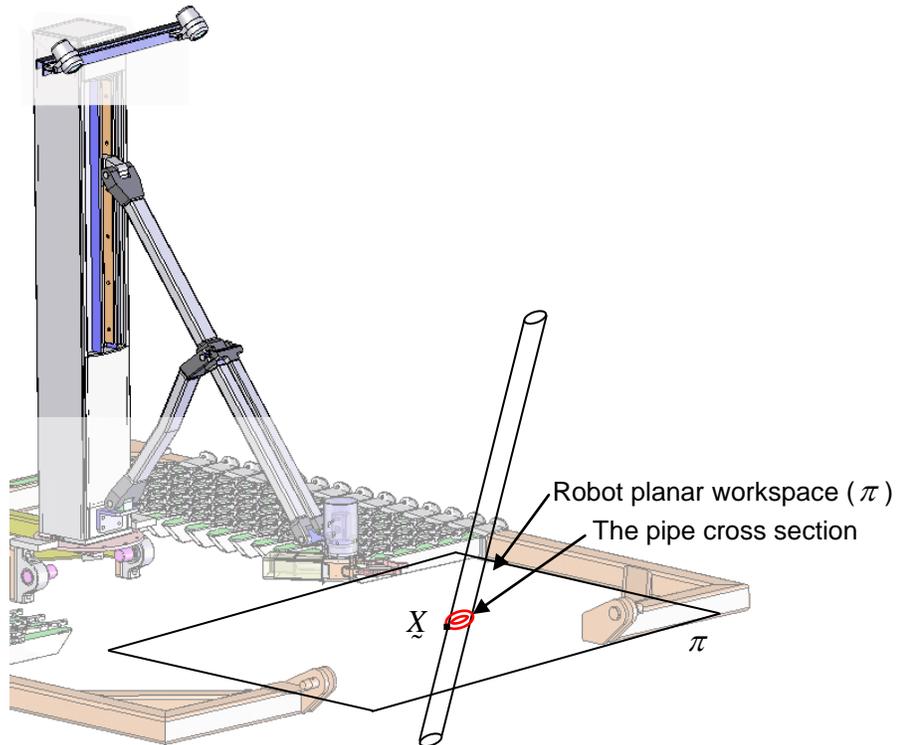


Figure 2-6. The Position-based (pipe cross section) servoing architecture

The block diagram depicted in Figure 2-7 shows the concept of *position-based look-and-move* architecture. The calculated joint-space commands (q_d) are sent to the motion controller unit.

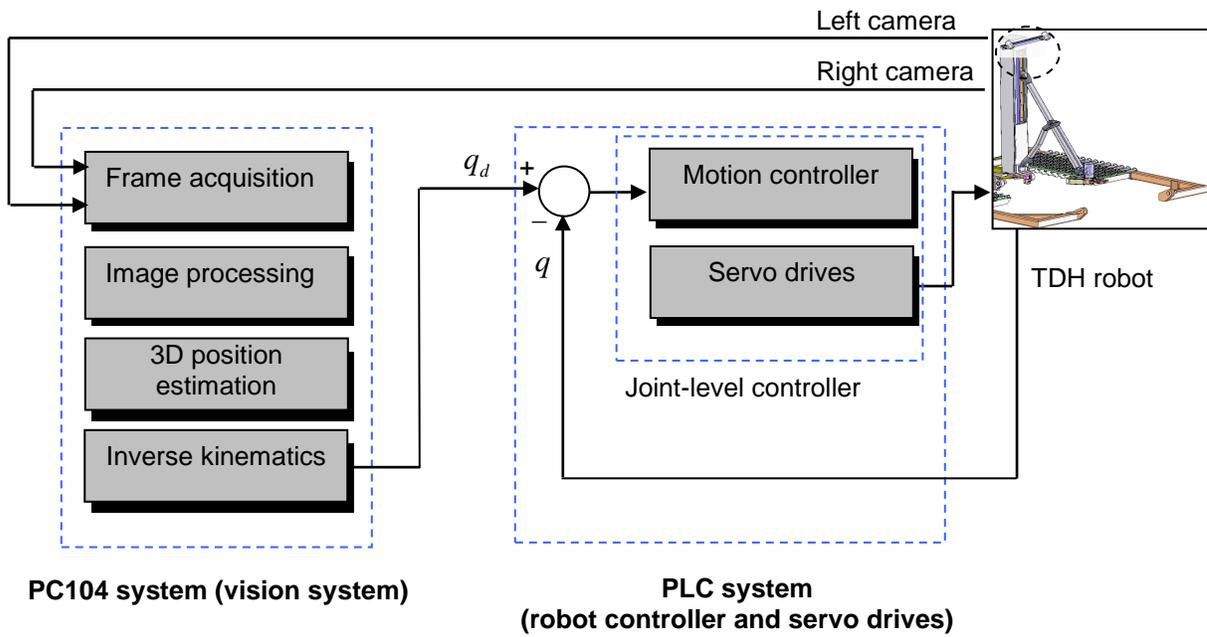


Figure 2-7. Position-based look-and-move visual servoing block diagram

2.4 Coordinate System Assignment

The Denavit-Hartenberg (DH) frame assignment of the TDH robotic manipulator has been shown in Figure 2-8.

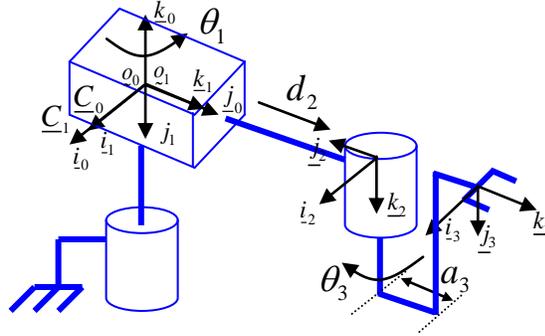


Figure 2-8. DH frame assignment for TDH robot

The frame assignment for the TDH robotic arm and the stereo cameras has been shown in Figure 2-9.

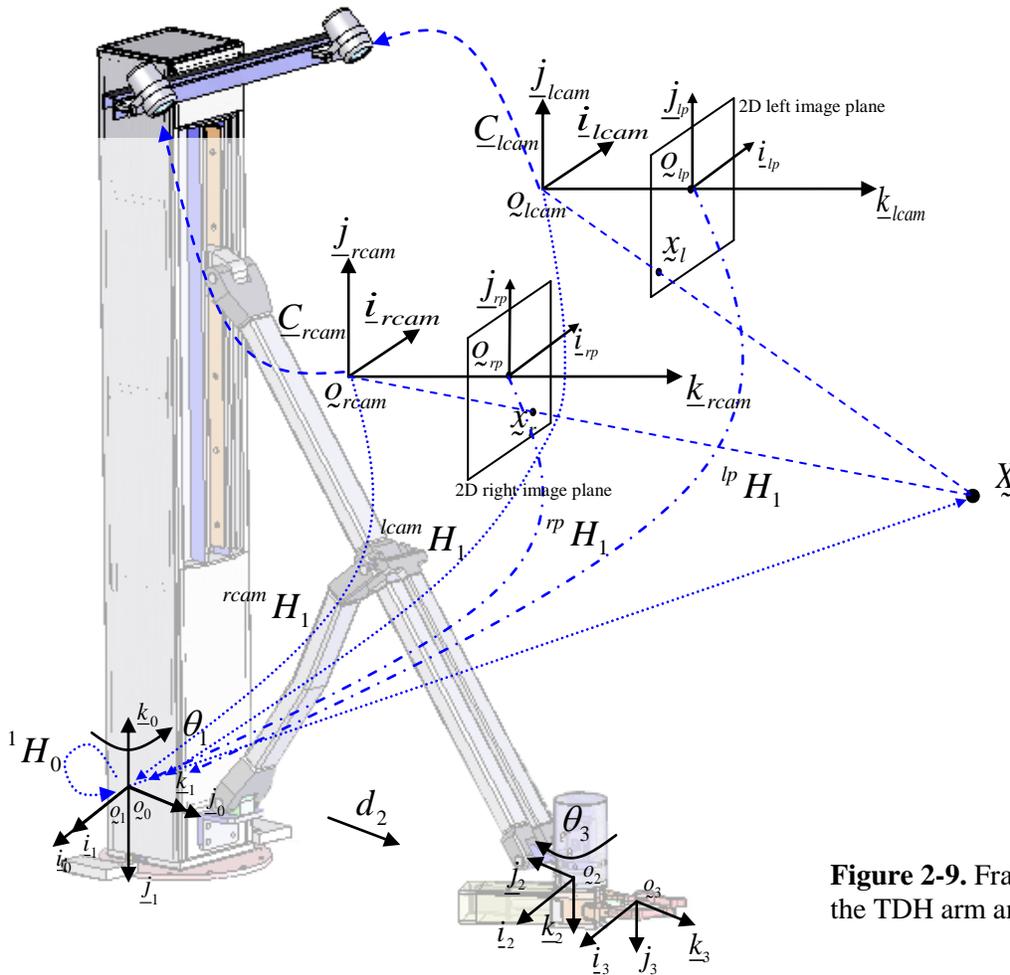


Figure 2-9. Frame assignment for the TDH arm and the stereo cameras

The description of the employed notation in Figure 2-9 is as follow:

- $\{q_0, \underline{C}_0\}$ is the base (world) coordinate system where $\underline{C}_0 = [\underline{i}_0 \quad \underline{j}_0 \quad \underline{k}_0]$ is a right-handed Cartesian frame with q_0 as its origin. This coordinate system is fixed to the monkey board.
- Point \underline{X} is represented in $\{q_0, \underline{C}_0\}$ by its homogenous coordinates ${}^0X = [x_0 \quad y_0 \quad z_0 \quad 1]^T$.
- $\{q_1, \underline{C}_1\}$ is assumed to be attached to the TDH robotic manipulator link 1 (the torso).
- $\{q_{lcam}, \underline{C}_{lcam}\}$ and $\{q_{rcam}, \underline{C}_{rcam}\}$ are the left and right camera coordinate systems respectively with q_{lcam} and q_{rcam} as the camera centers.
- Point \underline{X} is represented in the left and right camera coordinate systems by ${}^{lcam}X$ and ${}^{rcam}X$ respectively.
- The projection of point \underline{X} in the 2D image plane of the left camera is point \underline{x}_l and it is represented by the homogeneous coordinates ${}^l x = [x_l \quad y_l \quad 1]^T$ in $\{q_{lp}, \underline{C}_{lp}\}$.
- The projection of point \underline{X} in the 2D image plane of the right camera is point \underline{x}_r and it is represented by its homogeneous coordinates ${}^r x = [x_r \quad y_r \quad 1]^T$ in $\{q_{rp}, \underline{C}_{rp}\}$.
- The representation of a point in a coordinate system is referred to the coordinates of that point in that coordinates system in this thesis.

The DH table of parameters for the TDH robotic arm is as follow:

Table 2-1. DH table for the TDH robot

| Link | a_i | α_i | d_i | θ_i |
|------|-------|------------------|---------|--------------|
| 1 | 0 | $-\frac{\pi}{2}$ | 0 | θ_1^* |
| 2 | 0 | $-\frac{\pi}{2}$ | d_2^* | 0 |
| 3 | a_3 | $-\frac{\pi}{2}$ | 0 | θ_3^* |

* variable joint

The homogenous transformation between the robot gripper frame and the robot base frame can be written as follows:

$${}^0T_3 = {}^0T_1 {}^1T_2 {}^2T_3 = \begin{bmatrix} e^{\theta_1 k \times} e^{-\frac{\pi}{2} i \times} & 0 \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} e^{-\frac{\pi}{2} i \times} & d_2 k \\ 0^T & 1 \end{bmatrix} \begin{bmatrix} e^{\theta_3 k \times} e^{\frac{\pi}{2} i \times} & a_3 i \\ 0^T & 1 \end{bmatrix} \quad (2-1)$$

Point \tilde{X} is transformed to the camera coordinate systems using the homogenous transformations based on the pinhole camera model and the DLT approach (explained in Section 3.2:

$$w_l^l x = {}^r p H_1^1 H_0^0 X \quad (2-2)$$

$$w_r^r x = {}^l p H_1^1 H_0^0 X \quad (2-3)$$

Important note: In projective geometry w_l and w_r are arbitrary scale factors according to the pinhole camera model [62]. In fact, the quantities in projective geometry are determined up to a scale factor [63]. Ignoring scale factors w_l and w_r in (2-2) and (2-3) constructs a weak perspective under an affine camera model.

${}^1 H_0$ is a known transformation and can be written as follows:

$${}^1 H_0 = \begin{bmatrix} e^{-\theta_1 k x} e^{\frac{\pi}{2} i x} & 0 \\ 0^T & 1 \end{bmatrix} = ({}^0 T_1)^{-1} \quad (2-4)$$

In fact, (2-2) and (2-3) can be written in the following form considering the fact that vertical pixel coordinates in the stereo images are equal ($y_l = y_r = y$):

$$\begin{bmatrix} w_l x_l \\ (w_l + w_r) y \\ w_l \\ w_r x_r \\ w_r \end{bmatrix} = H {}^1 H_0^0 X \quad (2-5)$$

In Section 3.4, a stereo camera calibration method will be introduced to estimate the matrix H .

2.5 Hardware Interrupt Driven Data Communication

Based on the principles of the *position-based look-and-move* visual servoing architecture, the vision system (PC104) transmits the joint-space commands to the PLC and the motion controller units. The vision system does not perform the joint level control of the robotic arm as the motion controller unit performs this task. All data acquisition and transmission is via hardware interrupts. In particular the camera frame acquisition and joint-space command generation are performed in real-time with 30ms update rate. A delay of less than 30ms is introduced due to the algorithm processing load. This delay in its worst case may decrease the update rate of 30ms to approximately 28ms. Unlike the software-generated interrupts, the hardware interrupts guarantee a stationary update rate between the PC104 and the PLC systems.

A bilateral RS232 serial communication link has been established to provide the data communications between the PC104 and PLC systems.

Figure 2-10 shows the schematic of the hardware interrupt driven software threads running inside the proposed *position-based look-and-move* visual servoing system.

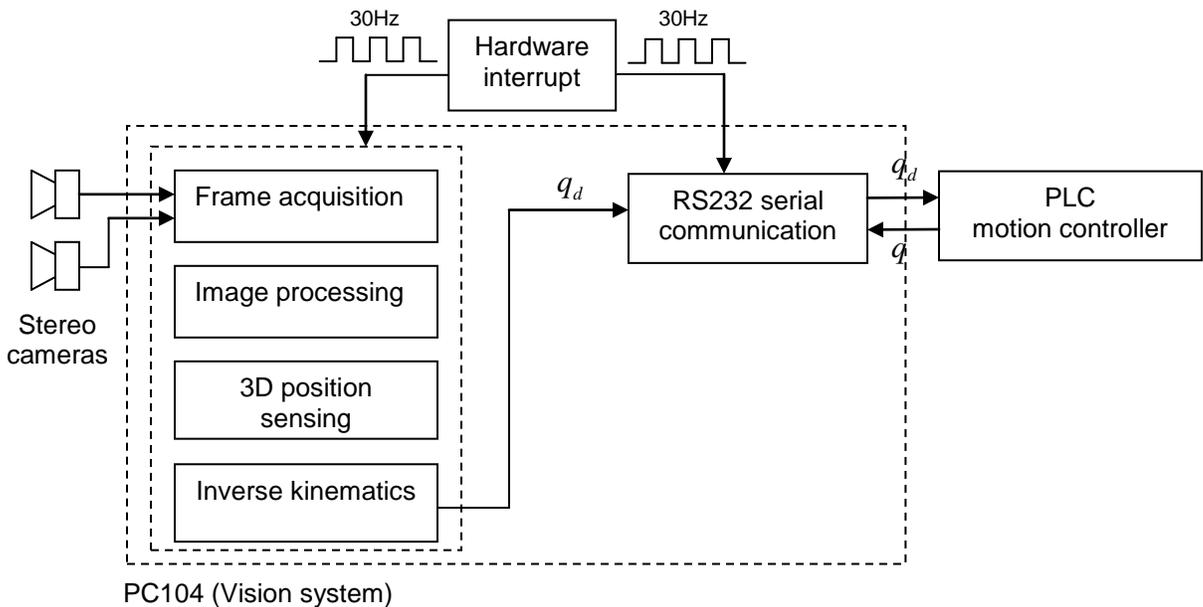


Figure 2-10. The hardware interrupt driven threads

Note: q_d is the desired robot joint angles.

Chapter 3. Stereo Camera Calibration

3.1 The Finite Camera Projection Model

3.1.1 Camera Extrinsic and Intrinsic Parameters

The general pinhole camera projection model is product of two matrices:

- 1- A 3x4 matrix of intrinsic parameters.
- 2- A 4x4 matrix of extrinsic parameters. This matrix holds the rotation and transformation parameters of the camera pose and orientation relative to an arbitrary coordinate system. In the proposed visual servoing system, the matrix of extrinsic parameters holds the rotation and transformation parameters between the camera coordinate system and the coordinate system attached to the robot torso ($\{\mathcal{Q}_1, \underline{C}_1\}$) indicated in Figure 2-9. In this figure such transformation is shown by

${}^{lcam}H_1$ and ${}^{rcam}H_1$ for the left and right cameras, respectively.

A matrix $P_{3 \times 4}$ relates the coordinates of a point \underline{X} to its image plane coordinates ${}^l x$ for the left camera:

$$w_l {}^l x = P_l {}^1 X \quad (3-1)$$

$$\text{where } P_l = K {}^{lcam}H_1 = \begin{bmatrix} K & 0 \end{bmatrix} \begin{bmatrix} \underline{R} & t \\ 0^T & 1 \end{bmatrix} = \begin{bmatrix} f & s & x_p & 0 \\ 0 & f & y_p & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 & t_x \\ r_4 & r_5 & r_6 & t_y \\ r_7 & r_8 & r_9 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3-2)$$

$K_{3 \times 3}$: Matrix of intrinsic camera parameters

$\underline{R}_{3 \times 3}$: Matrix of extrinsic orientation parameters

$t_{3 \times 1}$: Matrix of extrinsic translation parameters

f : The camera focal length

s : The angle between the x and y sensor axes of the CCD sensor

x_p and y_p : The principal point in x and y directions

${}^l x$: The image plane coordinates of \underline{X}

${}^1 X$: The homogeneous coordinates of \underline{X} in the robot frame \underline{C}_1

w_l : The scale factor according to pinhole camera model

3.2 Estimation of Camera Model Using DLT Method

3.2.1 Introduction

Estimation of the camera intrinsic and extrinsic parameters has been a long standing research in computer vision. In this thesis, extraction of the physical (intrinsic) and extrinsic parameters was not considered. Instead, the camera projection model in Equation (3-2) can be written in the format of a Direct Linear Transformation (DLT) which is a classic method in photogrammetry [74]-[76]. The DLT method was originally introduced by Abdel-Aziz and Karara in [71] and was later revised in several publications, e.g., [73]. DLT bridges the gap between photogrammetry and computer vision so that both areas can use DLT directly to solve the camera calibration problem [41]. In the process of obtaining the camera model using the DLT method, the lens distortion is ignored. It will be later shown in this chapter how the lens distortion correction can be employed to estimate the real pixel coordinates using their distorted image coordinates.

3.2.2 Implementation of the DLT Method

In the proposed visual servoing system with stereo vision, two camera models H_l and H_r should be obtained to map a point from $\{Q_1, C_1\}$ to the 2D image planes of the left and right cameras, respectively. Obviously, it is important to have the points in the FOV (Field of View) of the cameras. In fact, H_l and H_r are equal to homogeneous transformations ${}^l p H_1$ and ${}^r p H_1$, respectively. Hence the conditions below should be considered:

- $\{Q_1, C_1\}$ is attached to the column (link1) of the TDH manipulator. During the calibration process, this coordinate system is stationary with respect to $\{Q_0, C_0\}$ as the robot base angle does not change during calibration.
- ${}^0 X$ should be transformed to $\{Q_1, C_1\}$ using (2-4) and θ_1 is assumed to be zero.
- Point X is represented in $\{Q_1, C_1\}$ with its homogenous coordinates

$${}^1 X = [x_1 \quad y_1 \quad z_1 \quad 1]^T .$$

Transferring point \underline{X} from $\{\mathcal{Q}_0, \underline{C}_0\}$ to $\{\mathcal{Q}_1, \underline{C}_1\}$ can be written as follows:

$${}^1X = {}^1H_0^0X \quad (3-3)$$

where

$${}^1H_0 = \begin{bmatrix} \cos \theta_1 & 0 & -\sin \theta_1 & 0 \\ -\sin \theta_1 & 0 & -\cos \theta_1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \xrightarrow{\theta_1=0} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (3-4)$$

${}^l x$ and ${}^r x$ are the projection of point 1X in the 2D image planes of the stereo cameras.

As discussed before, the 3D coordinates of a point in $\{\mathcal{Q}_1, \underline{C}_1\}$ is mapped to 2D coordinate system of the left and right images ($\{\mathcal{Q}_{lp}, \underline{C}_{lp}\}$ and $\{\mathcal{Q}_{rp}, \underline{C}_{rp}\}$) using the equations below:

$$w_l {}^l x = {}^l H_1 {}^1 X = H_l {}^1 X \quad (3-5)$$

$$w_r {}^r x = {}^r H_1 {}^1 X = H_r {}^1 X \quad (3-6)$$

We start with the left camera to estimate the camera model parameters. The following equations can be obtained using the DLT method for the left and right cameras individually. A homogeneous 3x4 matrix can be utilized between the image and world coordinates of the calibration point. w_l and w_r are arbitrary scale factor according to pinhole camera model [61].

$$\begin{bmatrix} w_l x_l \\ w_l y_l \\ w_l \end{bmatrix} = H_l {}^1 X = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ \alpha_9 & \alpha_{10} & \alpha_{11} & \alpha_{12} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ \alpha_9 & \alpha_{10} & \alpha_{11} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} \quad (3-7)$$

$$\begin{bmatrix} w_r x_r \\ w_r y_r \\ w_r \end{bmatrix} = H_r {}^1 X = \begin{bmatrix} \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \beta_5 & \beta_6 & \beta_7 & \beta_8 \\ \beta_9 & \beta_{10} & \beta_{11} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} \quad (3-8)$$

In order to avoid a trivial solution ($\alpha_1 \cdots \alpha_{12} = 0$), a proper normalization must be applied [70]. Abdel-aziz and Karara [71] used constraint of $\alpha_{12} = 1$. The parameters ($\alpha_1 \cdots \alpha_{11}$) have no physical meaning, and this technique can be also considered as the

implicit camera calibration stage [70]. The same assumption should be utilized for the right camera in the system.

The parameter w_l can be eliminated by expanding (3-7) in the following form:

$$\begin{aligned} x_l &= \frac{\alpha_1 x_1 + \alpha_2 y_1 + \alpha_3 z_1 + \alpha_4}{\alpha_9 x_1 + \alpha_{10} y_1 + \alpha_{11} z_1 + 1} \\ y_l &= \frac{\alpha_5 x_1 + \alpha_6 y_1 + \alpha_7 z_1 + \alpha_8}{\alpha_9 x_1 + \alpha_{10} y_1 + \alpha_{11} z_1 + 1} \end{aligned} \quad (3-9)$$

From (3-8) the following equation can be obtained:

$$\begin{aligned} x_l &= \alpha_1 x_1 + \alpha_2 y_1 + \alpha_3 z_1 + \alpha_4 - \alpha_9 x_1 x_l - \alpha_{10} y_1 x_l - \alpha_{11} z_1 x_l \\ y_l &= \alpha_5 x_1 + \alpha_6 y_1 + \alpha_7 z_1 + \alpha_8 - \alpha_9 x_1 y_l - \alpha_{10} y_1 y_l - \alpha_{11} z_1 y_l \end{aligned} \quad (3-10)$$

It is important to consider that the above equations should be obtained for both left and right images in a stereo or non stereo vision system.

3.2.3 Estimation of Camera Model Parameters

The right hand side of (3-10) can be decomposed into two matrices of the known (A) and unknown (k) parameters of the camera model.

$$\underbrace{\begin{bmatrix} x_l \\ y_l \end{bmatrix}}_U = \underbrace{\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -x_l x_1 & -x_l y_1 & -x_l z_1 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -y_l x_1 & -y_l y_1 & -y_l z_1 \end{bmatrix}}_A k \quad (3-11)$$

where

$$k = [\alpha_1 \quad \alpha_2 \quad \alpha_3 \quad \alpha_4 \quad \alpha_5 \quad \alpha_6 \quad \alpha_7 \quad \alpha_8 \quad \alpha_9 \quad \alpha_{10} \quad \alpha_{11}]^T$$

So we write (3-11) in the form of the equation below.

$$U_{2 \times 1} = A_{2 \times 11} k_{11 \times 1} \quad (3-12)$$

The DLT camera calibration method solves the above equation to obtain the components of column matrix k of the left camera. The components of matrix $A_{2 \times 11}$ are constructed from the Cartesian coordinates ${}^1X = [x_1 \quad y_1 \quad z_1 \quad 1]^T$ and image coordinates (x_l, y_l) . Matrix $U_{2 \times 1}$ is column matrix containing the image coordinates of the calibration object in the left camera plane.

One approach to solve the 11 unknown parameters of matrix $k_{1 \times 1}$ is to construct a linear system of equations with at least 11 equations. Matrices $A_{2 \times 11}$ and $U_{2 \times 1}$ are constructed using only one calibration point, so employing six calibration points can lead to resizing these two matrices to $A_{12 \times 12}$ and $U_{12 \times 1}$. In fact the more points collected in the calibration process, the more overdetermined the linear system of equations. Collecting N points in the calibration process, (3-12) can be generalized as follows:

$$A_{2N \times 11} \cdot k_{1 \times 1} = U_{2N \times 1} \quad (3-13)$$

The equation below shows the expanded form of (3-13):

$$\begin{bmatrix} A_{12 \times 11} \\ A_{22 \times 11} \\ \vdots \\ A_{N2 \times 11} \end{bmatrix} \cdot k_{1 \times 1} = \begin{bmatrix} U_{12 \times 1} \\ U_{22 \times 1} \\ \vdots \\ U_{N2 \times 1} \end{bmatrix} \quad (3-14)$$

Obviously, to solve the 11 unknown parameters of $k_{1 \times 1}$, N should at least be equal to 6.

Consequently the *normal equation* of the below form can be solved to find k .

$$A^T A k = A^T U \Rightarrow k = (A^T A)^{-1} A^T U \quad (3-15)$$

In Section 3.3.3, an iterative least-squares method is provided to avoid the solution of the large matrices $A^T A$ and $A^T U$.

3.3 Least-Squares Solution to Obtain Camera Model Parameters

3.3.1 Parameter Estimation by Minimizing LS Loss Function

As explained in previous section, collecting a lot of data points in the calibration process constructs an overdetermined linear system of equations. This section describes the preparation of the known and unknown parameters in the calibration process to be solved using least-squares estimation.

Considering the least-squares loss function:

$$V(\theta, t) = \frac{1}{2} \sum_{i=1}^t (v(i) - \varphi^T(i)\theta)^2 \quad (3-16)$$

The measured variable v is linear in parameters θ and the least-squares criterion is quadratic. The notation below is used to associate the corresponding parameters:

The residuals $\varepsilon(i)$ are defined by:

$$\varepsilon(i) = v(i) - \hat{v}(i) = v(i) - \varphi^T(i)\theta \quad (3-17)$$

The loss function (3-16) can be written as:

$$V(\theta, t) = \frac{1}{2} \sum_{i=1}^t \varepsilon^2(i) = \frac{1}{2} \mathbf{E}^T \mathbf{E} = \frac{1}{2} \|\mathbf{E}\|^2 \quad (3-18)$$

$$\mathbf{E} = \mathbf{V} - \hat{\mathbf{V}} = \mathbf{V} - \Phi \theta \quad (3-19)$$

The function of (3-16) is minimal for parameters $\hat{\theta}$ considering:

$$\Phi^T \Phi \hat{\theta} = \Phi^T \mathbf{V} \quad (3-20)$$

If the matrix $\Phi^T \Phi$ is nonsingular, the minimum is unique and given by:

$$\hat{\theta} = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{V} \quad (3-21)$$

Equation (3-21) can be written as:

$$\hat{\theta} = \left(\sum_{i=1}^N \varphi^T(i)\varphi(i) \right)^{-1} \left(\sum_{i=1}^N \varphi^T(i)v(i) \right) = P(t) \left(\sum_{i=1}^N \varphi(i)v(i) \right) \quad (3-22)$$

The condition that the matrix $\Phi^T \Phi$ is invertible is called an *excitation* condition.

3.3.2 Associating the LS Components with Camera Parameters

We define the image coordinates in the camera plane as the measured variables for the least-squares approach. The least-squares approach should be applied to the left and right cameras separately to obtain the camera model parameters for each camera.

The measured variables can be the results of a successful template matching or any other image processing procedures providing the associated pixel coordinates to the calibration object in the robot base frame. The parameters of the known function Φ are constructed from the known world and image coordinates of the calibration object. Considering (3-16) and defining the column matrix Y as the measured variables in one of the cameras:

$$[v(1) \ \cdots \ v(N)]^T = [(x_l)_1 \ (y_l)_1 \ \cdots \ (x_l)_N \ (y_l)_N]^T \quad (3-23)$$

The matrix of the known functions $\Phi(N)$ can be formed for several data points in (3-10):

$$\begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_N \end{bmatrix} = \begin{bmatrix} (x_1)_1 & (y_1)_1 & (z_1)_1 & 1 & 0 & 0 & 0 & 0 & (-x_1 x_1)_1 & (-x_1 y_1)_1 & (-x_1 z_1)_1 \\ 0 & 0 & 0 & 0 & (x_1)_1 & (y_1)_1 & (z_1)_1 & 1 & (-y_1 x_1)_1 & (-y_1 y_1)_1 & (-y_1 z_1)_1 \\ \vdots & \vdots \\ (x_1)_N & (y_1)_N & (z_1)_N & 1 & 0 & 0 & 0 & 0 & (-x_1 x_1)_N & (-x_1 y_1)_N & (-x_1 z_1)_N \\ 0 & 0 & 0 & 0 & (x_1)_N & (y_1)_N & (z_1)_N & 1 & (-y_1 x_1)_N & (-y_1 y_1)_N & (-y_1 z_1)_N \end{bmatrix} \quad (3-24)$$

Like previous section, N is the number of collected points during the calibration process. The matrices of the known and measured parameters for both left and right cameras can be as follows when considering (3-12). Subscripts l and r denote the left and right cameras:

$$\varphi_{l_i}^T \Big|_{i=1 \dots N} = [\varphi_{l_1} \ \cdots \ \varphi_{l_N}]^T = [A_{l_1 \ 2 \times 12} \ A_{l_2 \ 2 \times 12} \ \cdots \ A_{l_N \ 2 \times 12}]^T \quad (3-25)$$

$$\varphi_{r_i}^T \Big|_{i=1 \dots N} = [\varphi_{r_1} \ \cdots \ \varphi_{r_N}]^T = [A_{r_1 \ 2 \times 12} \ A_{r_2 \ 2 \times 12} \ \cdots \ A_{r_N \ 2 \times 12}]^T \quad (3-26)$$

$$v_{l_i}^T \Big|_{i=1 \dots N} = [v_{l_1 \ 2 \times 1} \ \cdots \ v_{l_N \ 3 \times 1}]^T = [U_{l_1 \ 2 \times 1} \ U_{l_2 \ 2 \times 1} \ \cdots \ U_{l_N \ 2 \times 1}]^T \quad (3-27)$$

$$v_{r_i}^T \Big|_{i=1 \dots N} = [v_{r_1 \ 2 \times 1} \ \cdots \ v_{r_N \ 2 \times 1}]^T = [U_{r_1 \ 2 \times 1} \ U_{r_2 \ 2 \times 1} \ \cdots \ U_{r_N \ 2 \times 1}]^T \quad (3-28)$$

After collecting N calibration points in the FOV of the left and right cameras, φ_{l_i} , φ_{r_i} , v_{l_i} and v_{r_i} with $i=1 \dots N$ will be obtained to be used in (3-22). Hence the camera parameters for the left and right cameras can be estimated as follows:

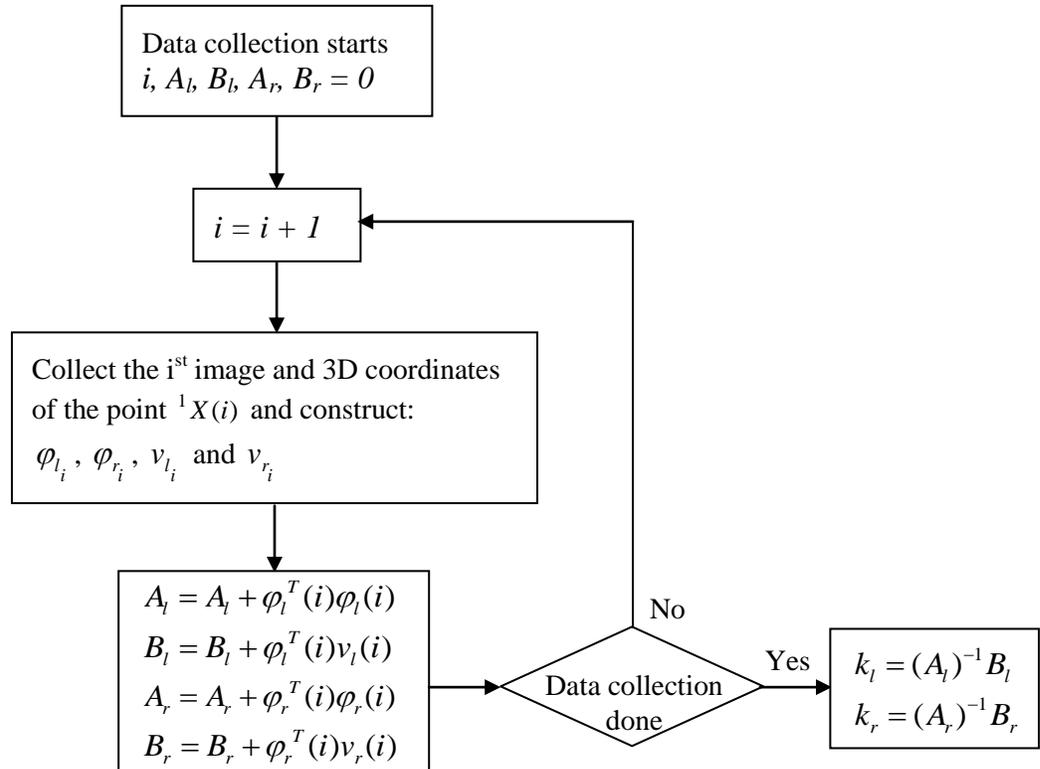
$$\text{Left camera parameters} = k_l = \hat{\theta}_l = \left(\sum_{i=1}^N \varphi_{l_i}^T(i) \varphi_{l_i}(i) \right)^{-1} \left(\sum_{i=1}^N \varphi_{l_i}^T(i) v_{l_i}(i) \right) \quad (3-29)$$

$$\text{Right camera parameters} = k_r = \hat{\theta}_r = \left(\sum_{i=1}^N \varphi_{r_i}^T(i) \varphi_{r_i}(i) \right)^{-1} \left(\sum_{i=1}^N \varphi_{r_i}^T(i) v_{r_i}(i) \right) \quad (3-30)$$

3.3.3 An Iterative LS Approach to Find the Camera Parameters

The least-squares estimation of the camera parameters explained in (3-29) and (3-30) requires the system to hold the matrices of the known and unknown parameters φ_{l_i} , φ_{r_i} , v_{l_i} and v_{r_i} (with $i=1 \dots N$) during the data collection process. Once this period has come to its end, the summation operators will be employed.

A very efficient way of solving the least-squares problem with the explained format is to perform the summation operation after each sample data is collected. See [77] for further discussion of iterative LS approach. The flowchart below shows this process:



The main advantages of the proposed iterative method over the least-squares solution:

- 1- The computation time in the iterative method is distributed between the data sampling process. The system does not need to solve (3-29) and (3-30) for $i = 1 \dots N$ after the data collection period.
- 2- The iterative method is memory efficient as the matrices φ_{l_i} , φ_{r_i} , v_{l_i} and v_{r_i} are not required to be held in the memory for $i = 1 \dots N$.

3.4 A Reduced Stereo Camera Model Calibration

3.4.1 Introduction

So far, the camera calibration method was performed to calibrate the left and right cameras individually. This method does not assume any geometric relationship between the image planes (e.g., parallel and coplanar image planes). The obtained models H_l and H_r will be later used in a triangulation method to provide the 3D estimation of the point in the FOV of the cameras.

In this section the fact that the left and right cameras obey the stereoscopic constraints is considered. As mentioned in Section 2.2.1, the most important property in systems with stereo images is to have equal vertical displacements on the stereo image coordinates $y_l = y_r$. If the cameras can be installed on the stereo rig to a given tolerance then one can add the middle rows of the equations (3-7) and (3-8) together considering $y_l = y_r = y$. It will be shown that fewer number of parameters will be introduced using this technique.

To distinguish the camera calibration methods performed in this thesis, each method has been named as following:

- The conventional method of camera calibration using the DLT camera model estimation remains unchanged as “DLT”.
- The term “Stereo” is added to “DLT” in this thesis to form “SDLT” method as for when a model is obtained for the stereo cameras.

3.4.2 System Parameterization

A 5x4 matrix can be utilized to relate the image coordinates of the calibration object (${}^l x = [x_l \ y_l \ 1]^T$ and ${}^r x = [x_r \ y_r \ 1]^T$) to its 3D homogenous coordinates ${}^1 X = [x_1 \ y_1 \ z_1 \ 1]^T$ in the robot torso coordinate system $\{q_1, \underline{C}_1\}$. (3-3) can be considered again to obtain ${}^1 X$. Considering (3-7) and (3-8), the following reduced equation can be utilized assuming $y_l = y_r = y$ while the second rows of the matrix equations (in (3-7) and (3-8)) have been added together.

$$\begin{bmatrix} w_l x_l \\ (w_l + w_r)y \\ w_l \\ w_r x_r \\ w_r \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ (\alpha_5 + \beta_5) & (\alpha_6 + \beta_6) & (\alpha_7 + \beta_7) & (\alpha_8 + \beta_8) \\ \alpha_9 & \alpha_{10} & \alpha_{11} & 1 \\ \beta_1 & \beta_2 & \beta_3 & \beta_4 \\ \beta_9 & \beta_{10} & \beta_{11} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = H^{-1}X \quad (3-31)$$

The matrix equation above can be reorganized by defining 18 parameters as follows:

$$\begin{bmatrix} w_l x_l \\ (w_l + w_r)y \\ w_l \\ w_r x_r \\ w_r \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 \\ \varphi_5 & \varphi_6 & \varphi_7 & \varphi_8 \\ \varphi_9 & \varphi_{10} & \varphi_{11} & 1 \\ \varphi_{12} & \varphi_{13} & \varphi_{14} & \varphi_{15} \\ \varphi_{16} & \varphi_{17} & \varphi_{18} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = H^{-1}X \quad (3-32)$$

Scale factors w_l and w_r can be eliminated by writing the equations in (3-32) as follows:

$$\begin{bmatrix} x_l \\ 2y \\ x_r \end{bmatrix} = \begin{bmatrix} \varphi_1 x_1 + \varphi_2 y_1 + \varphi_3 z_1 + \varphi_4 - \varphi_9 x_1 x_l - \varphi_{10} y_1 x_l - \varphi_{11} z_1 x_l \\ \varphi_5 x_1 + \varphi_6 y_1 + \varphi_7 z_1 + \varphi_8 - \varphi_9 x_1 y - \varphi_{16} x_1 y - \varphi_{10} y_1 y - \varphi_{17} y_1 y - \varphi_{11} z_1 y - \varphi_{18} z_1 y \\ \varphi_{12} x_1 + \varphi_{13} y_1 + \varphi_{14} z_1 + \varphi_{15} - \varphi_{16} x_1 x_r - \varphi_{17} y_1 x_r - \varphi_{18} z_1 x_r \end{bmatrix} \quad (3-33)$$

The three equations in (3-33) can be rearranged to yield

$$\underbrace{\begin{bmatrix} x_l \\ 2y \\ x_r \end{bmatrix}}_U = \underbrace{\begin{bmatrix} x_1 & y_1 & z_1 & 1 & 0 & 0 & 0 & 0 & -x_1 x_l - y_1 x_l - z_1 x_l & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -x_1 y - y_1 y - z_1 y & 0 & 0 & 0 & 0 & -x_1 y - y_1 y - z_1 y & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_1 & y_1 & z_1 & 1 & -x_1 x_r - y_1 x_r - z_1 x_r \end{bmatrix}}_D k \quad (3-34)$$

where $k = [\varphi_1 \ \varphi_2 \ \varphi_3 \ \varphi_4 \ \varphi_5 \ \varphi_6 \ \varphi_7 \ \varphi_8 \ \varphi_9 \ \varphi_{10} \ \varphi_{11} \ \varphi_{12} \ \varphi_{13} \ \varphi_{14} \ \varphi_{15} \ \varphi_{16} \ \varphi_{17} \ \varphi_{18}]^T$.

$$U_{3 \times 1} = D_{3 \times 18} k_{18 \times 1} \quad (3-35)$$

Collecting N number of calibration points resizes (3-35) to:

$$U_{3N \times 1} = D_{3N \times 18} k_{18 \times 1} \quad (3-36)$$

Using the iterative method in Section 3.3.3, the unknown parameters can be obtained as the equivalent of (3-15) except that both set of camera parameters are obtained at once.

$$k = \left(\sum_{i=1}^N D^T(i)D(i) \right)^{-1} \left(\sum_{i=1}^N D^T(i)U(i) \right) \quad (3-37)$$

3.5 Different Uses of the Camera Models

The DLT classic method of camera calibration calculates the camera models as stated in (3-5) and (3-6):

$$w_l {}^l x = H_l {}^1 X$$

$$w_r {}^r x = H_r {}^1 X$$

The SDLT method, in comparison, provides a single transformation for both stereo cameras as stated in (3-32):

$$\begin{bmatrix} w_l x_l \\ (w_l + w_r)y \\ w_l \\ w_r x_r \\ w_r \end{bmatrix} = H {}^1 X$$

Three different scenarios can be considered in Table 3-1 and Table 3-2 for the above equations.

Table 3-1. Camera calibration and 3D estimation procedures for the DLT method

| Given | Procedure | Find |
|---------------------------------|-------------------------------|-----------------------------------|
| ${}^l x, {}^r x$ and ${}^1 X$ | Camera calibration | H_l and H_r |
| ${}^l x, {}^r x, H_l$ and H_r | 3D position estimation | ${}^0 X$ (explained in Chapter 4) |
| H_l, H_r and ${}^1 X$ | Camera calibration evaluation | ${}^l x$ and ${}^r x$ |

Similar to what stated for the classic method, the table below indicates the three conditions for different known and unknown components:

Table 3-2. Camera calibration and 3D estimation procedures for SDLT method

| Given | Procedure | Find |
|-------------------------------|-------------------------------|-----------------------------------|
| ${}^l x, {}^r x$ and ${}^1 X$ | Camera calibration | H |
| ${}^l x, {}^r x$ and H | 3D position estimation | ${}^0 X$ (explained in Chapter 4) |
| H and ${}^1 X$ | Camera calibration evaluation | ${}^l x$ and ${}^r x$ |

3.6 Data Collection Techniques for Camera Calibration

In this thesis two different methods of data collection for camera calibration have been employed. As indicated in the first rows of Table 3-1 and Table 3-2, the camera calibration technique requires X as the coordinates of the calibration object and ${}^l x$ and ${}^r x$ as image coordinates. The two techniques for automatic data collection are:

- 1- Using a 3DOF Cartesian robot to move the calibration object in front of the cameras. In this thesis this method is called “*camera calibration using a Cartesian robot*”.
- 2- Using the torso-mounted TDH robotic manipulator itself to move the calibration object in the cameras FOV. This method is called “*In situ camera calibration*”.

3.6.1 Camera Calibration Using a Cartesian Robot

The volume of interest in the camera FOV (Field of View) should be considered for the data collection in the proposed camera calibration process. We assume this volume is a cube and will be referred to as the *calibration cube* in this thesis. The requirements of the application dictate the dimensions and position of this cube. Simply, the calibration cube is the volume where the 3D position sensing is happening. The calibration object should be moved inside the calibration cube. The corresponding coordinates in $\{Q_0, C_0\}$ and image planes are captured at the same time. As shown in Figure 3-1 the calibration robot has been placed within a certain distance from the TDH robot. The position for the calibration cube (where the pipe 3D detection is performed) is selected based on the application description explained in Sections 1.1 and 1.2. Extra care was taken to transfer the coordinates of the object to the robot base frame since the position of the calibration object is obtained relative to the calibration robot frame.

The calibration object is a bright LED which is mounted carefully on the Z actuator of the Cartesian calibration robot (shown in Figure 3-2). Increasing the number of LEDs helps to decrease the required motion on the calibration robot when the calibration cube is desired to be navigated. For the sake of simplicity only one LED case was considered for data collection in this thesis, but the system can be also calibrated using two LEDs mounted on an aluminum bracket (shown in Figure 3-2).

3.6.1.1 Camera Calibration Setup

The schematic in Figure 3-1 shows the camera calibration setup when a Cartesian robot is used to move the calibration object (LED) in the cameras FOV.

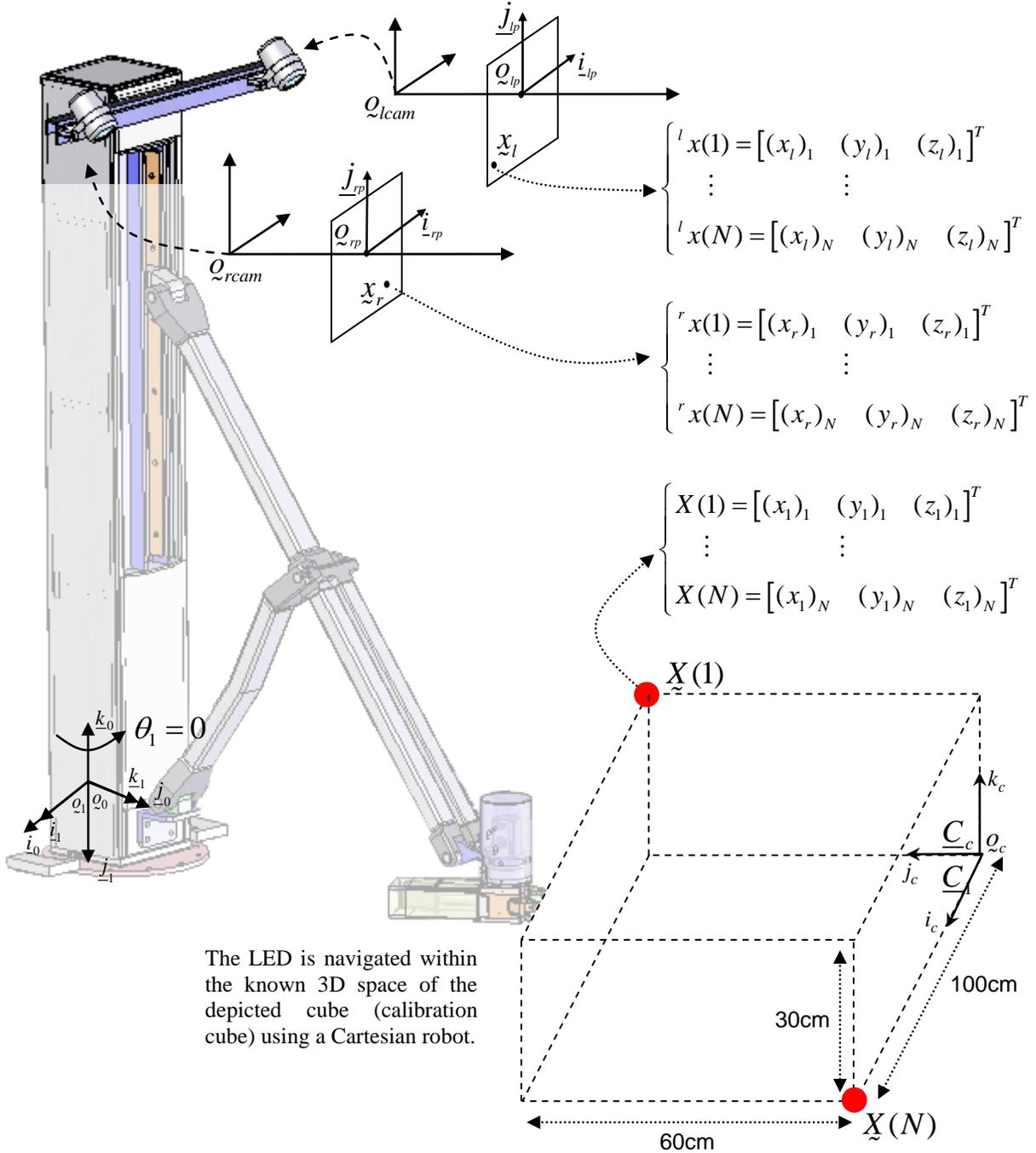


Figure 3-1. Schematic diagram of data collection using Cartesian robot

Notes regarding the schematic depicted in Figure 3-1:

- 1- $\{q_1, C_1\}$ is attached to the TDH robot column.
- 2- $X(1) \cdots X(N)$ is the set of collected coordinates of the position of the LED moved in the calibration cube using the employed Cartesian robot.
- 3- $\{q_c, C_c\}$ is the base coordinate system attached to the Cartesian robot. The required homogenous transformation required to transform the coordinates of the calibration point (LED) to $\{q_0, C_0\}$ is not mentioned in this thesis since it is a simple transformation.
- 4- The number of points collected from the scene depends on the sampling rate of the frame and data acquisition system used to read the configuration of the Cartesian robot.
- 5- $({}^l x(1) \cdots {}^l x(N))$ and $({}^r x(1) \cdots {}^r x(N))$ are the corresponding image coordinates of the data points $X(1) \cdots X(N)$ in the calibration cube.

3.6.1.2 The Specification of the Cartesian Camera Calibration Robot

As stated in the camera calibration schematic (Figure 3-1), a bright LED was navigated within the FOV of the cameras rigidly attached to the robot body. Image binerization was performed to distinguish the emitted LED light from the other dark objects. The emitted light from the LEDs appears as white or “1” in the binarized images, so the camera perception from the LED light could be mismatched with other white objects in the scene. To address this problem, two different approaches were considered in this work:

1. Perform the calibration in a dark environment to make the emitted lights from the LEDs as the only source of illumination. The binarized images were then searched to find the coordinates of the bright spots in the image frame.
2. Locally search for the LED lights in the binarized image while a black background was added to the LED structure. The black background is shown in Figure 3-2.

The Cartesian robot similar to a gantry robot moves the bright LED in the depicted cube in Figure 3-1. The properties of the employed Cartesian robot are as below:

1. The Cartesian robot (calibration robot) has three DOF. The robot moves the calibration object (bright LED) along its three linear actuators (x , y and z).
2. The workspace of the robot is limited to 50cm on the X axis, 30cm on the Y axis and 30cm on the Z axis.
3. Every axis is equipped with an encoder providing the real-time displacement on the respective dimension.
4. The real-time encoder values are transferred to the PC104 system via the RS232 serial communication. The serial communication update rate is 30ms.
5. The received encoder values of the 3D position of the LED will be matched with its corresponding image coordinates to be used in the camera calibration process.
6. The position of the origin of the calibration robot as well as its frame orientation is precisely measured from the origin of $\{q_0, C_0\}$.

7. A homogeneous transformation is required to transfer the LED coordinates from the calibration robot coordinate system $\{e_c, C_c\}$ to the TDH robot base coordinate system $\{e_0, C_0\}$.
8. The PLC/motion controller unit generates the motion commands for the calibration robot.
9. The PLC unit is programmed to send the required trajectory commands to the calibration robot. The LED trajectory covers different layers of the calibration cube in x and y directions. Raising the LED structure along the Z axis provides a new layer in the calibration cube. The calibration cube can be seen in Figure 5-1.
10. Due to the system delays, the update rate of the encoder values (in PC104) is slightly variable between 28ms and 30ms. This means that approximately 1800 data points of the calibration object (including the world and image coordinates) are collected every minute.

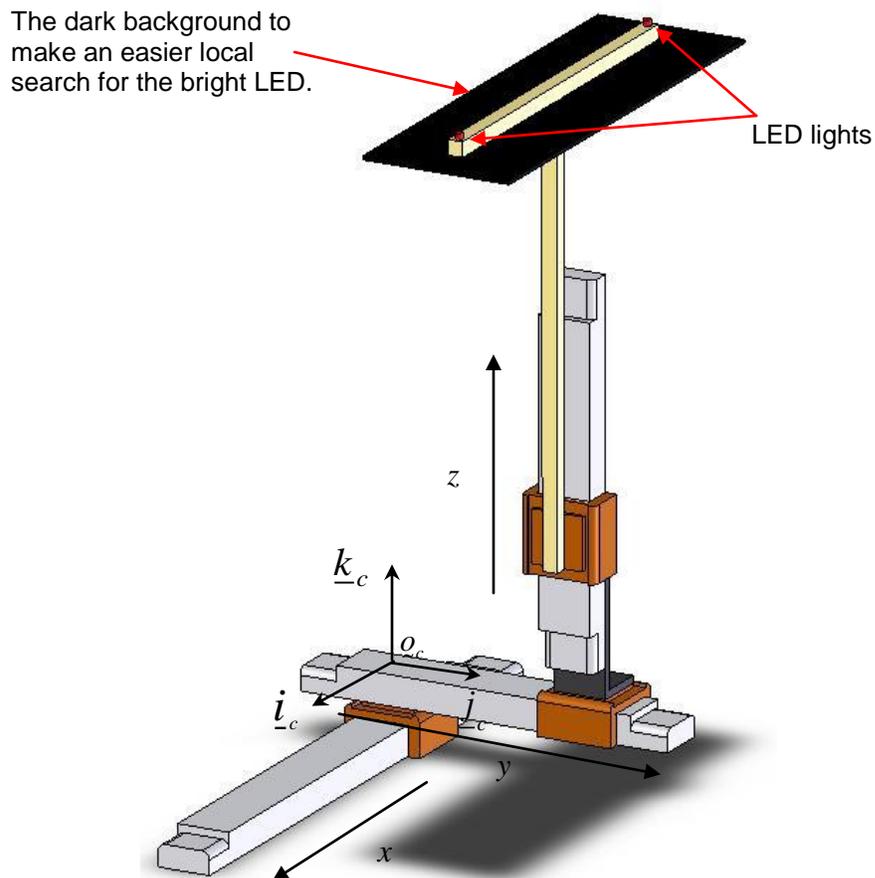


Figure 3-2. The Cartesian calibration robot

3.6.2 *In situ* camera Calibration

The term “*in situ*” refers to “in place” or “on site”. In this application the camera calibration is called “*in situ*” because the torso-mounted robot (TDH) is selected as the main manipulator to move the calibration object in the desired volume. The advantages of employing the so-called “*in situ*” calibration are:

- 1- No additional manipulator such as the Cartesian robot explained in previous section is required to move the calibration object.
- 2- The proposed visual servoing application can track and capture the object in the reachable space of the robot. The best manipulator to cover the reachable space of the TDH robot is the robot itself.
- 3- Employing another robot to move the calibration object requires the calibration robot to be placed precisely in a known position to the TDH robot base frame $\{q_0, C_0\}$. Thus one needs to transform the coordinates of the calibration object to the TDH robot base frame. Assigning a permanent location for the calibration robot while it is not attached to the TDH robot is not a straightforward job. By rigidly installing the *in situ* calibration tool on the end-effector of the robot, the coordinates of the calibration object (LED) can be calculated by solving the robot arm forward kinematic in joint-space.
- 4- The *in situ* approach is less expensive.
- 5- From the industrial point of view, when the system is installed on its real-world structure (top of the oil well mast), any damage to either of the stereo cameras perturbs the vision system accuracy and performance. Replacing the damaged camera requires another calibration process in the field since the camera model parameters may vary slightly with the previous settings. The *in situ* camera calibration provides this opportunity to calibrate the cameras using the TDH robot while it is placed on top of the oil well. There is no need to employ the mentioned Cartesian robot or any other manipulator to move the calibration object in the scene.
- 6- With proper training, the operators or the technicians at a mobile oil rig can perform the *in situ* calibration without high level knowledge of the vision system.

- 7- Unlike the case when another robot is employed for the calibration purpose, the *in situ* calibration process is independent of the orientation of the monkey board. The calibration tool orientation and position are local to the monkey board and the robot base frame $\{\mathcal{C}_0, \underline{C}_0\}$. Changes on orientation of the monkey board do not affect the transformation between the calibration tool and the TDH base frame. This is absolutely helpful as in the real-world application of this system there is no guaranty to have the monkey board in horizontal plane.

3.6.2.1 System Parameters for *In situ* Camera Calibration

The most important part of the in situ camera calibration system is the calibration tool that has been specifically designed and fabricated. The tool accommodates the objects to be tracked by the cameras while their 3D coordinates are known using the robot joint sensors. Figure 3-3 shows the *in situ* camera calibration concept.

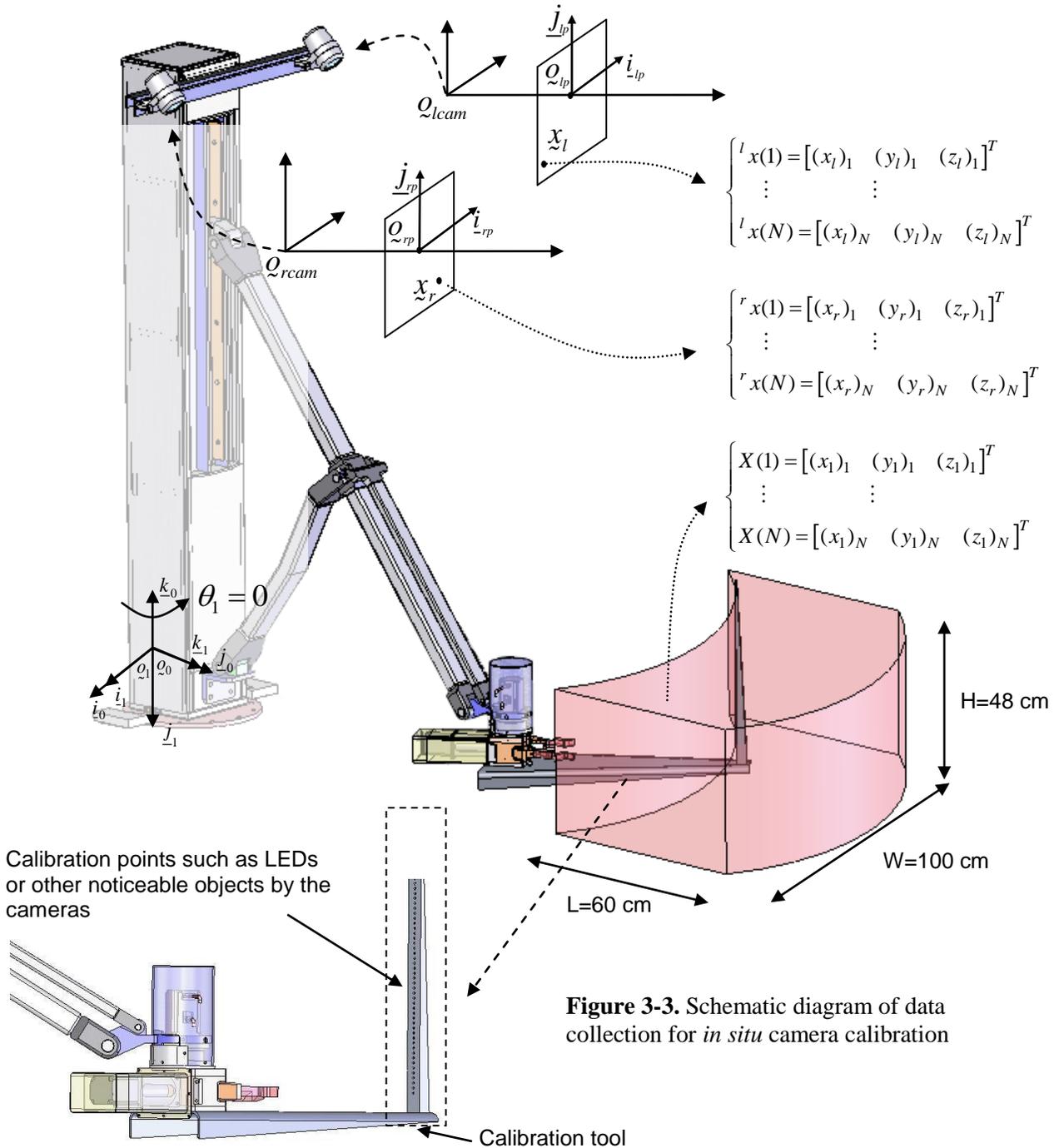


Figure 3-3. Schematic diagram of data collection for *in situ* camera calibration

3.6.2.2 Specification of the *In Situ* Camera Calibration Setup

As depicted in Figure 3-3, the calibration volume is a sector with parallel sides. The specifications below were used for the *in situ* calibration setup:

- 1- The calibration volume is created by moving the calibration tool containing the LEDs with equal distances on its vertical bar.
- 2- Only two degrees of freedom (d_2 and θ_3) have been used to move the calibration tool. The gripper radial and revolute motions are the two degrees of freedom required to create the depicted volume. Radial and revolute motions of 60cm and 45 degrees on d_2 and θ_3 respectively produce the depicted volume. The dimensions of the volume are stated in the Figure 3-3.
- 3- $\{q_1, \underline{C}_1\}$ is attached to the TDH robot column.
- 4- $X(1) \cdots X(N)$ is the set of collected coordinates of the position of the LED moved in the calibration volume. A forward kinematic process calculates these coordinates using the two active joint angles d_2 and θ_3 .
- 6- $({}^l x(1) \cdots {}^l x(N))$ and $({}^r x(1) \cdots {}^r x(N))$ are the corresponding image coordinates of the data points $X(1) \cdots X(N)$ in the calibration cube.
- 5- A dark background plate has been added to the calibration tool to simplify the local search for the lit LEDs. With this in place, the calibration process can be performed in a room with normal lighting.
- 6- The sampling rate of the data collection process is 25Hz. This means that 25 frames and 25 3D coordinates of the calibration object are captured every second during the process. Considering this sampling rate, if 40000 points were captured then the process takes approximately 2.6min.
- 7- The calibration tool is specifically designed and fabricated to be connected to the gripper structure with the highest possible precision. The tool is not a permanent part of the TDH robot and is solely used for camera calibration purposes.

- 8- As the calibration points (LEDs) have been located on a vertical line on the calibration tool, they all share the same coordinates along \underline{i}_1 and \underline{j}_1 with different z_1 coordinates along \underline{k}_1 as their heights.

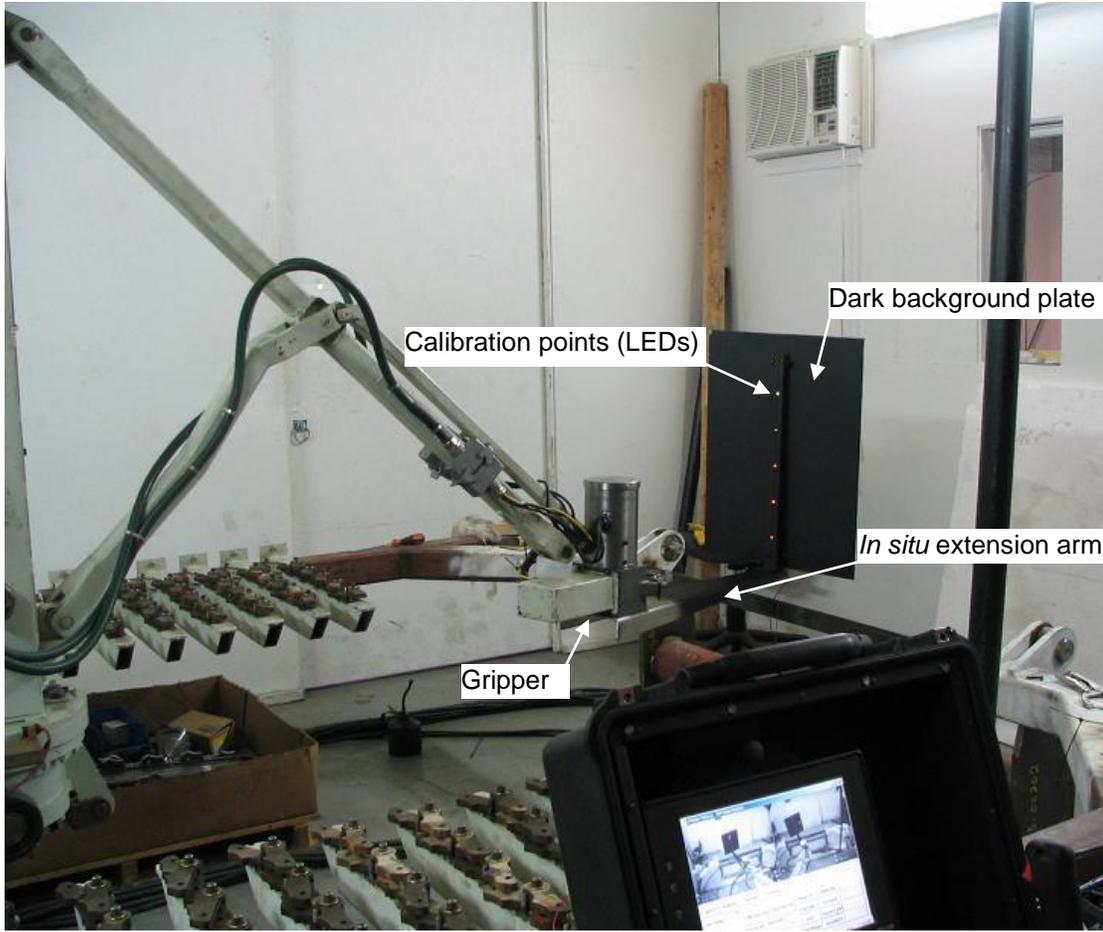


Figure 3-4. The *in situ* camera calibration setup

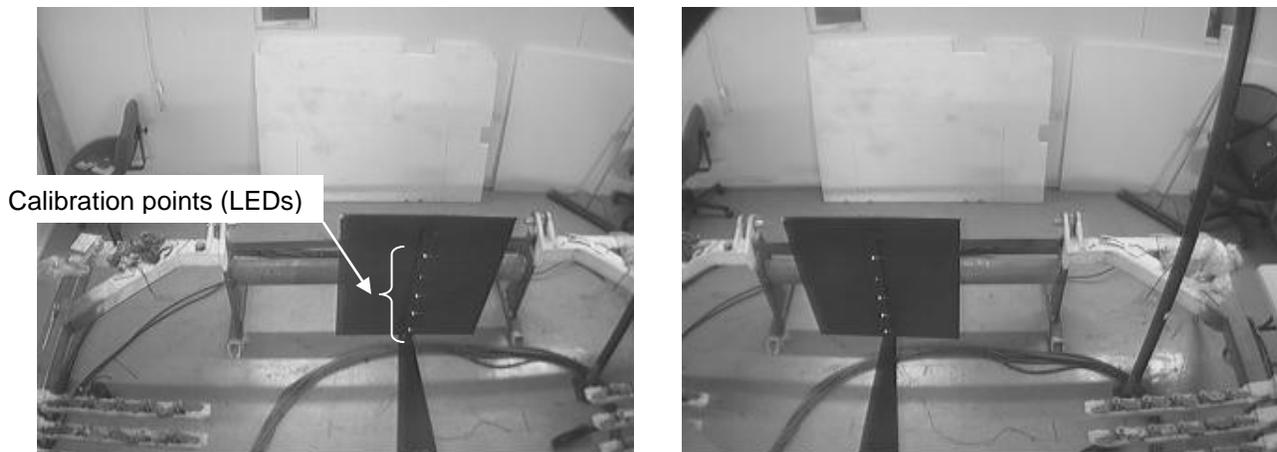


Figure 3-5. The left and right image views of the *in situ* calibration tool

3.7 Lens Distortion Correction

3.7.1 Introduction

Up to this point in the thesis, lens distortion has not been considered for the cameras in the stereo rig. The lenses employed in the cameras in this work are conventional non-linear micro lenses used in surveillance and security camera systems. In this section, only the radial lens distortion has been considered. The process to obtain the lens distortion model has been greatly influenced by the major contributions in this area by Zhang, Tsai and Heikkila et al.[62], [41] and[70], respectively. More elaborate models have been obtained in [64], [67]. The distortion model is totally dominated by radial components and more elaborate modeling not only would not help but may also introduce numerical instability [41], [68].

To obtain the lens distortion model of the cameras in a stereo system, the Matlab toolbox developed by Bouguet [68] has been employed. The lens distortion models obtained for each of the cameras are then applied to the image coordinates of the left and right camera planes individually. It has been determined that the accuracy of the 3D position sensing was improved by considering the lens distortion correction and pixel coordinates compensation.

Lens distortion has direct effects on the visual servoing applications. The accuracy measurements performed in Section 5.2 shows how much error the system experiences for the objects at the edges of the accuracy measurement area. However, the 3D positions of the objects placed in the middle of the robot workspace and consequently in the middle of the image planes are not influenced significantly by the lens distortion correction.

Since distortion calibration is only required once for a camera lens, and it cannot be changed in the field, it is not necessary to make this step part of the camera calibration parameters in the field.

3.7.2 Lens Distortion Extraction

3.7.2.1 Distortion Coefficients Calculation

The direct method of stereo camera calibration ignores the lens distortion components. In the proposed visual servoing system the lens distortion correction has been applied to the obtained (distorted) image coordinates to calculate their ideal (undistorted) image coordinates in the image plane. The first step in lens distortion correction is to obtain the distortion coefficients.

The Matlab calibration toolbox developed in [69] was used to obtain the distortion coefficients. The process requires images of a planar taken in different orientations and some user inputs to calculate the distortion coefficients.

In order to provide a background of the process the employed Matlab toolbox performs, only the left camera image components have been considered here.

The relation below is held between the real observed image coordinates and ideal (nonobservable distortion-free) image coordinates:

$${}^l x_u = {}^l x_d + ({}^l x_d - x_p)[k_1({}^l x'^2 + {}^l y'^2) + k_2({}^l x'^2 + {}^l y'^2)^2] \quad (3-38)$$

$${}^l y_u = {}^l y_d + ({}^l y_d - y_p)[k_1({}^l x'^2 + {}^l y'^2) + k_2({}^l x'^2 + {}^l y'^2)^2] \quad (3-39)$$

where:

$({}^l x_d, {}^l y_d)$: Real (distorted) image coordinates

$({}^l x_u, {}^l y_u)$: Ideal (undistorted) image coordinates

(x_p, y_p) : Principal point

$({}^l x', {}^l y')$: Normalized ideal image coordinates

k_1, k_2 : Distortion coefficients to be calibrated

The system of equations below can be formed to solve for the distortion parameters k_1 and k_2 :

$$\begin{bmatrix} ({}^l x_d - x_p)({}^l x'^2 + {}^l y'^2) & ({}^l x_d - x_p)({}^l x'^2 + {}^l y'^2)^2 \\ ({}^l y_d - y_p)({}^l x'^2 + {}^l y'^2) & ({}^l y_d - y_p)({}^l x'^2 + {}^l y'^2)^2 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \end{bmatrix} = \begin{bmatrix} {}^l x_u - {}^l x_d \\ {}^l y_u - {}^l y_d \end{bmatrix} \quad (3-40)$$

The distortion parameters then can be found by solving the above equations for n number of images and m points in every image.

3.7.2.2 Obtained Distortion Model Visualization

To visualize the distortion model of the left camera in the stereo rig, the radial distortion visualization graph has been plotted. As expected, the amount of radial disparity close to the edges of the image is more than the centre of it.

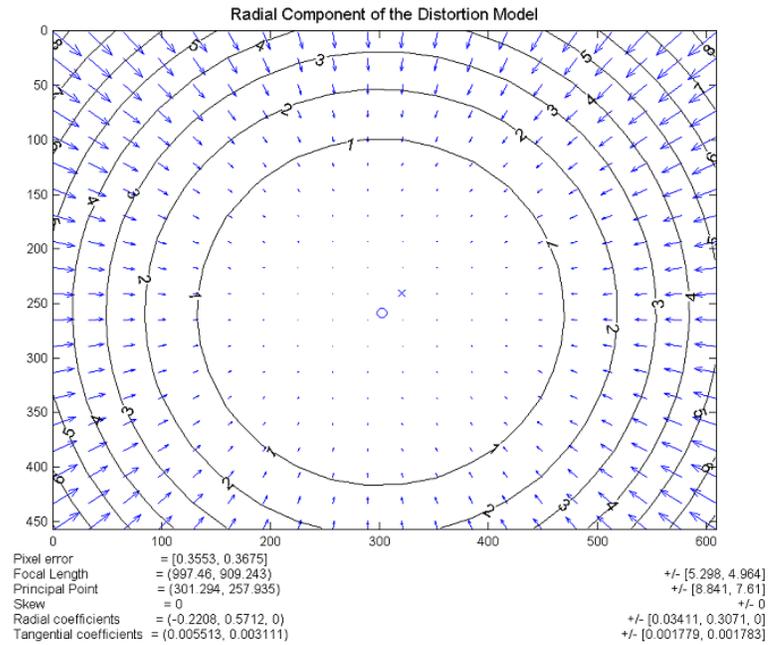


Figure 3-6. The radial distortion visualization

Now that the distortion coefficients (k_1 and k_2) have been calculated, the undistorted coordinates of the image coordinates can be calculated using (3-38) and (3-39).

Chapter 4. 3D Position Sensing Using Stereo Vision

4.1 Introduction

Chapter 3 discussed the proposed camera calibration techniques where two transformations were obtained using DLT and SDLT methods. As explained in Table 3-1 and Table 3-2, knowing the image and world coordinates of a set of calibration points leads to obtaining the camera models:

$$\{known : {}^l x, {}^r x \text{ and } {}^1 X\} \Rightarrow H_l \text{ and } H_r \text{ or } \{known : {}^l x, {}^r x \text{ and } {}^1 X\} \Rightarrow H$$

In this chapter the goal is to obtain the position of a 3D point using its perception in the image planes and the stereo camera models:

$$\{known : H_l, H_r, (x_l, y_l) \text{ and } (x_r, y_r)\} \Rightarrow {}^1 X \text{ or } \{known : H, (x_l, y_l) \text{ and } x_r\} \Rightarrow {}^1 X$$

Section 4.2 explains how a reliable perception of an object in the image planes can be obtained using the epipolar geometry principle. Section 4.3 explains how the acquired image coordinates and the camera models can be used in the 3D position sensing.

The proposed 3D position sensing process provides the estimated position of the object of interest in $\{q_1, \underline{C}_1\}$ attached to the robot torso. This estimation does take into account the robot base angle rotary motion. In order to transfer the estimated position to $\{q_0, \underline{C}_0\}$ while the robot base angle θ_1 changes, a homogeneous transformation should be applied to compensate the base angle and represent the estimated position in $\{q_0, \underline{C}_0\}$.

$${}^0 X = {}^0 T_1 {}^1 X \quad (4-1)$$

$$\text{Where } {}^0 T_1 = \begin{bmatrix} e^{\theta_1 k \times} e^{-\frac{\pi}{2} i \times} & & & \\ & 0 & & \\ & & 0^T & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} \cos \theta_1 & 0 & -\sin \theta_1 & 0 \\ \sin \theta_1 & 0 & \cos \theta_1 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4-2)$$

The obtained position of the object (${}^0 X$) will be labeled as the desired position in the proposed visual servoing system. The TDH robot end-effector will be guided to this destination.

4.2 Stereo Vision and Epipolar Geometry

4.2.1 Stereo Vision Principles

The epipolar geometry between two views is essentially the geometry of the intersection of the image planes with the pencil of planes having the baseline as axis (the baseline is the line joining the camera centers) [61].

The two possible geometries between the image planes will be reviewed. The first one deals with intersecting (non-parallel) planes and the second one with parallel image planes. The latter is the main focus of this chapter and all the 3D sensing processes consider this geometry.

- **Intersecting image planes**

We briefly review the essential epipolar geometry knowledge required in this thesis. More readings regarding the stereo and binocular vision are provided in [61].

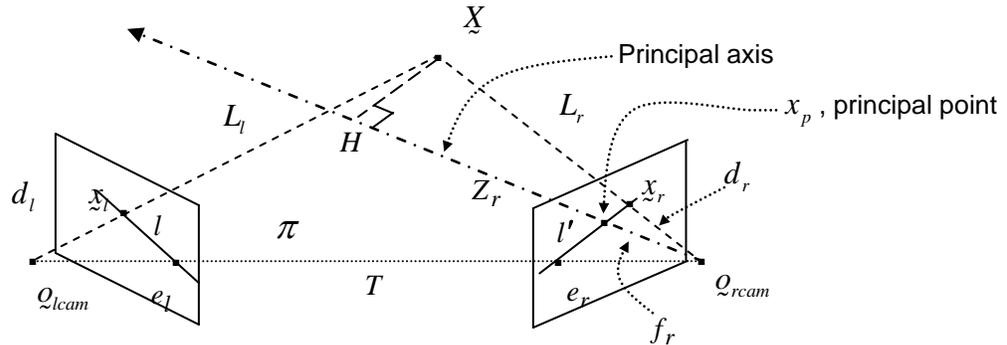


Figure 4-1. The epipolar geometry schematic

The geometry of a point in 3D space and its images in two CCD planes (image planes) has been depicted in Figure 4-1. The description of the components of this geometry is as below:

Q_{lcam}, Q_{rcam} : The camera centers

e_l, e_r : Epipoles. The intersection of the baseline (the line joining the camera centers) with the image planes

x_l, x_r : The images of the point X in left and right image planes

d_l : The distance between the left camera center (Q_{lcam}) and image point x_l .

d_r : The distance between the left camera center (o_{rcam}) and image point x_r .

L_l, L_r : The distance between the camera center and the point X

l, l' : Epipolar lines or pencils. The line between the epipole and the image point.

T : The stereo camera baseline

f_l, f_r : The focal length of the left and right cameras. The distance between the camera center and the image plane (the camera CCD).

Z_r : The depth of the point X in the right camera. The projection of the point X on the principal axis (The axis starts from the camera center and passes through the principal point which is the image plane center).

π : Epipolar plane, is defined as the plain formed by the point X in 3D space and the camera centers. The epipolar plane intersects with the image planes in two lines (the epipolar lines). It can be observed that the epipolar plane rotates about the baseline as the point X varies in 3D space. So, all the epipolar lines intersect at the epipole.

From the perspective projection:

$$\frac{f_r}{Z_r} = \frac{d_r}{L_r} \quad \frac{f_l}{Z_l} = \frac{d_l}{L_l} \quad (4-3)$$

The trajectory of point X on the epipolar plane lies on the epipolar lines that are obviously not parallel with each other. This fact constructs a very important phenomenon used to determine the corresponding image point of an object in 3D space when the camera planes are parallel with each other.

- **Parallel image planes**

We now discuss the case when the camera planes are parallel. In this thesis, the stereo camera system refers to the system with parallel camera planes.

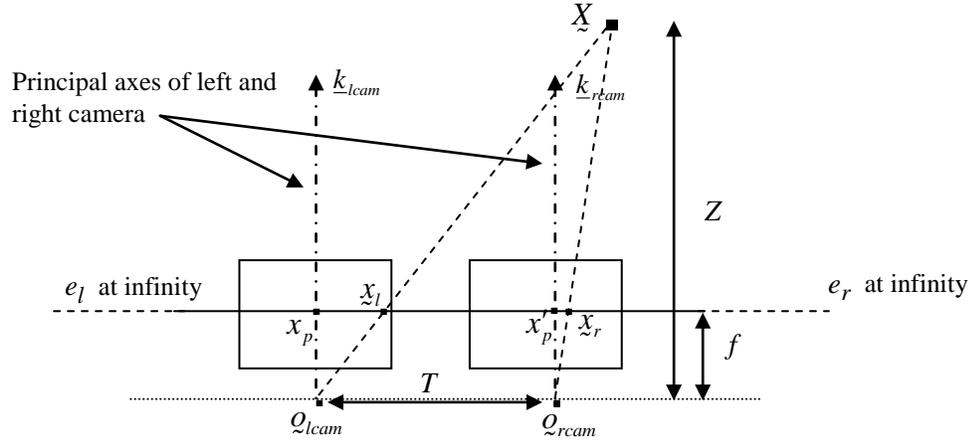


Figure 4-2. The stereo camera system with parallel image planes

$$\frac{T - |x_l - x_r|}{T} = \frac{Z - f}{Z} \Rightarrow Z = \frac{f}{d} T \quad (4-4)$$

$$d = |x_l - x_r| \quad (4-5)$$

Notes regarding Figure 4-2:

- 1- The epipoles e_l and e_r will be placed at infinity. This makes the epipolar line to be parallel to the baseline of the cameras.
- 2- The depth of the point X is inversely proportional to disparity d . This fact imposes that the farther the point in 3D space, the less disparity observed from its motion in the image planes.
- 3- The epipolar line l' is collinear with l .

Principle 4-1. Assuming a point in 3D space, the corresponding epipolar lines in the left and right images are collinear and must lie on a straight line parallel with the camera baseline if the following conditions are perfectly satisfied:

- The camera image planes are coplanar
- The image axes i_{lp} and i_{rp} are collinear

Corollary 4-1. When Principle 4-1 is held, assuming an image point x_l lies on l (in the left camera plane), an image point x_r can be obtained by continuing the epipolar line l towards the right image plane. Equivalently l' is along the l .

Proof: Consider the figure below:

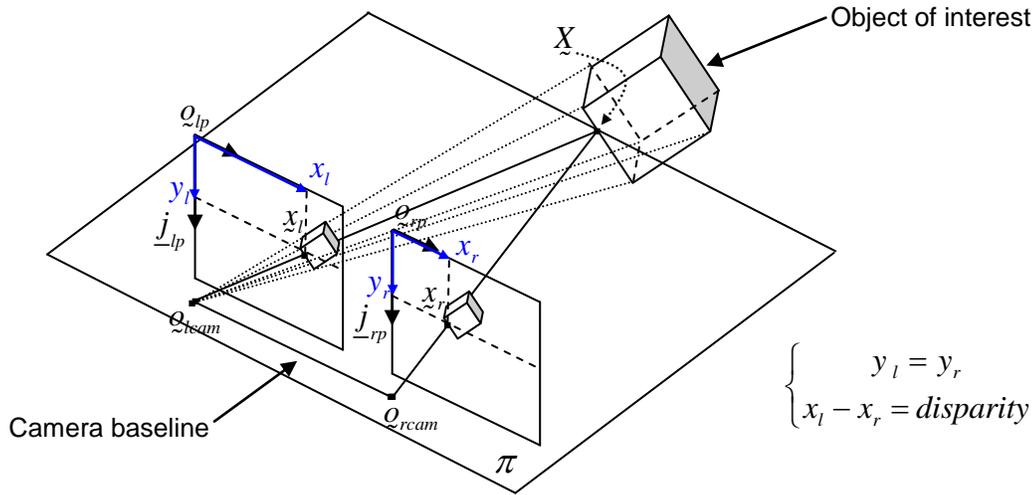


Figure 4-3. Stereo vision with parallel and coplanar image planes (The image origins have been shown on the image corners to avoid confusion)

Plane π contains points X, x_l, x_r, O_{lcam} and O_{rcam} . Consequently these points are coplanar.

If the camera planes are precisely parallel, then the epipolar lines must be collinear and have the same distance from the origin of the image planes (O_{lp} and O_{rp}) along \underline{j}_{lp} and \underline{j}_{rp} axes. This distance has been indicated as y_l and y_r respectively.

The relation below holds for every 3D point in the common FOV of the stereo camera system with parallel and coplanar image planes:

$$\left\{ \begin{array}{l} \text{When principle 4.1 is held :} \\ \forall X \in 3D \text{ space and in camera FOV, } \exists {}^l x = \begin{bmatrix} x_l \\ y_l \end{bmatrix}, {}^r x = \begin{bmatrix} x_r \\ y_r \end{bmatrix} \Big| y_l = y_r \end{array} \right\} \quad \text{Corollary 4-2}$$

The proof of the above corollary is provided in [61].

4.2.2 Facts Perturbing Parallel Stereo Vision Characteristics

Lens distortion: It is important to consider the lens distortion as a factor perturbing the facts in epipolar geometry with parallel image planes. Since the lens distortion is smaller in the central region of the image planes, points located in this area of the image are less subject to distortion. The detrimental effects of lens distortion show up especially while the visual servoing is in process. During the visual servoing process, the image coordinates of the object of interest may reside in the edges of the image planes due to changes in the robot base angle. The lens distortion model for both left and right lenses are obtained later in this chapter. The obtained model will be used to estimate the ideal (undistorted) coordinates of the image points.

Gradual mechanical misalignment: In order to satisfy the coplanar properties of the two cameras in a stereo rig, a rugged and solid mechanical structure should be manufactured to hold the cameras tightly in their place. Depending on where the cameras are installed, the environmental conditions may also affect the tightness of the mechanical fasteners and consequently the structure. Among those environmental conditions shock and vibration are the most important sources in losing the alignment of the cameras. Considering the items below can reduce the chance of misalignment on the adjusted stereo cameras:

1. Using a smaller number of fasteners and incorporating fixed mating materials (with no need for fasteners) in a stereo rig
2. Designing the stereo rig with few joints and adjustable components

4.3 3D Reconstruction and Position Estimation

4.3.1 Two Methods to Estimate the Object Position in 3D

As explained in the introduction section of this chapter, the goal is to find the 3D position of the object of interest using the known image coordinates and the camera models.

There are two methods in this section to obtain the 3D position of the target point in the FOV of cameras. In the first method, the coordinates of the desired point is obtained by solving an over-determined linear system of equations while the coordinates of the point in 3D space have been assumed to be the unknown components [61].

The second method uses Singular Value Decomposition (SVD) to obtain the desired coordinates of the point in the 3D space. This method has been introduced in [61].

The first and second methods can be employed for systems with camera calibration models obtained from the DLT and SDLT methods, respectively.

4.3.1.1 First method, matrix equations

Considering Equations (3-7) and (3-8) introduced in the DLT method, the matrix equations for the left and right cameras can be rearranged as follows:

$$\begin{bmatrix} x_l - \alpha_4 \\ y_l - \alpha_8 \end{bmatrix} = \begin{bmatrix} (\alpha_1 - \alpha_9) & (\alpha_2 - \alpha_{10}) & (\alpha_3 - \alpha_{11}) \\ (\alpha_5 - \alpha_9) & (\alpha_6 - \alpha_{10}) & (\alpha_7 - \alpha_{11}) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \quad (4-6)$$

$$\begin{bmatrix} x_r - \beta_4 \\ y_r - \beta_8 \end{bmatrix} = \begin{bmatrix} (\beta_1 - \beta_9) & (\beta_2 - \beta_{10}) & (\beta_3 - \beta_{11}) \\ (\beta_5 - \beta_9) & (\beta_6 - \beta_{10}) & (\beta_7 - \beta_{11}) \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} \quad (4-7)$$

An overdetermined linear system of equations can be solved to calculate the Cartesian coordinates of the target object given the camera model parameters $\{\alpha_i, \beta_i, i = 1, \dots, 11\}$ and the image coordinates (x_1, y_1, z_1) .

4.3.1.2 Second method, direct transformation

In Section 3.4 the introduced SDLT method was used to calculate the transformation matrix H to map the 3D coordinates of the target point to its image coordinates in the left and right cameras. Referring to (3-31) with the format:

$$\begin{bmatrix} w_l x_l \\ (w_l + w_r)y \\ w_l \\ w_r x_r \\ w_r \end{bmatrix} = \begin{bmatrix} \varphi_1 & \varphi_2 & \varphi_3 & \varphi_4 \\ \varphi_5 & \varphi_6 & \varphi_7 & \varphi_8 \\ \varphi_9 & \varphi_{10} & \varphi_{11} & 1 \\ \varphi_{12} & \varphi_{13} & \varphi_{14} & \varphi_{15} \\ \varphi_{16} & \varphi_{17} & \varphi_{18} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} = H^{-1} X$$

In the above equation, the known components are matrix H and the image coordinates.

The above equation can be formed in the system of homogenous equations as follows:

$$\begin{bmatrix} (\alpha_1 - \alpha_9 x_l) & (\alpha_2 - \alpha_{10} x_l) & (\alpha_3 - \alpha_{11} x_l) & (\alpha_4 - x_l) \\ (\alpha_5 - (\alpha_9 + \alpha_{16})y) & (\alpha_6 - (\alpha_{10} + \alpha_{17})y) & (\alpha_7 - (\alpha_{11} + \alpha_{18})y) & (\alpha_8 - 2y) \\ (\alpha_{12} - \alpha_{16} x_r) & (\alpha_{13} - \alpha_{17} x_r) & (\alpha_{14} - \alpha_{18} x_r) & (\alpha_{15} - x_r) \end{bmatrix} \begin{bmatrix} {}^1 X \\ 1 \end{bmatrix} = 0 \quad (4-8)$$

The components of the left hand side matrix in the above equation are known. We call this matrix Ω and re-write (4-8) as follows:

$$\Omega \begin{bmatrix} {}^1 X \\ 1 \end{bmatrix} = 0 \quad (4-9)$$

Singular Value Decomposition (SVD) now be employed to find the right null space vector of matrix Ω which is estimation for ${}^1 X$.

It is important to apply the transformation indicated in (4-2) to compensate for the TDH robot base angle θ_1 during the visual servoing process.

Important note: This method has been used in this thesis to estimate the 3D position of the target points.

4.4 Line and Plane Intersection in 3D

Finding the intersection of a line with a plane in 3D space has various applications in machine vision and visual servoing. As mentioned earlier, the TDH robotic manipulator is a planar robot in the sense that its gripper travels in a flat plane. Considering Figure 4-4 and $\{o_0, \underline{C}_0\}$ as the world coordinate system, the robot end-effector moves on a plane with \underline{k}_0 as its normal vector. The proposed visual servoing technique controls the robot end-effector to approach to the desired positions on this plane (π as shown in Figure 4-4). In fact the desired position on this plane is the intersection of the object and the robot planar workspace.

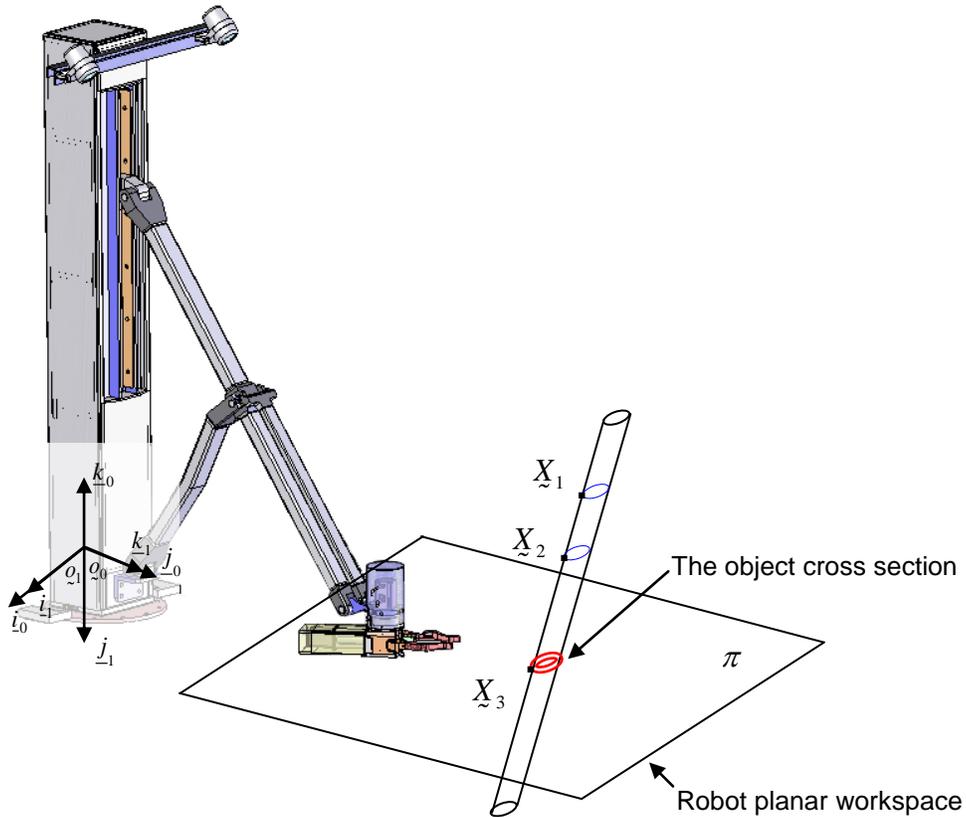


Figure 4-4. The cross section schematic of the pipe with gripper plane

Figure 4-4 shows the schematic of a pipe when it is placed in the robot workspace. In this case the object of interest has a cylindrical geometry and can be considered as a line in

3D space. Hence the problem becomes finding the equation of this line in 3D and then intersecting it with plane π . Algorithm 4-1 describes the process.

Algorithm 4-1. Cross section estimation of a line with a plane

Step1: Estimate the 3D position of two random points on the line. The two random points

X_1 and X_2 are represented by their coordinates in $\{Q_0, C_0\}$ as

${}^0X_1 = [(x_0)_1 \quad (y_0)_1 \quad (z_0)_1 \quad 1]^T$ and ${}^0X_2 = [(x_0)_2 \quad (y_0)_2 \quad (z_0)_2 \quad 1]^T$ respectively.

Step2: The equation of the straight line passing through these two points is as follows:

$$\frac{(x_0)_1 - (x_0)_2}{x - (x_0)_2} = \frac{(y_0)_1 - (y_0)_2}{y - (y_0)_2} = \frac{(z_0)_1 - (z_0)_2}{z - (z_0)_2}$$

Step3: Intersect the plane π with a certain height (h_0) with the equation of the line and obtain the intersection:

$$(x_0)_3 = x = \frac{z - (z_0)_2}{(z_0)_1 - (z_0)_2} ((x_0)_1 - (x_0)_2) + (x_0)_2$$

$$(y_0)_3 = y = \frac{z - (z_0)_2}{(z_0)_1 - (z_0)_2} ((y_0)_1 - (y_0)_2) + (y_0)_2$$

$$z = h_0 \Rightarrow {}^0X_3 = [(x_0)_3 \quad (y_0)_3 \quad h_0 \quad 1]^T$$

If $(z_0)_2 = (z_0)_1$ then:

- 1- if $h_0 = (z_0)_2 = (z_0)_1$ then the line is inside the plane and at least one solution and at most infinite solutions exists
- 2- if $h_0 \neq ((z_0)_2 = (z_0)_1)$ then no solution exists

Chapter 5. Experimental Results

5.1 Camera Calibration Experimental Results

In Chapter 3 two methods were discussed to relate the 3D coordinates of the calibration points to their image coordinates. The first method assumed individual camera models for the left and right cameras that are not necessarily placed in a stereo camera geometry with parallel and coplanar image planes. This method estimated the two camera models using the classic DLT method. The second method introduced a reduced format of the DLT method by assuming equal vertical coordinates in the observed stereo images. The single transformation matrix obtained can be used to relate (transform) the world coordinates of the calibration points to their image coordinates in both left and right image planes. So, the methods can be categorized as follows:

- 1- Obtaining individual camera models using the Classic DLT method (Section 3.2).
- 2- A single transformation using the SDLT method (Section 3.4).

After considering either of the above methods, the data collection process should be started to provide the coordinates of the calibration points in the 3D robot base frame and the 2D image planes. The iterative least-squares estimation will then be used to calculate the transformation matrix by estimating the required parameters. The two employed data collection methods (using a Cartesian robot and the *in situ* calibration tool) were discussed in Sections 3.6.1 and 3.6.2. In this section the experimental results have been categorized as follows.

- Data collection using the Cartesian robot
 - Validate the obtained camera models using the DLT classic method
 - Validate the obtained H matrix using the SDLT method
- Data collection using the *in situ* calibration tool
 - Validate the obtained camera models using the classic DLT method
 - Validate the obtained H matrix using the SDLT method

5.1.1 Data Collection using the Cartesian Robot

Figure 5-1 shows the trajectory of the LED (calibration object) coordinates in the calibration cube represented in $\{q_0, \underline{C}_0\}$. For the purpose of the calibration, 7500 data points have been collected. The first 4500 data points were employed for the camera calibration process and the remaining 3000 points were used to validate the obtained camera model(s) (cross validation).

Notes regarding the LED trajectory in the TDH robot base coordinate system ($\{q_0, \underline{C}_0\}$):

- 1- Considering the components shown in Figure 3-1 and what the plotted data represents, the LED travels from 149cm to 201cm along \underline{j}_0 axis, from -49cm to 49cm along \underline{i}_0 and from 10cm to 40cm along \underline{k}_0 . Hence, the calibration cube dimension is 98x52x30cm. This cube is only employed in this section and in the rest of this chapter the volume formed by the *in situ* tool is considered.
- 2- The robot piecewise motion in the calibration cube is not the same on all axes. The robot travels 52cm continuously along \underline{j}_0 , 5cm along \underline{i}_0 and so on. When \underline{i}_0 is traveled for 92cm then the calibration object is raised by 5cm along \underline{k}_0 .

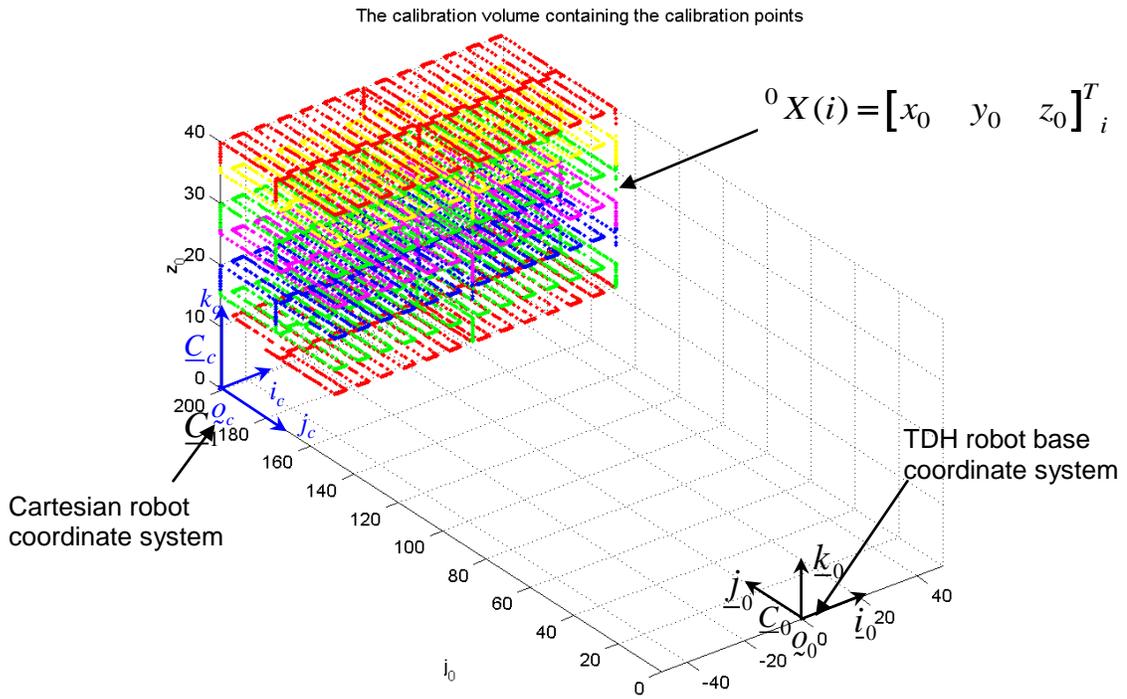


Figure 5-1. The calibration cube generated by the Cartesian robot

The plotted 3D trajectory are x_0 , y_0 and z_0 components in (3-24). All axes values are in cm.

Figure 5-2 shows the LED trajectory in the image planes in the left and right cameras.

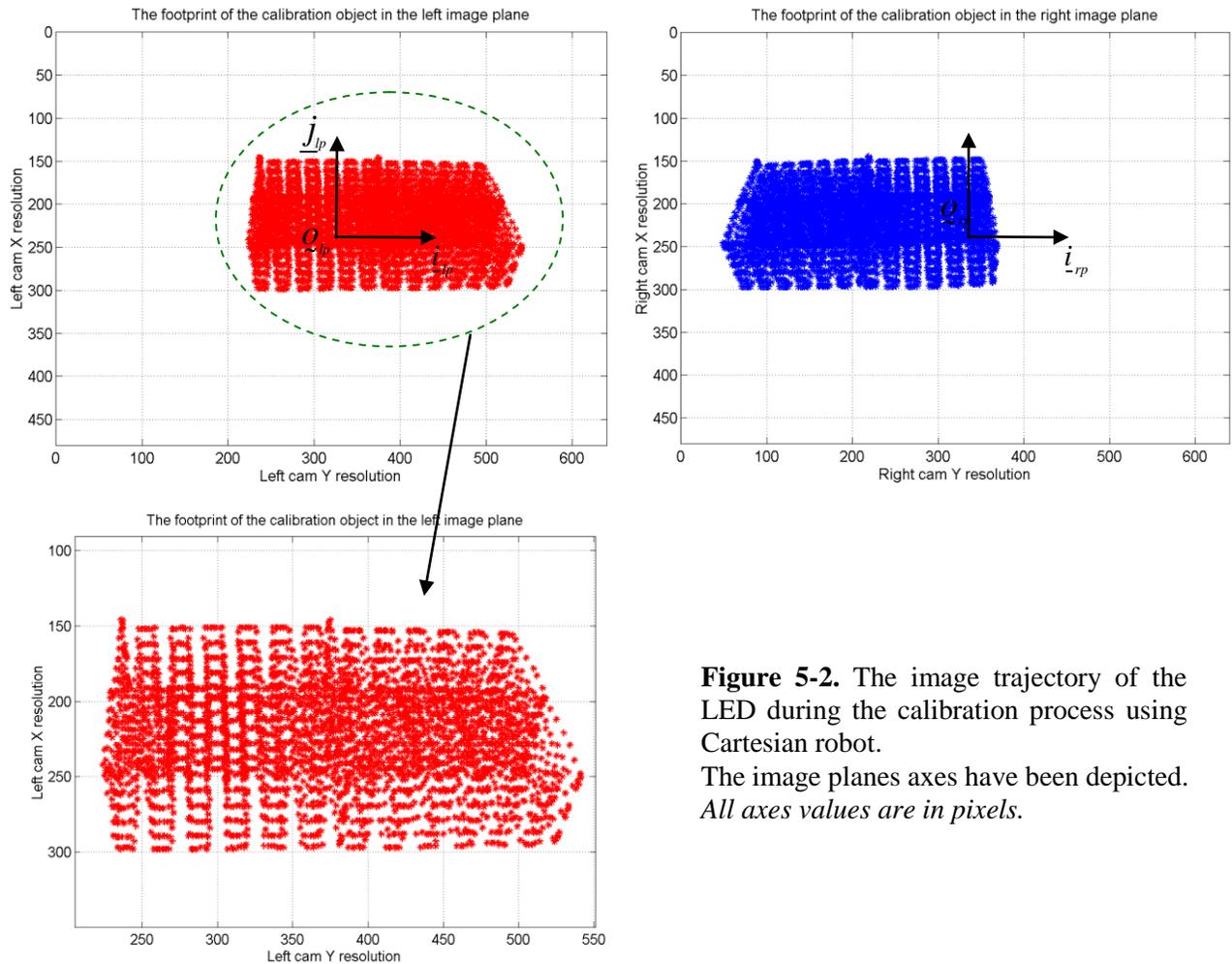


Figure 5-2. The image trajectory of the LED during the calibration process using Cartesian robot. The image planes axes have been depicted. All axes values are in pixels.

Notes regarding the trajectory of the LED in the image planes using the Cartesian robot:

- 3- As plotted above, the camera CCD resolution is 640x480 pixels.
- 4- The stereo camera rig has been designed to hold the cameras firmly with coplanar and horizontally shifted CCD planes. Hence all the points in the image planes share the same vertical displacement along \underline{j}_{lp} and \underline{j}_{rp} axes. As can be seen in the plots, the trajectories of the points are limited between 150 and 300 pixels.
- 5- As shown in the figure, the long camera baseline (53cm) has provided a significant disparity. The disparity depending on the LED distance to the cameras is variable between 161pixels and 218pixels.

5.1.1.1 Validate Camera Model using DLT and SDLT Methods

In this section the camera model(s) obtained using the DLT and SDLT methods are compared to investigate the obtained camera model(s).

The general conditions below as stated in the second row of Table 3-1 and Table 3-2 should be considered to verify the obtained camera model(s):

$$\{known: H_l, H_r, {}^l x, {}^r x\} \Rightarrow {}^0 \hat{X}$$

$$\{known: H, {}^l x, {}^r x\} \Rightarrow {}^0 \hat{X}$$

The estimated 3D Cartesian coordinates will be compared with their actual values. The plots below show the error between the actual and estimated 3D Cartesian coordinates while the above conditions were considered. The DLT and SDLT methods were employed to obtain the camera model(s) using the collected 3000 points reserved for camera model validation.

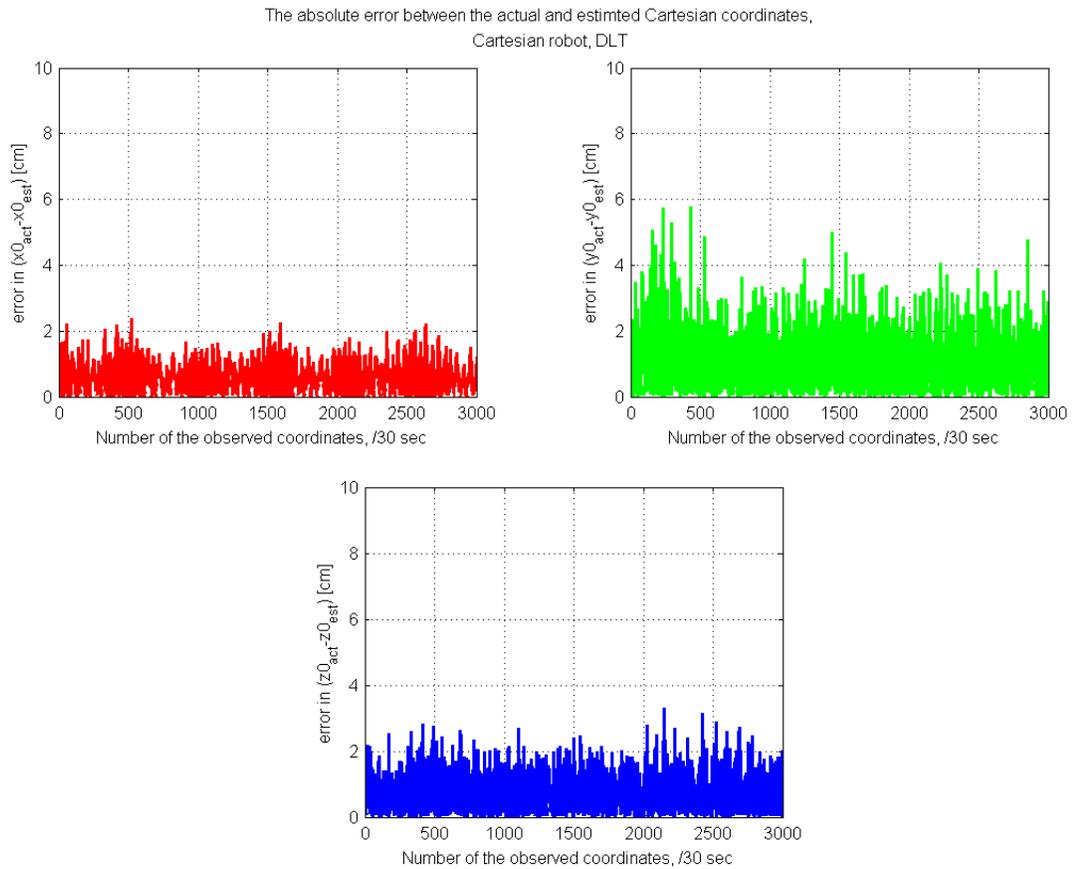


Figure 5-3. The absolute error between est and act 3D Cartesian coordinates
 The method of data collection: using the Cartesian robot
 The camera model is obtained: using DLT method

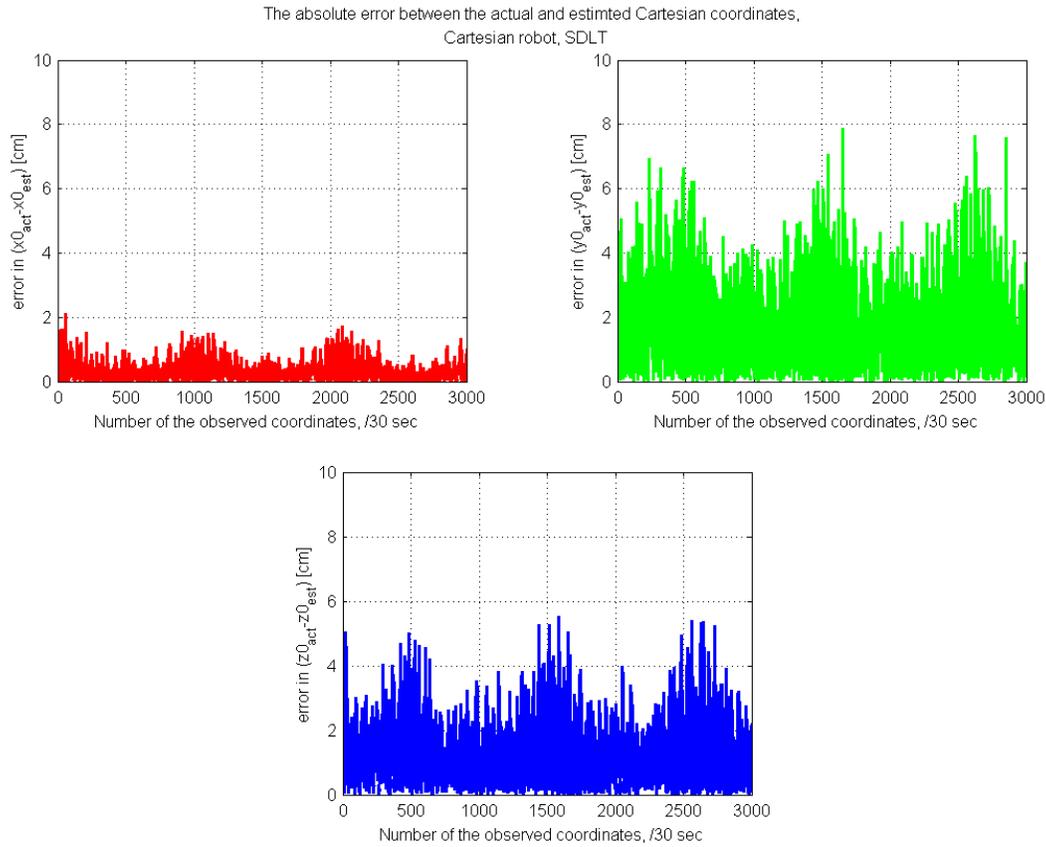


Figure 5-4. The absolute error between est and act 3D Cartesian coordinates
The method of data collection: using the Cartesian robot
The camera model is obtained: using SDLT

Table 5-1. Comparing the Mean Absolute Error (MAE) and Maximum Error (MAXE) between the actual and estimated Cartesian coordinates obtained using the classic and direct method

| 3D points | MAE-Cartesian DLT (cm) | MAXE-Cartesian DLT (cm) | MAE-Cartesian SDLT (cm) | MAXE-Cartesian SDLT (cm) |
|------------------|------------------------|-------------------------|-------------------------|--------------------------|
| x_0, \hat{x}_0 | 0.5296 | 2.3698 | 0.3199 | 2.1045 |
| y_0, \hat{y}_0 | 1.0258 | 5.7620 | 1.6410 | 7.8731 |
| z_0, \hat{z}_0 | 0.7337 | 3.2906 | 1.2059 | 5.5190 |

The results show that the mean absolute errors between the actual and estimated 3D Cartesian coordinates are not the same. The calculated mean Euclidean distance between the actual and estimated 3D coordinates in the DLT method is 0.99cm whereas in the SLDT method is 1.26cm. This shows that the DLT method provides a slightly more accurate estimation of the camera model(s).

5.1.2 Data Collection using the *In situ* Calibration Technique

The trajectory of the calibration objects (bright LEDs) have been depicted in Figure 5-5. During the *in situ* camera calibration, 7500 points were collected. Unlike the Cartesian calibration robot, the coordinates of the 3D calibration points are obtained using the TDH robot encoders and a forward kinematics process. The first 4500 data points have been used for camera calibration and the remaining 3000 data points were used to validate the obtained camera model(s). In this thesis it has been assumed that the encoders provide the closest joint angles to their actual values. All the linkage measurements have been assumed to be accurate. Hence the 3D coordinates of the calibration points (using a forward kinematics process) is assumed to be as close as possible to their actual coordinates.

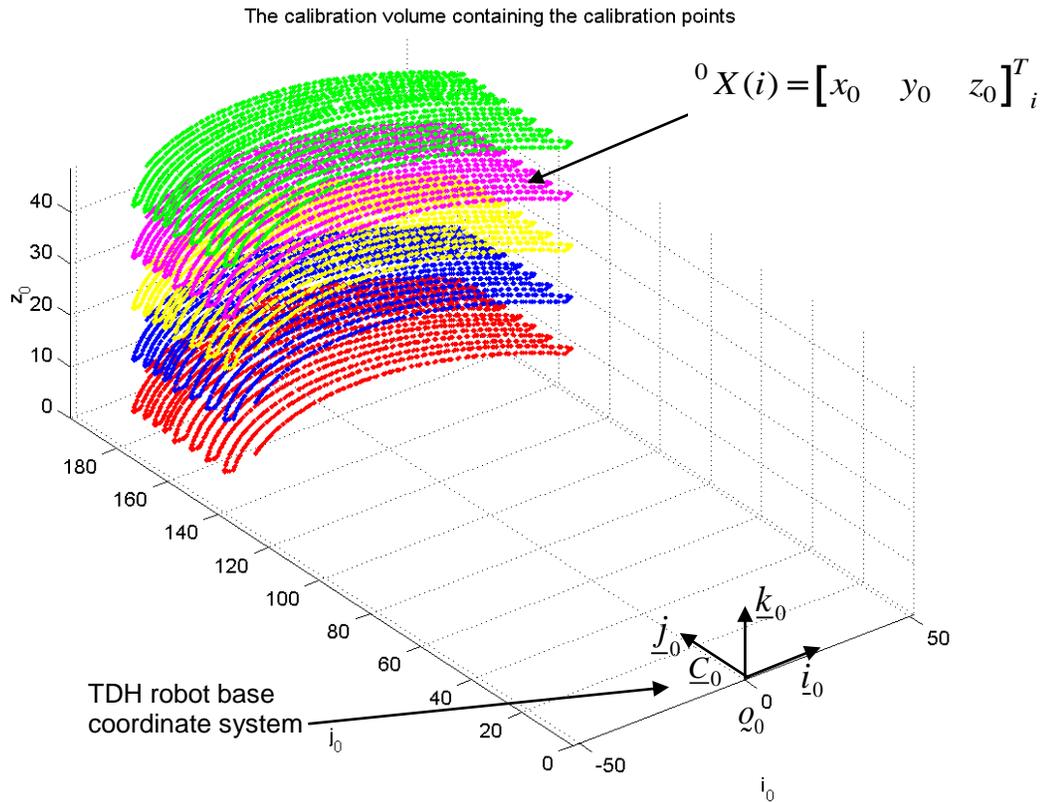


Figure 5-5. The calibration volume generated by *in situ* calibration. The plotted 3D trajectory are x_0 , y_0 and z_0 components in (3-24). All axes values are in cm.

5.1.2.1 Validate Camera Model using DLT and SDLT Methods

The obtained camera model(s) from the *in situ* data collection technique is validated and compared when the DLT and SDLT methods were used. The plotted information shows the error between the actual and estimated 3D Cartesian coordinates for the 3000 collected points reserved for camera model validation.

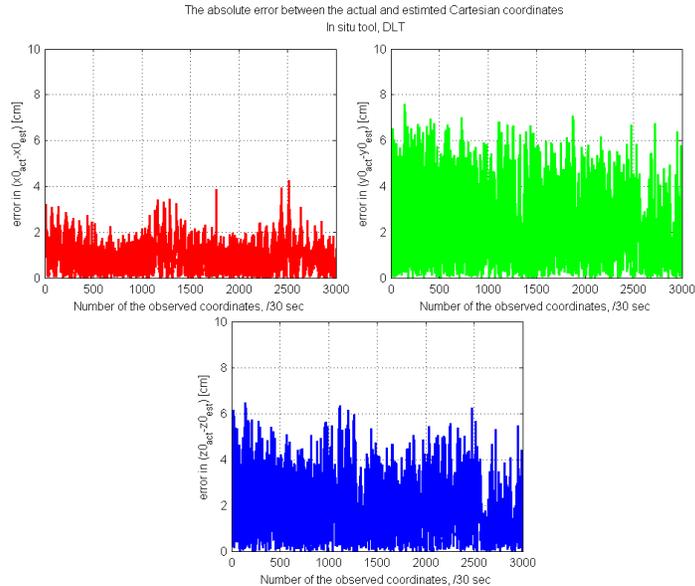


Figure 5-6. The absolute error between est and act 3D Cartesian coordinates
The method of data collection: using the *in situ* calibration, using the DLT method

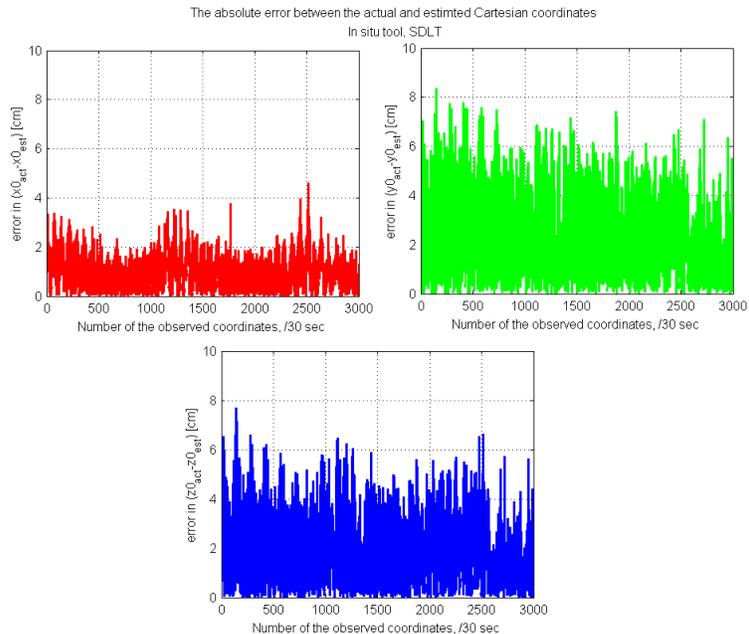


Figure 5-7. The absolute error between est and act 3D Cartesian coordinates
The method of data collection: using the *in situ* calibration, using the SDLT method

Table 5-2. Comparing the Mean Absolute Error (MAE) and Maximum Error (MAXE) between the actual and estimated Cartesian coordinates obtained using the classic and direct method

| 3D points | MAE-Cartesian DLT (cm) | MAXE-Cartesian DLT (cm) | MAE-Cartesian SDLT (cm) | MAXE-Cartesian SDLT (cm) |
|------------------|-----------------------------------|------------------------------------|------------------------------------|-------------------------------------|
| x_0, \hat{x}_0 | 0.8589 | 5.2523 | 0.9005 | 5.8542 |
| y_0, \hat{y}_0 | 2.0240 | 7.7411 | 2.0646 | 8.3629 |
| z_0, \hat{z}_0 | 1.7215 | 8.2598 | 1.8228 | 9.6080 |

The results show that the mean absolute errors between the actual and estimated 3D Cartesian coordinates are not the same. The calculated mean Euclidean distance between the actual and estimated 3D coordinates in the DLT method is 1.0343cm whereas in the SLDT method is 1.0445cm. This shows that the DLT method provides slightly a more accurate estimation of the camera model.

5.2 Experimental Results for 3D Position Estimation

In this section the camera models are constructed using the collected data in the *in situ* data collection technique. The last 3D reconstruction algorithm explained in Section 4.3 is employed to estimate the 3D position of the target object. The accuracy of this reconstruction is verified in two tables when the DLT and SDLT methods are employed.

The TDH robot has been assigned to move a LED within the robot workspace. The forward kinematics process using the robot encoders provides the position of the LED installed at the end-effector of the robot. The LED is installed with 8.5cm height from the robot base origin Q_0 . An image processing algorithm is able to track the LED in the image planes. The results will be the image coordinates in the left and right images (${}^l x, {}^r x$). Hence 2nd condition in Table 3-2 has been satisfied:

$$\{known : H, x_l, x_r, y_l = y_r = y, \Omega\} \Rightarrow {}^0 \hat{X}$$

The actual positions of the LED (${}^0 X_{act} = [x_0 \ y_0 \ z_0 \ 1]^T_{act}$) will be compared with the estimated positions (${}^0 \hat{X} = [x_0 \ y_0 \ z_0 \ 1]^T_{est}$).

Important Note: In the performed experiments it has been assumed that the real 3D position of the LED, obtained from the robot encoders, is as close as possible to its actual position in $\{Q_0, C_0\}$.

In order to visualize the accuracy of the 3D position estimation process, color accuracy maps have been prepared. These color maps are useful to verify the accuracy of the 3D estimation in different locations within the area of the LED trajectory. The experimental results have been organized as follows considering both DLT and SDLT camera models:

- 3D positioning without radial lens distortion correction
 - 3D positioning for the area larger than calibration area
 - 3D positioning within the calibration area
- 3D positioning with radial lens distortion correction
 - 3D positioning for area the larger than calibration area
 - 3D positioning within the calibration area

- 3D positioning for area smaller than the calibration area

Figure 5-8 shows the areas that the experimental results were performed inside them:

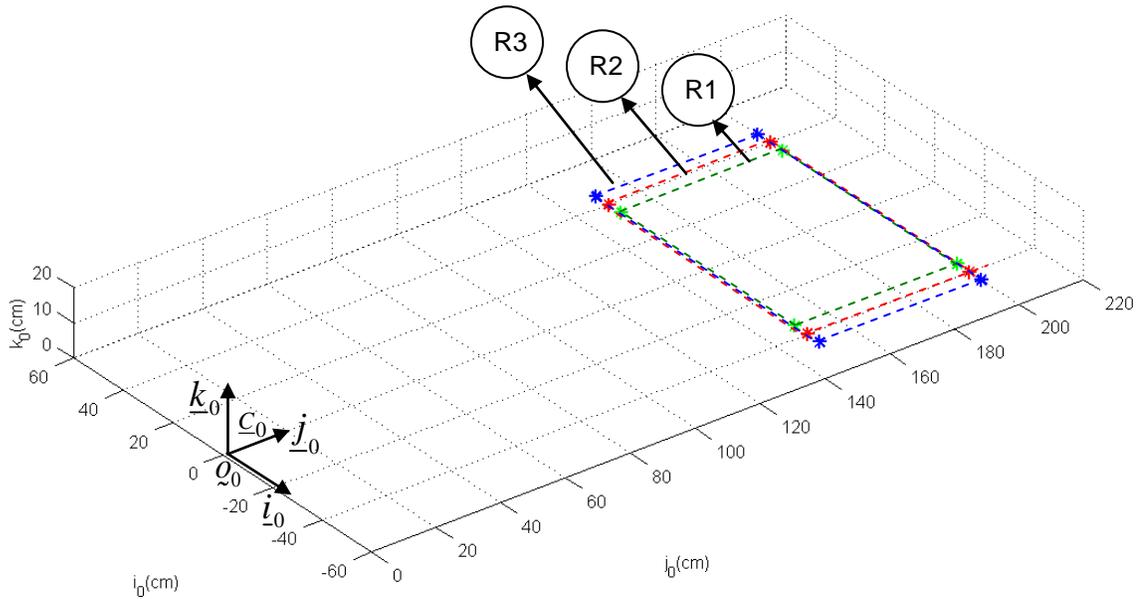


Figure 5-8. The 3D positioning validation areas

From now, Accuracy Measurement Area (AMA) is referred to the area that the accuracy measurement is performed within this area. The descriptions and dimension of the areas are as follow:

R1: The area smaller than the calibration area. The dimension of this area is 70x50cm which is from -35cm to 35cm along i_0 and 150cm to 200cm along j_0 .

R2: The stereo camera calibration is performed within this area. This area is in fact the first layer of the LED trajectory shown in Figure 5-1. The LED trajectory during the calibration is between -40cm to 40cm along i_0 and 150cm to 200cm along j_0 on different heights. Hence the dimension for area R2 is 80x50cm.

R3: The area larger than the calibration area. The dimension of this area is 90x50cm which is from -45cm to 45cm along i_0 and 150cm to 200cm along j_0 .

The Mean Absolute Error (MAE) of the actual and estimated 3D positions of the LED trajectory is calculated. The MAE is calculated separately for the Cartesian coordinates (x_0 , y_0 and z_0). The MAE is also calculated for the Euclidean distance of the Cartesian coordinates.

5.2.1 3D Positioning without Distortion Correction

5.2.1.1 3D Positioning for Area Larger than Calibration Area

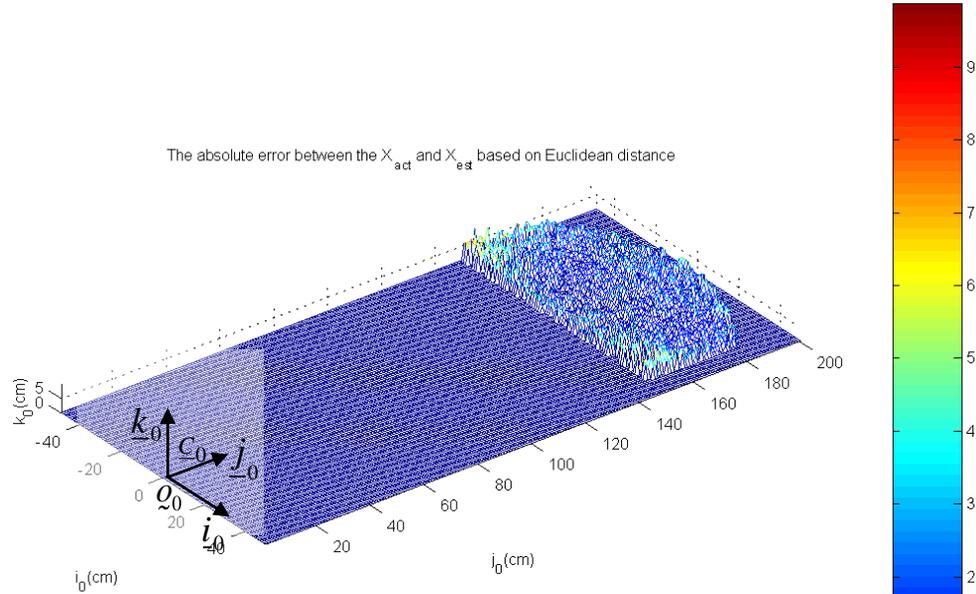


Figure 5-9. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R3 without lens distortion correction has been selected.

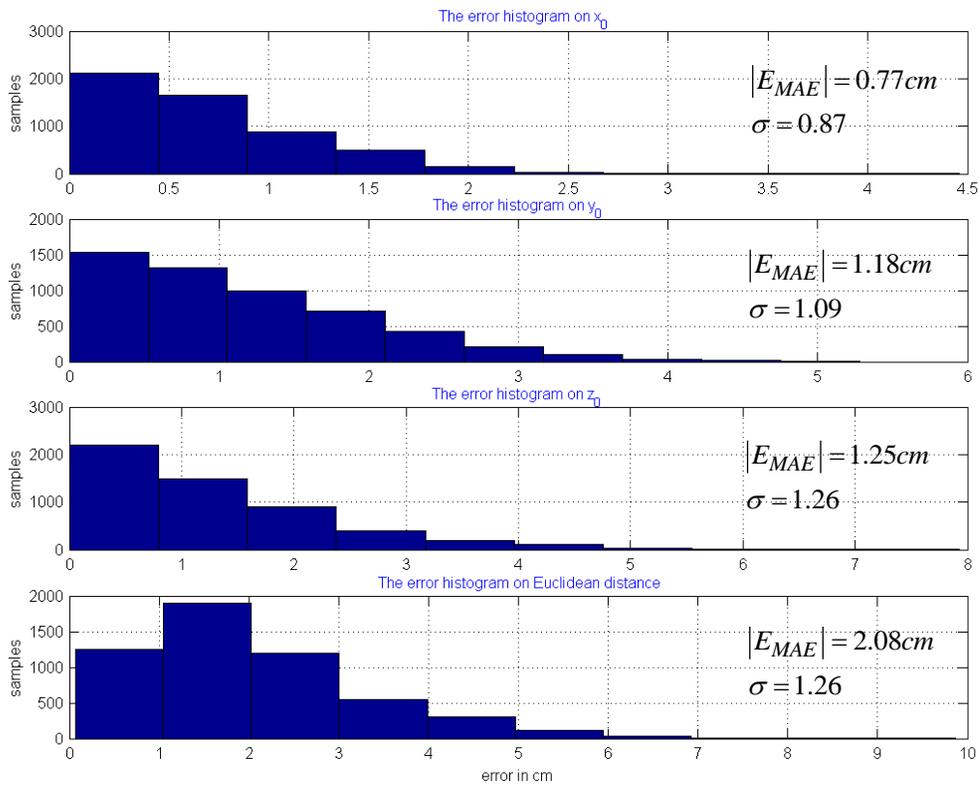


Figure 5-10. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions.

5.2.1.2 3D Positioning within Calibration Area

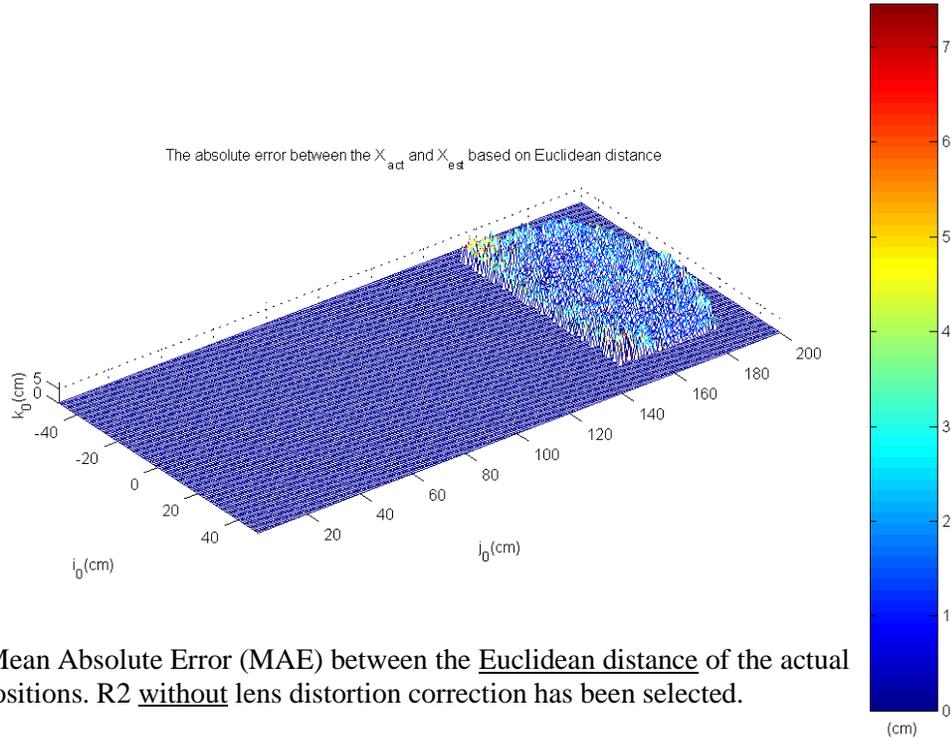


Figure 5-11. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R2 without lens distortion correction has been selected.

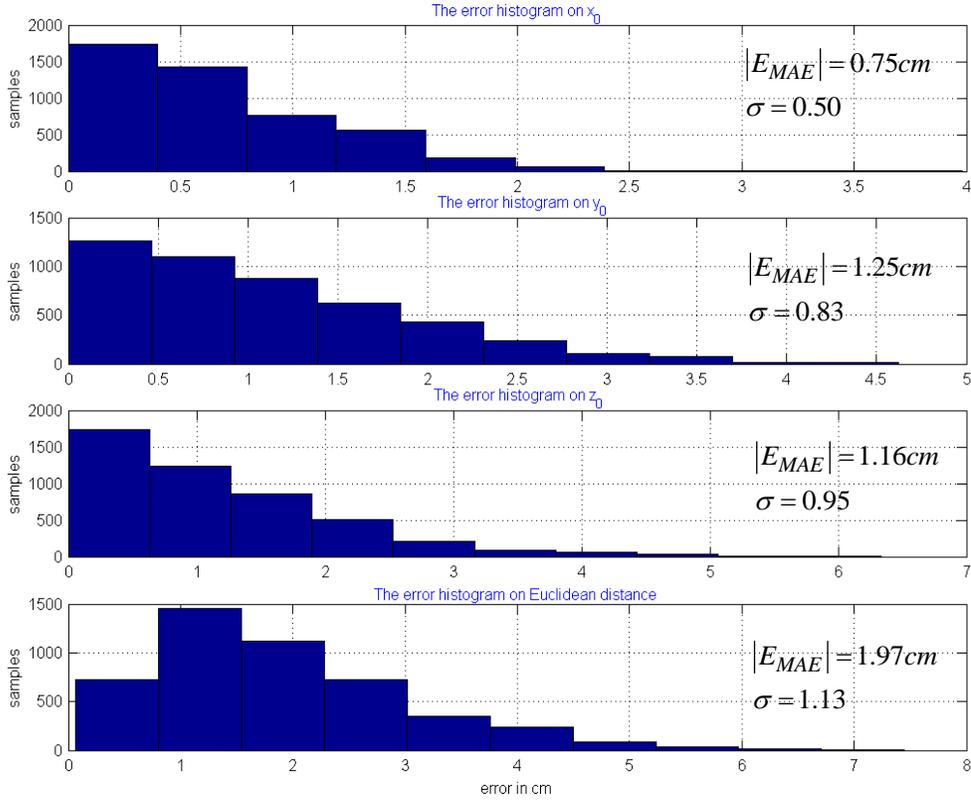


Figure 5-12. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions.

5.2.2 3D positioning with Distortion Correction

5.2.2.1 3D Positioning for Area Larger than Calibration Area

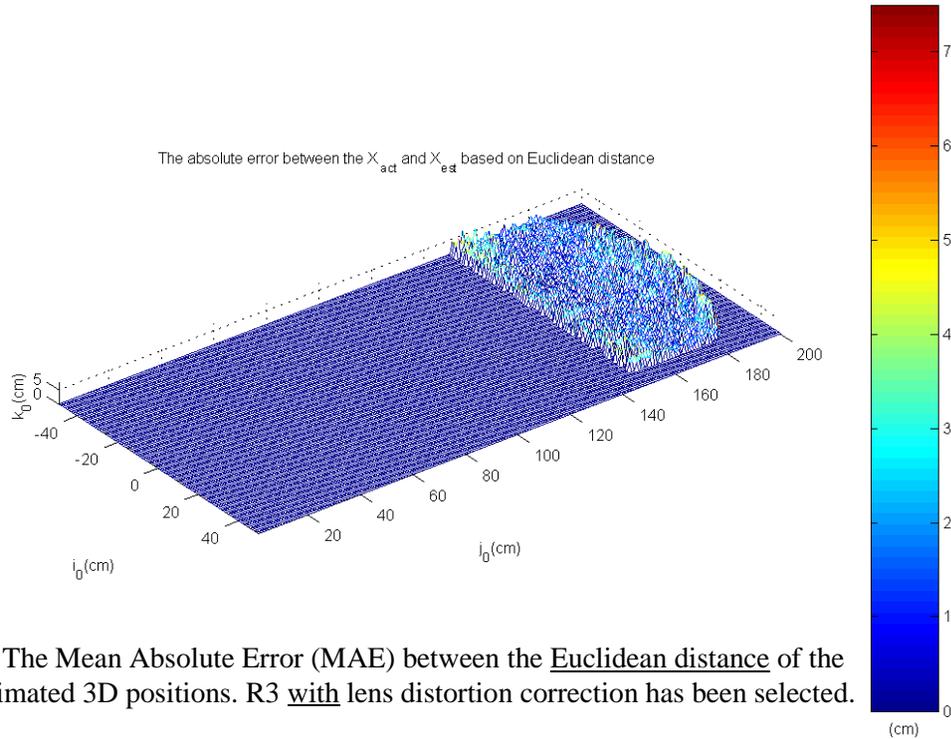


Figure 5-13. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R3 with lens distortion correction has been selected.

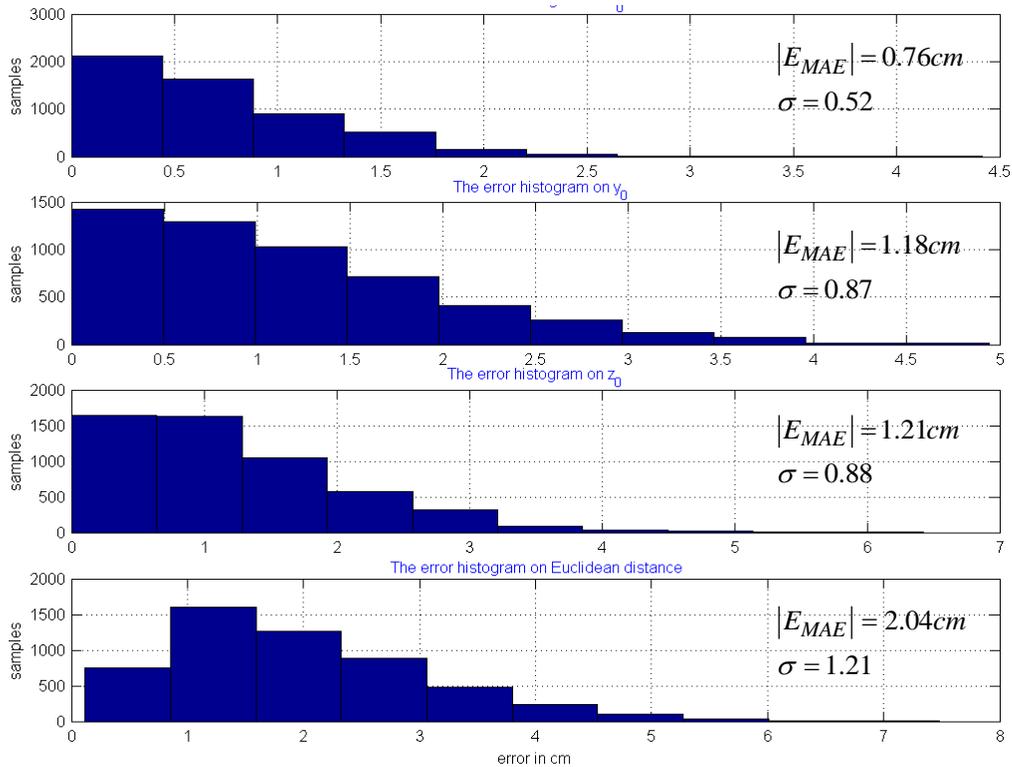


Figure 5-14. The histogram of the MAE between the Cartesian coordinates (x_0 , y_0 , z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions.

5.2.2.2 3D Positioning within Calibration Area

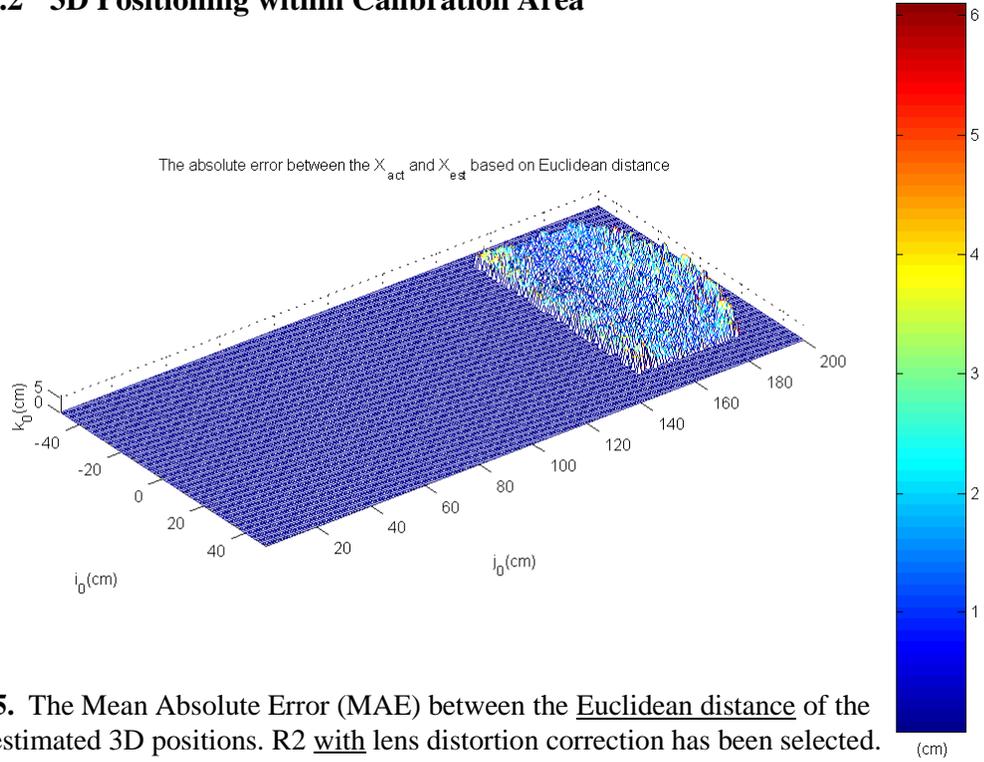


Figure 5-15. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R2 with lens distortion correction has been selected.

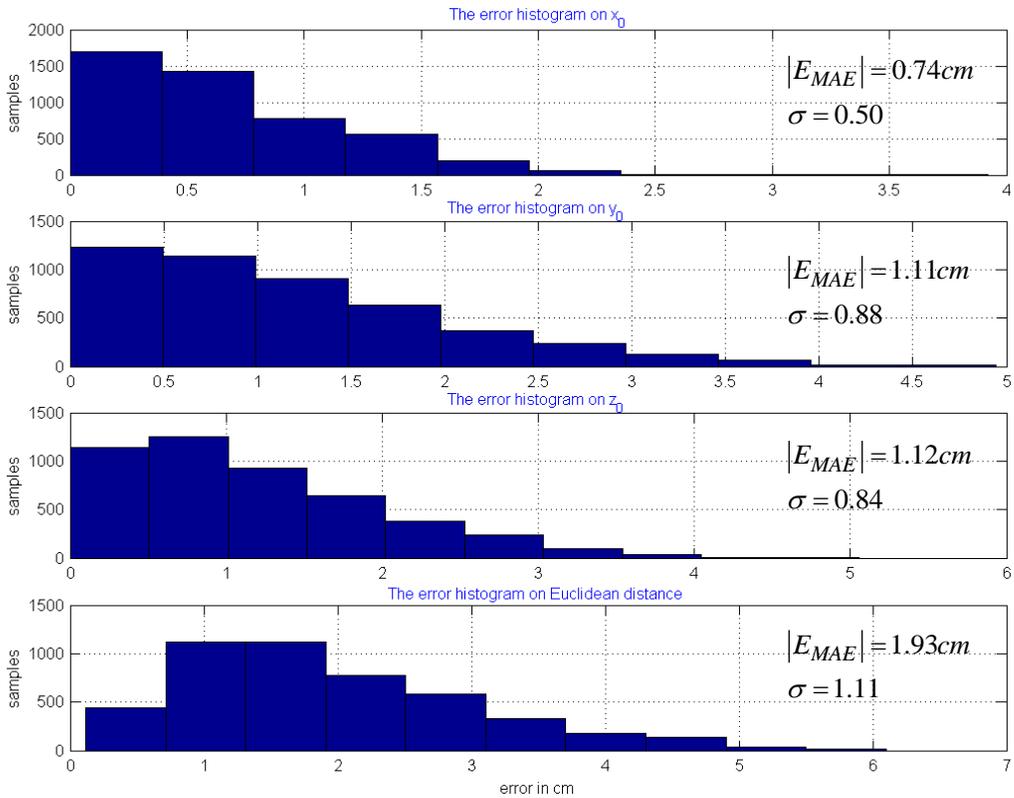


Figure 5-16. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions.

5.2.2.3 3D Positioning for Area Smaller than Calibration Area

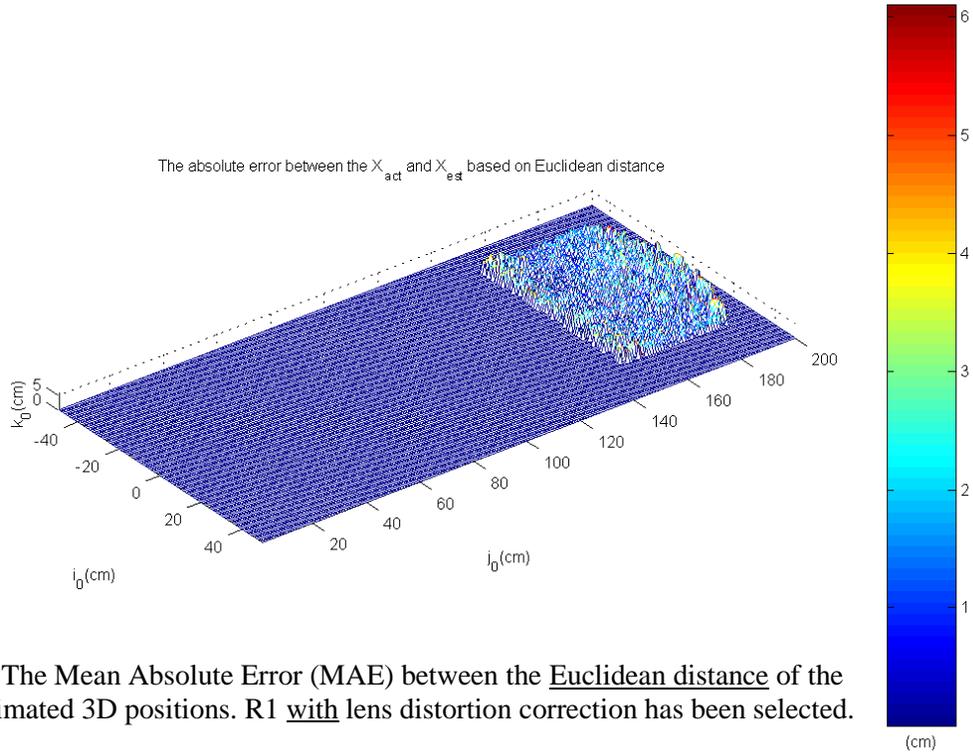


Figure 5-17. The Mean Absolute Error (MAE) between the Euclidean distance of the actual and estimated 3D positions. R1 with lens distortion correction has been selected.

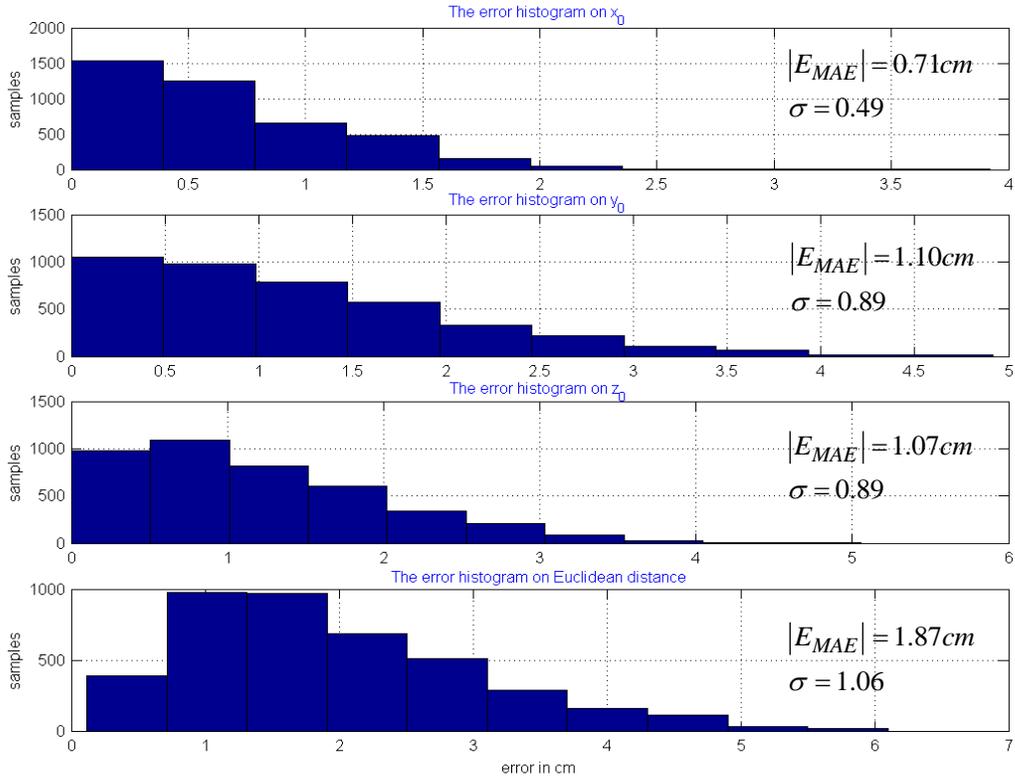


Figure 5-18. The histogram of the MAE between the Cartesian coordinates (x_0, y_0, z_0) of the actual and estimated 3D positions (first three plots). The MAE between the Euclidean distances of the actual and estimated 3D positions.

5.2.3 Comparing the Accuracy Measurement Results Using DLT Method

Table 5-3 shows the accuracy of the 3D positioning system for the mentioned areas in the robot workspace. The first three rows show the Mean Absolute Error (MAE) between the actual and estimated 3D positions on the Cartesian coordinates. The fourth row shows the MAE between the actual and estimated 3D positions in Euclidean distance. The last row indicates the standard deviation of the Euclidean distances.

Table 5-3. The accuracy measurement comparison in three different AMA[†] using DLT.

| <u>Unit = cm</u> | Area of R3 90x50cm NLDC | Area of R2 80x50cm NLDC | Area of R3 90x50cm WLDC | Area of R2 80x50cm WLDC | Area of R1 70x50cm WLDC |
|----------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| MAE (x₀) | 0.8517 | 0.7982 | 0.8554 | 0.8009 | 0.7602 |
| MAE (y₀) | 1.0929 | 1.0485 | 1.0612 | 1.0521 | 1.0488 |
| MAE (z₀) | 1.2362 | 1.1102 | 1.1408 | 1.0682 | 1.0420 |
| MAE (ED) | 2.0485 | 1.9100 | 1.9707 | 1.8842 | 1.8397 |
| STD (ED) | 1.3217 | 1.1791 | 1.2113 | 1.0855 | 1.0826 |

NLDC = No Lens Distortion Correction, WLDC = With Lens Distortion Correction

† AMA = Accuracy Measurement Area, ED = Euclidean Distance

In another attempt the worst case error between the actual and estimated 3D positions in different AMA has been provided in table below:

Table 5-4. The worst case error between the actual and estimated 3D positions using DLT

| <u>Unit = cm</u> | Area of R3 90x50cm NLDC | Area of R2 80x50cm NLDC | Area of R3 90x50cm WLDC | Area of R2 80x50cm WLDC | Area of R1 70x50cm WLDC |
|-----------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|-------------------------------|
| MAXE (x₀) | 4.9006 | 4.9006 | 4.9524 | 4.9524 | 3.9415 |
| MAXE (y₀) | 6.5344 | 6.5344 | 7.3571 | 7.3571 | 5.8666 |
| MAXE (z₀) | 8.4882 | 6.9241 | 7.7484 | 6.2757 | 6.1975 |
| MAXE (ED) | 10.3217 | 8.1945 | 9.2595 | 8.1456 | 6.9405 |

MAXE = Maximum Error

Important note: It can be confirmed from the provided histograms that the worst case error provided in Table 5-4 has the lowest frequency magnitude in the distribution.

5.2.4 Comparing the Accuracy Measurement Results using SDLT Method

In this section the accuracy measurement has been performed when the 3D positioning algorithm uses the camera model obtained from SDLT process. Same tables as represented in the previous section are provided here.

Table 5-5. The accuracy measurement comparison in three different AMA using SDLT.

| Unit = cm | Area of R3 90x50cm NLDC | Area of R2 80x50cm NLDC | Area of R3 90x50cm WLDC | Area of R2 80x50cm WLDC | Area of R1 70x50cm WLDC |
|-------------------------------|--|--|--|--|--|
| MAE (x_0) | 0.7707 | 0.7553 | 0.7618 | 0.7454 | 0.7180 |
| MAE (y_0) | 1.1884 | 1.2575 | 1.1856 | 1.1194 | 1.1028 |
| MAE (z_0) | 1.2591 | 1.1675 | 1.2118 | 1.1269 | 1.0718 |
| MAE (ED) | 2.0825 | 1.9778 | 2.0404 | 1.9381 | 1.8750 |
| STD (ED) | 1.2632 | 1.1304 | 1.2185 | 1.1155 | 1.0676 |

Table 5-6. The worst case error between the actual and estimated 3D positions using SDLT

| Unit = cm | Area of R3 90x50cm NLDC | Area of R2 80x50cm NLDC | Area of R3 90x50cm WLDC | Area of R2 80x50cm WLDC | Area of R1 70x50cm WLDC |
|--------------------------------|--|--|--|--|--|
| MAXE (x_0) | 4.6939 | 4.1995 | 4.6309 | 4.2001 | 4.2001 |
| MAXE (y_0) | 5.5089 | 4.8104 | 5.4740 | 4.8405 | 4.8405 |
| MAXE (z_0) | 7.5376 | 7.0118 | 7.5885 | 6.4095 | 5.9955 |
| MAXE (ED) | 9.8757 | 8.3349 | 9.9154 | 7.6061 | 6.8671 |

It can be concluded, for both DLT and SDLT methods that the lens distortion correction slightly improves the accuracy of the 3D positioning as the average error is reduced in the columns of the tables that the lens distortion correction was in effect. E.g., the average error for the area R2 has been reduced by 1.35% and 2% when DLT and SLDT methods were used, respectively. In Section 5.2.5 the minimal effects of the lens correction distortion is discussed. Limiting the accuracy measurement area to the area that the camera calibration was performed within that area (R2), less average error is introduced than the area larger than the calibration area (R3). This is due to the fact that the system is dealing with similar data set that was used during the camera calibration process.

5.2.5 Effects of Lens Distortion Correction

The first two columns of Table 5-3 and Table 5-4 are the results of the accuracy measurement experiment when the radial lens distortion is not corrected. The third to fifth columns show the same type of experiment when the lens distortion correction was employed. One can simply verify that the results are close to each other. Presumably, lens distortion correction should improve the system accuracy and it does it slightly in this case. This correction could be crucial especially when the camera calibration deals with the image coordinates in the nonlinear or distorted areas of the left and right images. In order to investigate the reason behind the similarity of the accuracy measurements with/without lens distortion correction, the plot below have been prepared.

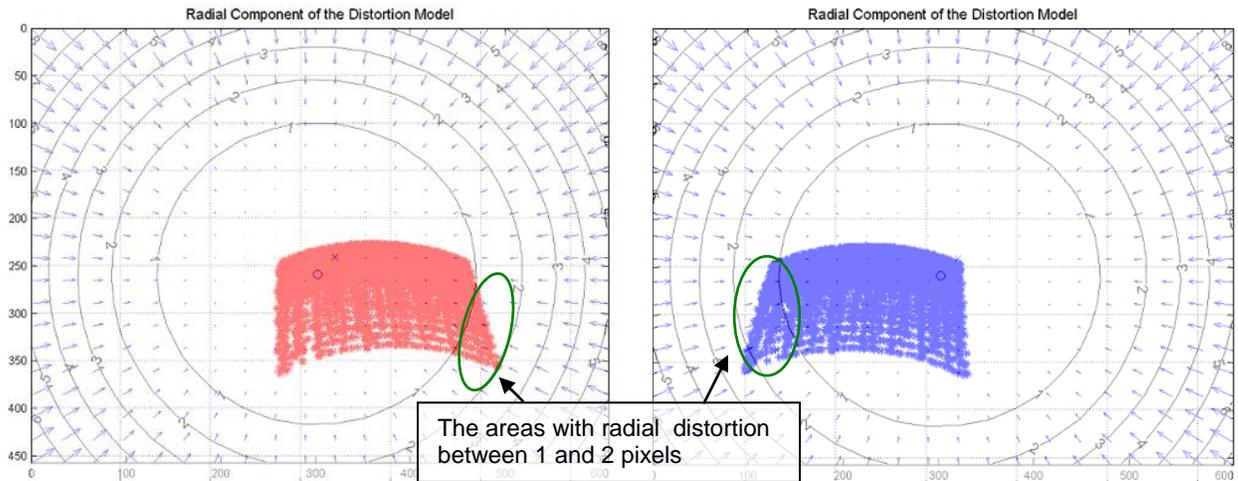


Figure 5-19. The image coordinates of the target object coincided with the radial distortion map

The above figure shows the footprint of the image coordinates of the target object used in the accuracy measurement experiment. It can be seen that only small portions of the image coordinates in both left and right image planes are placed in the areas with radial distortion between 1 and 2 pixels. The lens distortion for the areas of the images with the most population of the image coordinates is in fact less than 1 pixel. This provides some evidence of why the lens distortion correction (at most by 2 pixels) does not significantly improve the system accuracy. However, considering the portion of the data in the distortion area of (1 to 2 pixels), it can be confirmed from Table 5-3 that the accuracy has been slightly improved when the lens distortion correction is applied.

5.3 Sensitivity Analysis of the 3D Position Sensing

5.3.1 Introduction

As described earlier in this chapter, the 3D positioning system requires the image coordinates and camera models to be determined. The camera models are assumed to be accurately calculated as the result of a successful stereo camera calibration. The stereo matching process correlates the stereo coordinates in the left and right images. This process is vulnerable to number of phenomena that may disturb the correct detection of the stereo points in the image planes. Among all the possible problems, the items below are the most common:

- 1- Slight errors in the baseline distance (The distance between the camera origins). If one of the cameras moves horizontally along the baseline, the original disparity of the image coordinates changes. This introduces a horizontal displacement error between the stereo matching points.
- 2- Unlike the presented work in this thesis, some applications require the image processing to be performed in areas close to the edges of the stereo images. In such applications, stereo point matching process is supposed to provide the corresponding image coordinates in all areas of the image planes if linear lenses are employed for the cameras. This is not usually the case as the radial lens distortion is the inevitable part of the lenses. The radial lens distortion can disturb the stereo matching principle which has considered equal vertical displacement for the stereo points. However, correction of the radial lens distortion can significantly improve the distortions in those applications.
- 3- Phenomena such as occlusion and camouflage can affect the accuracy of the stereo matching algorithm. In this case the stereo matching process is not able to detect the corresponding image coordinates in the left and right planes correctly.

In this section, the sensitivity analysis experiment has been performed to verify the sensitivity of the system to camera physical motions and stereo matching errors.

5.3.2 Horizontal Pixel Error

In this section, an intentional displacement error has been applied to the horizontal displacement of the obtained image coordinate in the right camera (x_r). A cumulative error starting from 1 pixel to 50 pixels has been considered to disturb the actual image coordinates as shown in Figure 5-20. This error affects the estimated 3D Cartesian coordinates of the target point as shown in Figure 5-21.

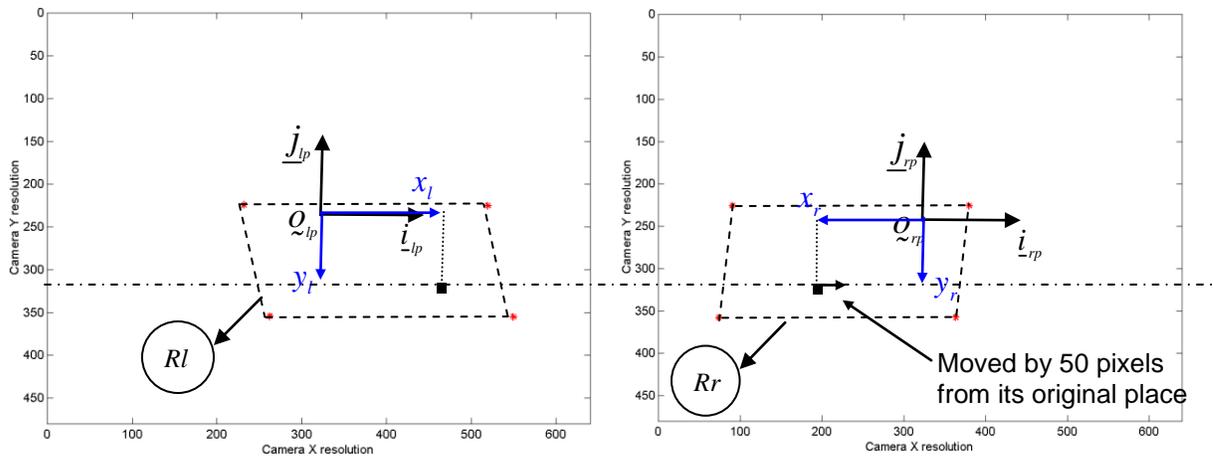


Figure 5-20. The region of interest in the left and right image planes

The dotted regions (Rl and Rr) depicted in Figure 5-20, indicates the projection of the calibration points trajectory on the first layer of the calibration volume. Only the layer with $z_0=8.5\text{cm}$ has been selected. The origin of the image coordinate system is in the centre of each image plane.

The observed error in Cartesian coordinates at a randomly selected target point ($X_t = [-14 \ 187 \ 7]^T$), has been plotted in Figure 5-21 when 1 to 50 pixel error has been added to its horizontal coordinate (x_r) in the right image plane. The plotted blue circles show the amount of errors in cm between the original coordinates of the target point and the estimated coordinates in the presence of the pixel error.

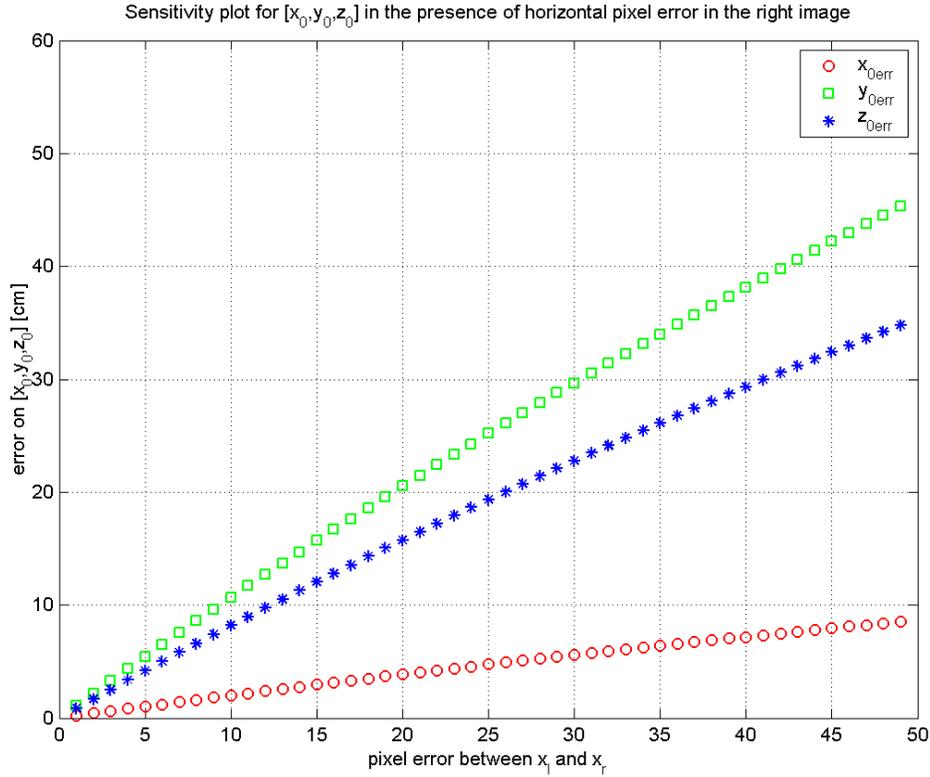


Figure 5-21. The error on the Cartesian coordinates of the target point occurred by the misplacement of the horizontal pixel in the right image plane.

The original coordinates of the arbitrary point in $\{C_0, C_0\}$ is $X_t = [-14 \ 187 \ 7]^T$ in cm.

Based on the plotted errors in Figure 5-21, one can verify that if there is 10 pixels error on horizontal displacement of the target image coordinate in the right camera, the amount of error introduced on the target point ($X_t = [-14 \ 187 \ 7]^T$) are approximately 2cm, 10.6cm and 8.16cm on its Cartesian coordinates x_0 , y_0 and z_0 , respectively. It has been verified that one pixel error on the horizontal coordinates in the right image introduces 0.17cm, 0.91cm and 0.70cm error in average on the on x_0 , y_0 and z_0 , respectively. Considering the frame C_0 in Figure 5-1, the errors are introduced as follows: 0.17cm along \underline{i}_0 (side), 0.91cm along \underline{j}_0 (depth) and 0.70cm along \underline{k}_0 (up) for every pixel error.

5.3.3 Vertical Pixel Error

In another attempt, the cumulative pixel error has been applied to the vertical displacement of the image coordinates in the right image plane.

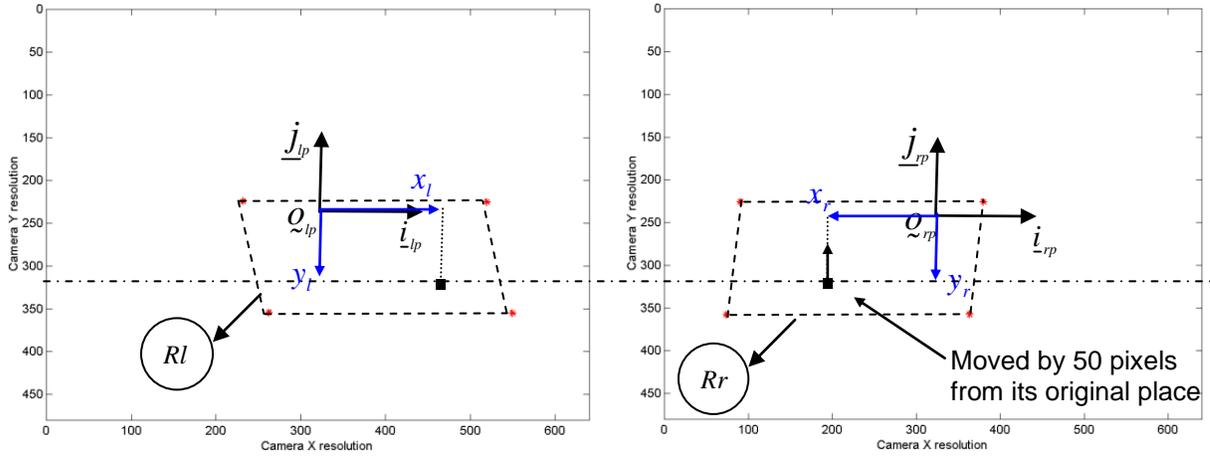


Figure 5-22. The region of interest in the left and right image planes

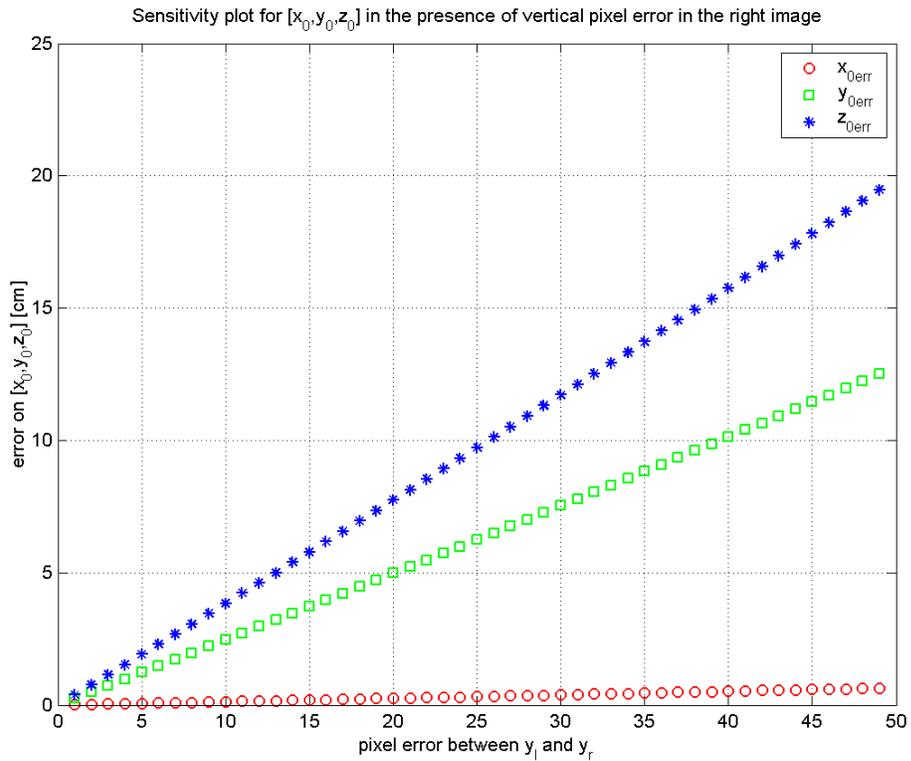


Figure 5-23. The error on the Cartesian coordinates of the target point occurred by the misplacement of the vertical pixel in the right image plane.

The original coordinates of the point in $\{Q_0, C_0\}$ is $X_t = [-14 \ 187 \ 7]^T$ in cm.

It has been verified that one pixel error on the vertical coordinates in the right image introduces 0.01cm, 0.25cm and 0.39cm error in average on the Cartesian coordinates x_0 , y_0 and z_0 , respectively.

Based on the simulation performed for the arbitrary point in this section, the three discussions below can be considered:

- The horizontal pixel misplacement introduces larger error on Cartesian coordinates of the estimated 3D positions compare to the case when the vertical pixel misplacement is applied. This shows that the 3D position sensing system is more sensitive to pixel errors occurred horizontally. In fact, the horizontal pixel misplacements can be interpreted as disparity changes in the stereo images.
- The pixel errors on the image coordinates can be interpreted as random vertical or horizontal camera motions. The errors appeared in the image planes should be interpreted as very small physical camera motions. If one considers the width of the camera CCD plane approximately equal to 1.5cm while it contains 640 pixels in horizontal resolution, every 10 pixels are equal to 0.23mm. This shows how sensitive the camera system could be to the image noises and camera vibration.
- To reduce the effects of the errors on the image coordinates, it is important to employ a highly accurate image processing algorithm for precise detection of the image coordinates. Image noise removal and moving average techniques can be considered as well to reduce the observed pixel errors.

5.4 Joint Angle Independent Visual Servoing

As explained in Chapter 2, a torso-mounted camera location has been considered for the proposed visual servoing system. The torso-mounted position constructs one DOF for the stereo cameras. This degree of freedom is in fact the TDH robot base angle (θ_1). Reliable visual servoing should be performed when the robot base angle is changing while the target object remains in its stationary 3D position. In this section, an experiment was run to verify the accuracy of the vision system during random robot motion. An LED was selected to be the target object. The target object was placed in a known 3D position to the base coordinate system $\{q_0, \underline{C}_0\}$. The system was programmed to rotate the robot torso with joint angle (θ_1). Obviously to obtain the 3D position of the stationary target object, the base angle of the robot should be compensated during the angular motion of the robot torso. If the estimated 3D position does not change during the robot torso angular displacement, the 3D positing system can be considered as a reliable system independent of the changes on the robot base angle.

Figure 5-24 shows the explained process as follows:

The top subplot: the image coordinates of the stationary object in the left and right cameras. As is expected the vertical displacements along \underline{j}_l and \underline{j}_r in the left and right image planes is minimal and approximately equal in both image planes. This is due to the constraint explained in Section 4.2.1 for the cameras with parallel image planes.

The middle subplot: The estimated 3D Cartesian coordinates of the target object is plotted. The dotted line close to each curve shows the actual coordinates of the target object measured accurately in coordinate system $\{q_0, \underline{C}_0\}$. The Cartesian coordinates of the target object is: $X_t = [-44 \quad 200 \quad 40]^T$.

The bottom subplot: The angular motion of the TDH robot base angle (θ_1) has been plotted. The maximum and minimum base angle in degrees is as follows:

$$\theta_{1\min} = -27.8^\circ \text{ and } \theta_{1\max} = 5.6^\circ$$

The maximum and minimum base angle limits are set to always provide the projection of the target object in both image planes. Such constraint on the base angle displacement guarantees that the image coordinates of the target object in both image planes are obtainable. Four elliptical dotted curves in Figure 5-24 relates the global extremum points of the TDH robot base angle displacement to the corresponding image coordinates and the 3D Cartesian coordinates of the target object. Lens distortion correction is employed in this experiment. In the magnified plot, maximum errors of 5.5cm, 7.2cm and 8.1cm are obtained for the stationary coordinates of $x_0 = -44$, $y_0 = 200$ and $z_0 = 40$ respectively.

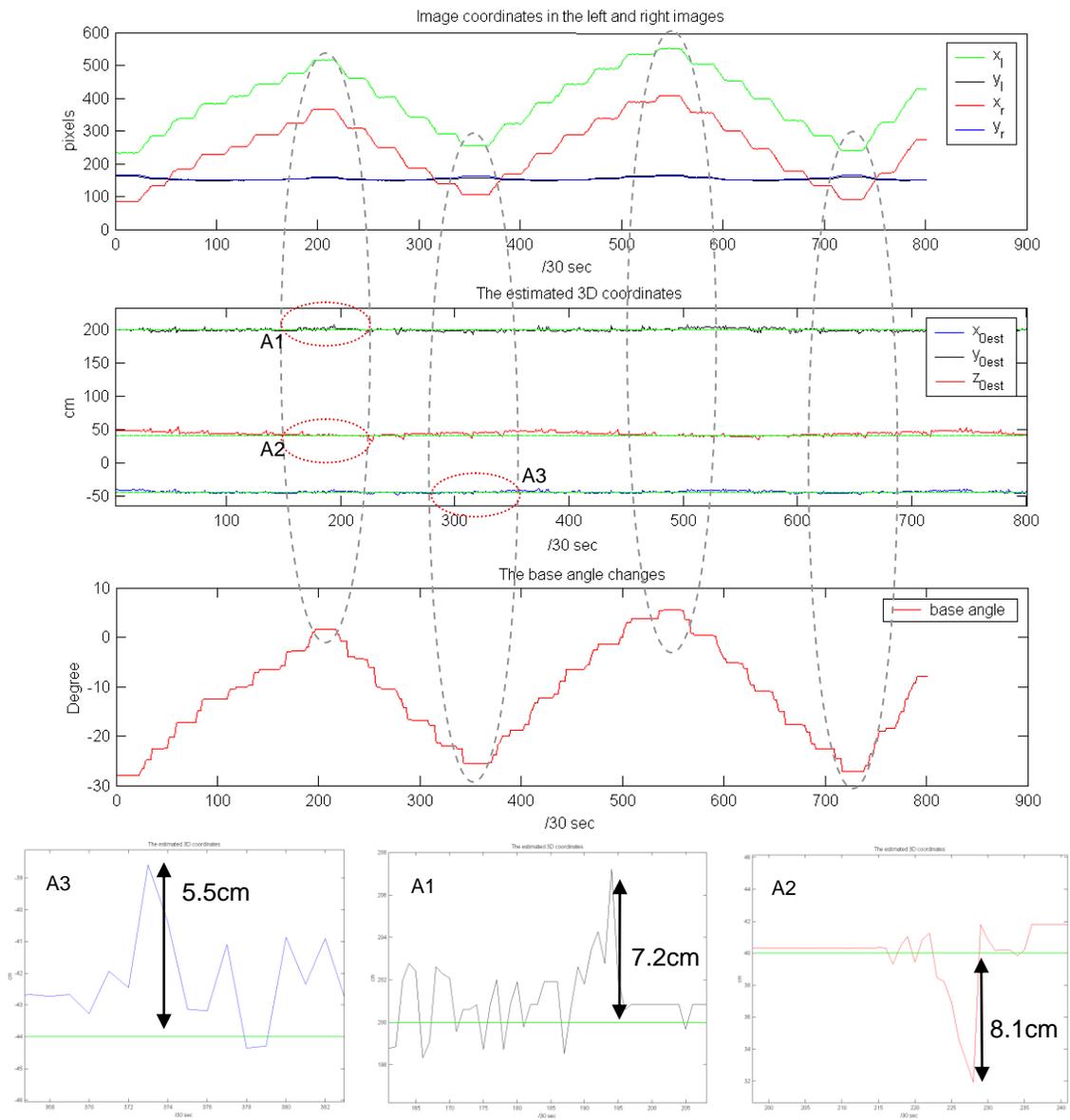


Figure 5-24. The 3D position estimation of a stationary object during the angular motion of the robot base angle (θ_1 , third subplot form top)

The table below shows the MAE and the maximum error between the estimated and actual 3D coordinates of the target object during the explained servoing task.

Table 5-7. The errors between the actual and estimated stationary coordinates

| | |
|--|--------|
| MAE on x_0, (cm) | 2.0004 |
| MAE on y_0, (cm) | 1.3593 |
| MAE on z_0, (cm) | 1.5629 |
| MAX error on x_0, (cm) | 5.5883 |
| MAX error on y_0, (cm) | 7.2053 |
| MAX error on z_0, (cm) | 8.1532 |

Chapter 6. Conclusions and Future Work

The stereo camera calibration presented in this thesis is a practical method with applications in industry. Accurate 3D position sensing can be obtained if the 3D and 2D positions of the data points in the world and image frames are provided accurately for the purpose of camera calibration and position estimation. The stereo camera pair is mounted on the robot torso using highly accurate mechanical fasteners. This allows the calibrated camera pair to be removed from the robot torso for shipping purposes. In the design and manufacturing process of the stereo camera pair and the accommodating surface on the robot torso, extra care was practiced to be able to mount the stereo rig back on its original position. However, the stereo camera pair can be recalibrated using the introduced *in situ* camera calibration technique to fine tune the camera model parameters in the case of misplacements on the robot torso. In this chapter a summary of the system field trials, problem areas and future refinements are provided.

6.1 System Evaluation

6.1.1 3D Position Sensing Accuracy

In Sections 5.2.3 and 5.2.4, the 3D position sensing accuracy was evaluated. In these two sections, the errors between the actual and the estimated 3D positions were compared when the DLT and SDLT camera calibration methods (explained in Sections 3.2 and 3.4) were used. The average error increased by 2.8% when SDLT method was used. However, the maximum error was reduced by 6.6% with the DLT method. ED stands for Euclidean Distance in the tables below.

Table 6-2. Average error in DLT and SDLT

| <u>Unit = cm</u> | Area of R2 80x50cm (DLT) | Area of R2 80x50cm (Stereo DLT) |
|-------------------------------|--------------------------------|---------------------------------------|
| MAE (x_0) | 0.8009 | 0.7454 |
| MAE (y_0) | 1.0521 | 1.1194 |
| MAE (z_0) | 1.0682 | 1.1269 |
| MAE (ED) | 1.8842 | 1.9381 |

Table 6-1.Maximum error in DLT and SDLT

| <u>Unit = cm</u> | Area of R2 80x50cm (DLT) | Area of R2 80x50cm (Stereo DLT) |
|--------------------------------|--------------------------------|---------------------------------------|
| MAXE (x_0) | 4.9524 | 4.2001 |
| MAXE (y_0) | 7.3571 | 4.8405 |
| MAXE (z_0) | 6.2757 | 6.4095 |
| MAXE (ED) | 8.1456 | 7.6061 |

6.1.2 Lens Distortion Correction Effects

The effects of lens distortion correction were shown in the accuracy measurement tables in Sections 5.2.3 and 5.2.4. The tables below are the replicas of the associated rows in the accuracy measurement tables indicating the effects of Radial Lens Distortion Correction (RLDC).

Table 6-4. Effects of RLDC in DLT method

| <u>Unit = cm</u> | Area of R2 80x50cm (DLT) NLDC | Area of R2 80x50cm (DLT) WLDC |
|------------------|--|--|
| MAE (x_0) | 0.7982 | 0.8009 |
| MAE (y_0) | 1.0485 | 1.0521 |
| MAE (z_0) | 1.1102 | 1.0682 |
| MAE (ED) | 1.9100 | 1.8842 |

Table 6-3. Effects of RLDC in SDLT method

| <u>Unit = cm</u> | Area of R2 80x50cm (Stereo DLT) NLDC | Area of R2 80x50cm (Stereo DLT) WLDC |
|------------------|---|---|
| MAE (x_0) | 0.7553 | 0.7454 |
| MAE (y_0) | 1.2575 | 1.1194 |
| MAE (z_0) | 1.1675 | 1.1269 |
| MAE (ED) | 1.9778 | 1.9381 |

NLDC = No Lens Distortion Correction, WLDC = With Lens Distortion Correction

Table 6-5. Max error with RLDC (DLT)

| <u>Unit = cm</u> | Area of R2 80x50cm (DLT) NLDC | Area of R2 80x50cm (DLT) WLDC |
|------------------|--|--|
| MAXE (x_0) | 4.9006 | 4.9524 |
| MAXE (y_0) | 6.5344 | 7.3571 |
| MAXE (z_0) | 6.9241 | 6.2757 |
| MAXE (ED) | 8.1945 | 8.1456 |

Table 6-6. Max error with RLDC (SDLT)

| <u>Unit = cm</u> | Area of R2 80x50cm (Stereo DLT) NLDC | Area of R2 80x50cm (Stereo DLT) WLDC |
|------------------|---|---|
| MAXE (x_0) | 4.1995 | 4.2001 |
| MAXE (y_0) | 4.8104 | 4.8405 |
| MAXE (z_0) | 7.0118 | 6.4095 |
| MAXE (ED) | 8.3349 | 7.6061 |

MAXE = Maximum Error

As stated in Section 5.2.5, the radial lens distortion correction does not significantly decrease the 3D position sensing average errors for both DLT and SDLT methods. The lens distortion correction reduces the average error by 1.35% when the camera model is obtained using the DLT method. This reduction is 2% when the camera model is obtained using SDLT method. However, the maximum error has dropped by 18% when the lens distortion correction is employed in the SDLT method. This shows

that the stereo assumption ($y_l = y_r = y$) has improved the worst case error (only for the z component shown in Table 6-6) when the SDLT method was employed. In fact, the system suffered from the inequality in the pixel vertical displacement (due to image processing) when the DLT method was used.

6.1.3 Real-world System Test

In the course of development of the system, three site visits to the final installation location of the system were performed. The first two trips were arranged with the main goals below:

- 1- Get familiar with the industrial requirements of the system
- 2- Examine different locations for the stereo camera pair
- 3- Verify the environmental conditions to meet the outdoor requirements

The third visit was performed to install the first prototype of the completed visual servoing system on the original location of the system. The performance of the visual servoing system was fully investigated and very good results were obtained during the system servoing tasks. The system shortcomings and the problem areas were documented to be used in the future improvements of the system.

6.1.4 Vision System Errors

Detection of the target image coordinates is required in both camera calibration and 3D position sensing. Hence, accurate correspondence between the image points must be performed during all vision related algorithms. The accuracy of the image processing algorithm becomes worse in the existence of image noise. With the cost of the system speed (dropping the image frame acquisition rate), algorithms such as template matching can be used to detect the target points in the stereo images.

6.1.5 Mechanical Inaccuracies

As in any mechanical system, there is inaccuracy in endpoint motion due to sensor error and servoing errors. These errors contribute to the overall system error. E.g. a maximum error of approximately 6cm was measured mechanically on the end-effector height relative to the robot base coordinate system when the end-effector was fully

retracted. Considering Figure 2-9, this error occurs during the radial motion of the robot arm.

6.1.6 Comparing with Affine (Weak Perspective) Stereo Systems

The stereo camera calibration method implemented in this thesis employs the pinhole camera model with full perspective transformation. Hollinghurst and Cipolla [78] have introduced a camera calibration technique based on weak perspective camera modeling. Their method is fast and the camera model requires fewer parameters than the implemented model in this thesis. The weak perspective camera model was evaluated for use in the proposed visual servoing system. It was observed that the average error in 3D position estimation using the full perspective (pinhole camera) is 57% of the average error when the weak perspective was used.

6.2 Future Refinements

6.2.1 An Automatic Accuracy Test for Obtained Camera Model

There is no specific testing procedure in this thesis to verify the accuracy of the obtained camera model. This verification can be accomplished by estimation of 3D positions of some known target objects and comparing them with their actual positions. E.g. one can program the robot to move its end-effector to a few known positions in the stereo camera FOV. The 3D estimation algorithm detects the position of the end-effector and it will be compared with the actual value obtained from the arm forward kinematics process. If the 3D positioning system introduces errors more than what the system tolerates, the *in situ* camera calibration process can be repeated to recalibrate the stereo cameras. It is important to make sure that during this process the image coordinates are detected accurately.

6.2.2 Accurate Camera Calibration Platform

As stated in Section 6.1.6, mechanical error disturbs the robot endpoint motion and this introduces unexpected error in positioning the end-effector. The camera model can be estimated with high accuracy if the 3D positions of the control points are obtained accurately. The first solution to address this problem is to request that the robot

manufacture improves the mechanical problems of the robot. This solution could be costly and specially may not be considered by the manufacture as a necessary fix since the errors reside in the system accuracy tolerance. A simple and cheap procedure is to fit polynomial functions to estimate the position of the end-effector. This requires sample known positions of the end-effector together with accurate encoders reading.

References

- [1] S. Hutchinson, , G.D. Hager, and P.I. Corke, “A Tutorial on Visual Servo Control,” *Robotics and Automation, IEEE Transactions*, Volume 12, Issue 5, Oct. 1996 Page(s):651 - 670 Digital Object Identifier 10.1109/70.538972
- [2] C. Brenneke, O. Wulf, and B. Wagner, “Using 3D Laser Range Data for SLAM in Outdoor Environments,” *Processing of the 2003 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems*, Las Vegas, Nevada. October 2003
- [3] O. Wulf and B.Wagner, “Fast 3D scanning Methods for Laser Measurement Systems,” *International conference on Control Systems and Computer Science*, vol 1, Bucharest, Romania, July 2003, pp 312-317
- [4] T. C. Ng, J. Ibanez, J. Chang Tan, “Development of a 3D LADAR System for Autonomous Navigation,” *Proceeding of the 2004 IEEE Conference on Robotics, Automation and Mechatronics*, Singapore, 1-3 December. 2004
- [5] A. Nuchter, K. Lingemann, J. Hertzberg and Hartmut Surmann, “Accurate Object Localization in 3D Laser Range Scans,” *Advanced Robotics, 2005. ICAR '05. Proceedings.*, 12th International Conference on July 18-20, 2005 Page(s):665 – 672
- [6] Datong Chen, Wen Gao, and Xilin Chen, “A new approach of recovering 3-D shape from structure-lighting,” *Signal Processing, 1996.*, 3rd International Conference on Volume 2, 14-18 Oct. 1996 Page(s):839 - 842 vol.2 Digital Object Identifier 10.1109/ICSIGP.1996.566216
- [7] J. Park, G.N. DeSouza, and A.C. Kak, “Dual Beam Structured-light Scanning for 3-D Object Modeling,” *3-D Digital Imaging and Modeling, 2001. Proceedings.* 28 May-1 June 2001 Page(s):65 - 72 Digital Object Identifier 10.1109/IM.2001.924399
- [8] N.G. Durdle, J. Thayoor, V.J. Raso,” An improved structured light technique for surface reconstruction of the human trunk,” *Electrical and Computer Engineering, 1998. IEEE Canadian Conference*, Volume 2, 24-28 May 1998 Page(s):874 - 877 vol.2 Digital Object Identifier 10.1109/CCECE.1998.685637

- [9] H. -G Maas, "Robust Automatic Surface Reconstruction with Structured light," International Archives of Photogrammetry and Remote Sensing Vol. XXIX (1992), Part B5, pp. 709-713
- [10] Y. Shirai and H. Inoue, "Guiding a robot by visual feedback in assembling tasks," *Pattern Recognition.*, vol. 5. pp. 99-108, 1973.
- [11] A. Muis, O. Kouhei, "An iterative approach in pose measurement through hand-eye calibration," Control Applications, 2003. CCA 2003. Proceedings of 2003 IEEE Conference, Volume 2, 23-25 June 2003 Page(s):983 - 988 vol.2 Digital Object Identifier 10.1109/CCA.2003.1223144
- [12] P. K. Allen, B. Yoshimi, A. Timcenko, and P. Michelman, " Automated Tracking and Grasping of a Moving Object with a Robotic Hand-eye System," 152 IEEE IEEE TRANSACTIONS ON ROBOTICS AND AUTOMATION, VOL. 9, NO. 2, APRIL 1993
- [13] T.I.J. Tsay, C.J. Chang, "Pose control of mobile manipulators with an uncalibrated eye-in-hand vision system," Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference, Volume 3, 28 Sept.-2 Oct. 2004 Page(s):3033 - 3038 vol.3 Digital Object Identifier 10.1109/IROS.2004.1389871
- [14] L. Bascetta, P. Rocco, "Task space visual servoing of eye-in-hand flexible manipulators," Advanced Intelligent Mechatronics, 2003. AIM 2003. Proceedings. 2003 IEEE/ASME International Conference, Volume 2, 20-24 July 2003 Page(s):1442 - 1448 vol.2 Digital Object Identifier 10.1109/AIM.2003.1225555
- [15] C.E. Smith, S.A. Brandt, and N.P. Papanikolopoulos, "Eye-in-hand robotic tasks in uncalibrated environments," Robotics and Automation, IEEE Transactions, Volume 13, Issue 6, Dec. 1997 Page(s):903 – 914, Digital Object Identifier 10.1109/70.650169
- [16] C. Kulpate, M. Mehrazdeh, and R. Paranjape, "An eye-to-hand visual servoing structure for 3D positioning of a robotic arm using one camera and a flat mirror," Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference, 2-6 Aug. 2005 Page(s):1464 - 1470 Digital Object Identifier 10.1109/IROS.2005.1545348
- [17] C. Kulpate, M. Mehrazdeh, and R. Paranjape, "An eye-to-hand visual servoing structure for 3D positioning of a robotic arm using one camera and a flat mirror," Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference, 2-6 Aug. 2005 Page(s):1464 - 1470 Digital Object Identifier 10.1109/IROS.2005.1545348
- [18] S. Jafari, R. Jarvis, and T. Sivahumaran, "Relative Visual Servoing," Robotics, Automation and Mechatronics, 2004 IEEE Conference, Volume 2, 1-3 Dec. 2004 Page(s):880 - 885 vol.2
- [19] A. Muis, K. Ohnishi, "Eye-to-hand approach on eye-in-hand configuration within real-time visual servoing," Mechatronics, IEEE/ASME Transactions, Volume 10, Issue 4, Aug. 2005 Page(s):404 - 410 Digital Object Identifier 10.1109/TMECH.2005.852397
- [20] G. Flandin, F. Chaumette, and E. Marchand, "Eye-in-hand/eye-to-hand cooperation for visual servoing," Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE

International Conference, Volume 3, 24-28 April 2000 Page(s):2741 - 2746 vol.3
Digital Object Identifier 10.1109/ROBOT.2000.846442

- [21] P. Martinet, J. Gallice, "Position based visual servoing using a non-linear approach," Intelligent Robots and Systems, 1999. IROS '99. Proceedings. 1999 IEEE/RSJ International Conference, Volume 1, 17-21 Oct. 1999 Page(s):531 - 536 vol.1
Digital Object Identifier 10.1109/IROS.1999.813058
- [22] G. Taylor, L. Kleeman, "Hybrid *position-based* visual servoing with online calibration for a humanoid robot," Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference, Volume 1, 28 Sept.-2 Oct. 2004 Page(s):686 - 691 vol.1
Digital Object Identifier 10.1109/IROS.2004.1389432
- [23] W.J. Wilson, W.J. Williams, C.C. Hulls, G.S. Bell, "Relative end-effector control using Cartesian position based visual servoing," Robotics and Automation, IEEE Transactions, Volume 12, Issue 5, Oct. 1996 Page(s):684 – 696
Digital Object Identifier 10.1109/ROBOT.1996.131968
- [24] K. Hashimoto, T. Kimoto, T. Ebine, and H. Kimura, "Manipulator control with image-based visual servo," Robotics and Automation, 1991. Proceedings., 1991 IEEE International Conference, 9-11 April 1991 Page(s):2267 - 2271 vol.3
Digital Object Identifier 10.1109/ROBOT.1991.131968
- [25] Y.T. Kim, S.C. Fan, S.H. Han, and H.S. Go, "Image-based visual feedback control of a dual-arm robot," Industrial Electronics, 2001. Proceedings. ISIE 2001. IEEE International Symposium, Volume 3, 12-16 June 2001 Page(s):1603 - 1608 vol.3
Digital Object Identifier 10.1109/ISIE.2001.931946
- [26] S.H. Han, W.H. See, J. Lee, M.H. Lee, and H. Hashimoto, "Image-based visual servoing control of a SCARA type dual-arm robot," Industrial Electronics, 2000. ISIE 2000. Proceedings of the 2000 IEEE International Symposium, Volume 2, 4-8 Dec. 2000 Page(s):517 - 522 vol.2
Digital Object Identifier 10.1109/ISIE.2000.930351
- [27] K.G. Stanley, Q.M.J Wu, and W.A. Gruver, "An intelligent dual mode vision guided robotic system," Systems, Man and Cybernetics, 2005 IEEE International Conference, Volume 3, 10-12 Oct. 2005 Page(s):2248 - 2253 Vol. 3
Digital Object Identifier 10.1109/ICSMC.2005.1571483
- [28] F. Conticelli, B. Allotta, "Two-level visual control of dynamic look-and-move systems," Robotics and Automation, 2000. Proceedings. ICRA '00. IEEE International Conference, Volume 4, 24-28 April 2000 Page(s):3784 - 3789 vol.4
Digital Object Identifier 10.1109/ROBOT.2000.845321
- [29] T.I.J Tsay, C.J. Chang, "Pose control of mobile manipulators with an uncalibrated eye-in-hand vision system," Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference, Volume 3, 28 Sept.-2 Oct. 2004 Page(s):3033 - 3038 vol.3
Digital Object Identifier 10.1109/IROS.2004.1389871
- [30] M. Bachiller, A. Adam, V. Feliu, and C. Cerrada, "Well structured robot positioning control strategy for position based visual servoing," Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference, Volume 3, 2001 Page(s):2541 - 2546 vol.3
Digital Object Identifier 10.1109/ROBOT.2001.933005

- [31] R. Mahony, T. Hamel, "Image-based visual servo control of aerial robotic systems using linear image features," *Robotics, IEEE Transactions on* [see also *Robotics and Automation, IEEE Transactions*, Volume 21, Issue 2, April 2005 Page(s):227 - 239 Digital Object Identifier 10.1109/TRO.2004.835446
- [32] L. E. Weiss, A. C. Sanderson and C. P. Neuman, "Visual Servo Control of Robots, an adaptive image-based approach," *Robotics and Automation. Proceedings. 1985 IEEE International Conference*, volume 2, Mar 1985 Page(s):662 - 668
- [33] W.E. Dixon, "Teach by zooming: a camera independent alternative to teach by showing visual servo control," *Intelligent Robots and Systems, 2003. (IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference*, Volume 1, 27-31 Oct. 2003 Page(s):749 - 754 vol.1 Digital Object Identifier 10.1109/IROS.2003.1250719
- [34] A. Ruf, M. Tonko, R. Horaud, and H.-H Nagel, "Visual tracking of an end-effector by adaptive kinematic prediction," *Intelligent Robots and Systems, 1997. IROS '97., Proceedings of the 1997 IEEE/RSJ International Conference*, Volume 2, 7-11 Sept. 1997
- [35] Kang Qingsheng, Meng Zhengda, and Dai Xianzhong, "A new visual servo scheme without calibration," *Mechatronics and Automation, 2005 IEEE International Conference*, Volume 4, 29 July-1 Aug. 2005 Page(s):2134 - 2137 Vol. 4
- [36] K. Toyama, G.D. Hager, and J. Wang, "Servomatic: a modular system for robust positioning using stereo visual servoing," *Robotics and Automation, 1996. Proceedings., 1996 IEEE International Conference*, Volume 3, 22-28 April 1996 Page(s):2636 - 2642 vol.3 Digital Object Identifier 10.1109/ROBOT.1996.506560
- [37] Y. Kuno, H. Numagami, M. Ishikawa, H. Hoshino, and M. Kidode, "Three-dimensional vision techniques for an advanced robot system," *Robotics and Automation. Proceedings. 1985 IEEE International Conference*, Volume 2, Mar 1985 Page(s):11 - 16
- [38] L.C. Silva, A. Petraglia, M.R. Petraglia, "Stereo vision system for real time inspection and 3D reconstruction," *Industrial Electronics, 2003. ISIE '03. 2003 IEEE International Symposium*, Volume 1, 9-11 June 2003 Page(s):607 - 611 vol. 1
- [39] Y. Yamamoto, I. Yoda, K. Sakaue, "Arm-pointing gesture interface using surrounded stereo cameras system," *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference*, Volume 4, 23-26 Aug. 2004 Page(s):965 - 970 Vol.4 Digital Object Identifier 10.1109/ICPR.2004.1333934
- [40] Lee DooHyun, Kweon InSo, "A novel stereo camera system by a biprism," *Robotics and Automation, IEEE Transactions*, Volume 16, Issue 5, Oct. 2000 Page(s):528 - 541 Digital Object Identifier 10.1109/70.880803
- [41] R. Tsai, "A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses," *Robotics and Automation, IEEE Journal of* [legacy, pre - 1988], Volume 3, Issue 4, Aug 1987 Page(s):323 - 344
- [42] Yongjie Yan, Qidan Zhu, Zhuang Lin, and Quanfu Chen, " Camera Calibration in Binocular Stereo Vision of Moving Robot," *Intelligent Control and Automation, 2006.*

WCICA 2006. The Sixth World Congress, Volume 2, 21-23 June 2006 Page(s):9257 - 9261 Digital Object Identifier 10.1109/WCICA.2006.1713792

- [43] E.E. Hemayed, "A survey of camera self-calibration," Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2003., 21-22 July 2003 Page(s):351 - 357 Digital Object Identifier 10.1109/AVSS.2003.1217942
- [44] R.Y. Tsai, R.K. Lenz, "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration," Robotics and Automation, IEEE Transactions, Volume 5, Issue 3, June 1989 Page(s):345 - 358 Digital Object Identifier 10.1109/70.34770
- [45] N. Pessel, J. Opderbecke, M.-J Aldon, "An experimental study of a robust self-calibration method for a single camera," Image and Signal Processing and Analysis, 2003. ISPA 2003. Proceedings of the 3rd International Symposium, Volume 1, 18-20 Sept. 2003 Page(s):522 - 527 Vol.1 Digital Object Identifier 10.1109/ISPA.2003.1296952
- [46] Hongshan Yu, Yaonan Wang, "An Improved Self-calibration Method for Active Stereo Camera," Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress, Volume 1, 21-23 June 2006 Page(s):5186 - 5190 Digital Object Identifier 10.1109/WCICA.2006.1713380
- [47] M.A. Sid-Ahmed, M.T. Boraie, "Dual camera calibration for 3-D machine vision metrology," Instrumentation and Measurement, IEEE Transactions, Volume 39, Issue 3, June 1990 Page(s):512 - 516 Digital Object Identifier 10.1109/19.106283
- [48] Marita, T.; Oniga, F.; Nedevschi, S.; Graf, T.; Schmidt, R., "Camera Calibration Method for Far Range Stereovision Sensors Used in Vehicles," Intelligent Vehicles Symposium, 2006 IEEE, 13-15 June 2006 Page(s):356 - 363
- [49] Zhengyou Zhang, "Camera calibration with one-dimensional objects," Pattern Analysis and Machine Intelligence, IEEE Transactions, Volume 26, Issue 7, July 2004 Page(s):892 - 899 Digital Object Identifier 10.1109/TPAMI.2004.21
- [50] M. Malle, M. Shaheen, F. Chavand, "Automatic camera calibration based on robot calibration," Instrumentation and Measurement Technology Conference, 1999. IMTC/99. Proceedings of the 16th IEEE, Volume 2, 24-26 May 1999 Page(s):1278 - 1282 vol.2 Digital Object Identifier 10.1109/IMTC.1999.777060
- [51] Christof Ridder, Olaf Munkelt, and Harald Kirchner, "Adaptive Background Estimation and Foreground Detection using Kalman-Filtering," Proceedings of International Conference on recent Advances in Mechatronics, ICRAM'95, UNESCO Chair on Mechatronics, 193-199, -1995.
- [52] Christopher R. Wren, Ali Azarbayejani, Trevor Darrell, and Alex Pentland. "Pfunder: Real-Time Tracking of the Human Body," In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, July 1997, vol 19, no 7, pp. 780-785.
- [53] C. Stauffer, W.E.L Grimson, "Adaptive background mixture models for real-time tracking," Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference, Volume 2, 23-25 June 1999 Digital Object Identifier 10.1109/CVPR.1999.784637

- [54] Zang Qi, R. Klette, "Robust background subtraction and maintenance," Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference, Volume 2, 23-26 Aug. 2004 Page(s):90 - 93 Vol.2 Digital Object Identifier 10.1109/ICPR.2004.1334047
- [55] M. Piccardi, "Background subtraction techniques: a review," Systems, Man and Cybernetics, 2004 IEEE International Conference, Volume 4, 10-13 Oct. 2004 Page(s):3099 - 3104 vol.4 Digital Object Identifier 10.1109/ICSMC.2004.1400815
- [56] Lim Ser-Nam, A. Mittal, L.S. Davis, and N. Paragios, "Fast illumination-invariant background subtraction using two views: error analysis, sensor placement and applications," Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference, Volume 1, 20-25 June 2005 Page(s):1071 - 1078 vol. 1 Digital Object Identifier 10.1109/CVPR.2005.155
- [57] A. Yoneyama, C.H. Yeh, C.-C.J. Kuo, "Moving cast shadow elimination for robust vehicle extraction based on 2D joint vehicle/shadow models," Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance, 2003. 21-22 July 2003 Page(s):229 - 236 Digital Object Identifier 10.1109/AVSS.2003.1217926
- [58] Y. Matsushita, K. Nishino, K. Ikeuchi, M. Sakauchi, "Shadow elimination for robust video surveillance," Motion and Video Computing, 2002. Proceedings. Workshop, 5-6 Dec. 2002 Page(s):15 - 21 Digital Object Identifier 10.1109/MOTION.2002.1182208
- [59] Chang Chia-Jung, Hu Wen-Fong, Hsieh Jun-Wei, and Chen Yung-Sheng, "Shadow elimination for effective moving object detection with Gaussian models," Pattern Recognition, 2002. Proceedings. 16th International Conference, Volume 2, 11-15 Aug. 2002 Page(s):540 - 543 vol.2 Digital Object Identifier 10.1109/ICPR.2002.1048359
- [60] G.D. Finlayson, S.D. Hordley, Lu Cheng, and M.S. Drew, "On the removal of shadows from images," Pattern Analysis and Machine Intelligence, IEEE Transactions, Volume 28, Issue 1, Jan. 2006 Page(s):59 - 68 Digital Object Identifier 10.1109/TPAMI.2006.18
- [61] Richard Hartley, Andrew Zisserman., *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [62] Z.Y. Zhang. "Flexible Camera Calibration by Viewing a Plane from Unknown Orientations,". In: IEEE kim.L.Boyer,Sudeep.Sarkar, eds.IEEE International Conference on Computer Vision, Greece. Washington: IEEE Computer Society, 1999_666-673P.
- [63] O. Faugeras, *Three Dimensional Computer Vision*. MIT press, 1992.
- [64] D. C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.
- [65] W. Faig. "Calibration of close-range photogrammetry systems: Mathematical formulation," *Photogrammetric Engineering and Remote Sensing*, 41(12):1479–1486, 1975.

- [66] C. C. Slama, editor. *Manual of Photogrammetry*. American Society of Photogrammetry, 4th ed., 1980.
- [67] J. Weng, P. Cohen, and M. Herniou. "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(10):965–980, Oct. 1992.
- [68] G. Wei and S. Ma. "Implicit and explicit camera calibration: Theory and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(5):469–480, 1994.
- [69] J-Y Bouguet, Camera Calibration Toolbox for Matlab,
http://www.vision.caltech.edu/bouguetj/calib_doc/index.html#examples
- [70] J. Heikkila, O. Silven, "A four-step camera calibration procedure with implicit image correction," *Computer Vision and Pattern Recognition*, 1997. Proceedings., 1997 IEEE Computer Society Conference on 17-19 June 1997 Page(s):1106 – 1112 Digital Object Identifier 10.1109/CVPR.1997.609468
- [71] Y.I. Abdel-Aziz, H.M. Karara, (1971), "Direct linear transformation into object space coordinates in close-range photogrammetry," *Proc. Symposium on Close-Range Photogrammetry*, Urbana, Illinois, p. 1 - 18.
- [72] Modular sensor platform robotic vehicle, US patent 7,144,057
- [73] Q.-T Luong, O.D. Faugeras, "Self-calibration of a camera using multiple images," *Pattern Recognition*, 1992 . Vol.1. Conference A: Computer Vision and Applications, Proceedings., 11th IAPR International Conference on 30 Aug 1992 Page(s):9 – 12
- [74] D. H. Ballard and Broxwn, C M. *Computer Vision*. Prentice-Hall, 1982.
- [75] H.A. Beyer, "Some Aspects of the Geometric Calibration of CCD-Camera," in *ISPRiS Fast Processing of photogrannetry Data.interlake*, 1987.
- [76] Li Jing Chen, "CCD Cam:era Calibration with DLT," in *Journal of CoMtter*, 1990, no.3.
- [77] K. J. Astrom and B. Wittenmark, *Adaptive Control*, Boston: Addison Wesley, 1995.
- [78] N. Hollinghurst and R. Cipolla, "Uncalibrated stereo hand eye coordination," *Image und Vision Computing*. vol. 12, no. 3, pp. 187-192, 1994.

Appendix A

System Components

The proposed visual servoing system contains the main hardware components below:

1- TDH robotic manipulator:

A 3DOF custom made robot to be installed on the so-called Monkey Board (MB) structure. The robot is designed and manufactured by Roboweld, Coquitlam, BC. The stereo cameras are installed on the robot torso as shown in Figure 2-1. The robotic arm is equipped with the following components:

- *Two AC servo motors*: The first servo motor provides the revolute motion of the robot torso. This DOF is called TDH robot base angle indicated by θ_1 in Figure 2-1. The second servo motor provides the radial motion on the robot arm by driving a linear actuator installed inside the robot column (torso). This DOF is indicated by d_2 .
- *Stepper motor*: Provides the revolute motion on the robot end-effector (gripper). The DOF is indicated by θ_3 .
- *Absolute encoders*: The servo and stepper motors are equipped with absolute encoders to be read by the PLC unit. The robot joint angles are obtained by reading the employed absolute encoders.

2- PC104 embedded system with X86 CPU architecture:

A heavy-duty rugged PC104 embedded system to transfer the calculated joint angle commands to the (Programmable Logic Control) PLC and motion controller units. The PC104 Embedded CPU box has been designed and manufactured by Motion Metrics International, Vancouver, BC. The CPU box contains the hardware below:

- *CPU board*: The image processing and machine vision algorithms are executed using its 900MHz processor.

- *Frame grabber*: The NTSC signals (analog video) from the stereo cameras are acquired and converted to digital frames using a high performance four channel frame grabber with maximum of 30Hz frame rate.
- *DAQ board*: The main purpose of using this board is to benefit from its onboard hardware interrupt. Delivering a constant timing between different software threads requires the interrupts to be generated using software independent routines. The other use of this board is to read the analog voltages of a joystick employed for manual motion control of the TDH robot as well as reading of other analog sensor signals such as those of accelerometers. The mentioned joystick and analog sensors have not been discussed in this thesis.

-

3- Programmable Logic Control (PLC) and Motion Controller:

An industrial PLC and motion controller systems employed for robot motion control using the joint angle commands sent from the PC104 system. The system contains the following components:

- *PLC unit*: Two primary tasks are performed by this unit. The PLC system is a bridge between the PC104 embedded system and the motion control unit. It transfers the desired joint angles calculated by the PC104 to the motion controller unit. The other main task of this unit is to control/execute all the required pipe handling tasks defined by the application requirements.
- *Motion Controller unit*: This component performs the joint level closed-loop motion control of the robot.

4- Servo motor drives:

These units receive the control signals from the motion controller unit and generate the required amplified voltages to be used by the servo motors. Each servo motor requires a servo drive unit.

5- Stereo camera rig:

The stereo camera rig accommodates the two monochrome analog cameras. The cameras are heavy-duty and specially designed and manufactured for harsh

environment such as mining and underwater industries. This unit is designed and fabricated by Motion Metrics International and Roboweld.

6- LCD touch screen monitor:

The touch screen monitor is used to show the interactive Graphical User Interface (GUI) and to provide a visual perception of the tripping actions using the real-time video signals from the stereo cameras.



**Torso-mounted Visual Servoing of an Outdoor Robotic
Manipulator**

by

Nima Zirknejad

April 2007

**Department of Electrical and Computer Engineering
Faculty of Applied Science
The University of British Columbia**