STEREO VISION USING COMPENSATION TECHNIQUE FOR IMPROVEMENT OF DISTANCE MEASUREMENT ACCURACY FOR ROBOT ARM APPLICATIONS

ENG SWEE KHENG

UNIVERSITI SAINS MALAYSIA
2012

STEREO VISION USING COMPENSATION TECHNIQUE FOR IMPROVEMENT OF DISTANCE MEASUREMENT ACCURACY FOR ROBOT ARM APPLICATIONS

by

ENG SWEE KHENG

Thesis is submitted in fulfillment of the requirements
for the Degree of
Master of Science (Electrical & Electronic Engineering)

February 2012

ACKNOWLEDGEMENTS

First of all, I would like to thank Universiti Sains Malaysia and School of Electrical and Electronic Engineering for providing me the postgraduate project and sponsor me to implement my project by Postgraduate Research Grant Scheme (PRGS), letting me to gain exposure regarding to the designation of this project that I am going to cope with in the future.

I would like to thank my supervisor, Dr. Anwar Hasni Abu Hassan for providing me the chance to accomplish my postgraduate project. I am grateful to my supervisor for his guidance, encouragement, permission and approval along the project designation. Special thanks to him for the opportunity given to me to complete my postgraduate project and for the hospitality extended to me. He gave me a lot of attention, encouragement, guidance, and willing to share his experience with me. Besides that, I would like to express my deepest gratitude to my lecturer, Professor Dr. Lim Chee Peng. He always encourages me and providing me many opinions during project designation and implementation.

Furthermore, I would like to give my special thanks to the technicians i.e. Mr. Jamaluddin bin Che Mat, Mr. Hairul Nizam bin Abdul Rahman, Pn. Mazlina bt Ghani, Mr. Nor Azhar bin Zabidin and Mr. Amir bin Hamid for their patience and willingness to impart their knowledge and to share their experience with me. I also would like to thanks to my friends, Mr. Tiang Tow Leong, Mr. Hafiz Johar, Mr. Ali Ranjbaran, Mr. Lee Wee Chuen, Mr. Ting Shyue Siong, Mr. Ong Hui Yeh, Miss Loh Yee Koon, Mr. Chan Pui Weng for their willingness to help me and encourage me during my project sessions.

Last but not least, my thanks go to my family members who offered me unlimited moral support. Without them, I can never complete my project successfully. A word also goes for those who I inevitably forgot to mention. I am no less grateful to them, and I hope that they will forgive my oversight.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ii

TABLE OF CONTENTS iv

LIST OF TABLES ix

LIST OF FIGURES x

ABSTRAK xiii

ABSTRACT xv

CHAPTER 1 INTRODUCTION 1

1.1 Research Background 1

1.2 Problem Statement 3

1.3 Scope of the Research 4

1.4 Objectives of the Research 5

1.5 Outline of the Thesis 5

CHAPTER 2 LITERATURE REVIEW 7

2.1 Introduction 7

2.2 Stereo Vision System 8

2.3 Stereo Vision in Robotic Applications 8

2.3.1 Review of Other Vision Technique in Robotic Applications ... 9

2.4 Comparison of Camera Calibration Methods 10

2.5 Stereo Matching 12

2.6 Image Segmentation 13

2.7 Morphological Operations 14

2.8 Summary 15

CHAPTER 3 THEORETICAL BACKGROUND 16

3.1 Introduction 16

3.2 Camera Calibration and Modelling 16

3.2.1 Camera Model 17

3.2.1.1 Camera Position and Orientation 18

3.2.1.2 Perspective Projection 19

3.2.1.3 Lens Distortion 19

3.2.1.4 Computer Image Frame 20

3.2.2 The Method of Tsai 20

3.3 Rectified Configuration of Two Cameras 27

3.3.1 3D Point Calculation based on Triangulation Method 28

3.4 Edge-based Segmentation 30

3.4.1 Edge Image Thresholding 31

3.5 Binary Morphology 31

3.5.1 Dilation 32

3.5.2 Erosion 32

3.5.3 Opening 32

3.5.4 Closing 32

3.6 Manipulator Modeling 33

3.6.1 Denavit-Hartenberg Convention 33

3.6.2 Kinematic Modeling 34

3.6.3 Dynamic Modeling 35

3.7 Summary 36

CHAPTER 4 METHODOLOGY 37

4.1 Introduction 37

4.2 Hardware Overview 37

4.3 System Implementation 39

4.4 Camera Calibration Method 42

4.5 Stereo Acquisition 45

4.6 Image Rectification Process 45

4.7 Stereo Matching Algorithm 46

4.7.1 Proposed Image Segmentation Technique 47

4.7.2 Centroid Coordinate 49

4.7.3 Disparity 49

4.8 3D Coordinate Computation 49

4.8.1 Curve Fitting Tool 50

4.8.2 Triangulation Formula 51

4.9 Proposed Inverse Kinematic Calculation 51

4.10 Robot Movement 55

4.11 Summary 55

CHAPTER 5 EXPERIMENTAL RESULTS AND DISCUSSIONS 56

5.1 Introduction 56

5.2 Camera Calibration Parameters 56

5.3 Image Processing 58

5.3.1 Converted Grayscale Images 58

5.3.2 Image Rectification 64

5.3.3 Image Segmentation 69

5.4 3D Position Measurement 77

5.5 Robot Angle 86

5.6 Implementation 87

CHAPTER 6 CONCLUSIONS AND RECOMMENDATION 89

6.1 Conclusions 89

6.2 Recommendation 90

REFERENCES 91

APPENDICES 98

APPENDIX A – Photo of Hardware 98

APPENDIX B – Image Processing Results for Other Shape of the Object 99

LIST OF PUBLICATIONS 101

LIST OF TABLES

		Page
Table 5.1	Intrinsic parameters of the left and right cameras	57
Table 5.2	Extrinsic parameters of the left and right cameras	57
Table 5.3	Summary of the Y position values for all the grayscale results	63
Table 5.4	Summary of the Y position values for all the rectification results	68
Table 5.5	Pixel to centimeter (cm) conversions for *x* coordinates	77
Table 5.6	Pixel to centimeter (cm) conversions for *y* coordinates	78
Table 5.7	Disparity to distance conversions	79
Table 5.8	Position measurement	84
Table 5.9	Newly computed results using the compensation technique	85
Table 5.10	Computed angle for motor movement	87

LIST OF FIGURES

		Page
Figure 3.1	The geometric relation between a 3D object point and its 2D image projection	17
Figure 3.2	Illustration of the radial alignment constraint	22
Figure 3.3	Flowchart of the Tsai method	26
Figure 3.4	The rectified stereo configuration where the epipolar lines are parallel in the image, and epipoles move to infinity	27
Figure 3.5	Elementary stereo geometry in the rectified configuration. The depth z of the point X in the 3D scene can be calculated from the disparity $d = u\text{-}u'$. Values of u, u' are measured at the same height, i.e , $v = v'$	29
Figure 3.6	The direct and inverse kinematic models	34
Figure 3.7	Schematic overview of the RRR-manipulator	36
Figure 4.1	Interfacing of the hardware components	38
Figure 4.2	Layout of the experimental set-up	39
Figure 4.3	System implementation	41
Figure 4.4	Left images of checkerboard (with the area of a square 8.41 cm^2)	43
Figure 4.5	Right images of checkerboard (with the area of a square 8.41 cm^2)	43
Figure 4.6	Camera calibration process	44
Figure 4.7	Image rectification process	46
Figure 4.8	Image segmentation process	48
Figure 4.9	First joint of the robotic arm	52
Figure 4.10	Illustration showing the second and third joint robotic arms with the two angles, θ_2 and	52
Figure 5.1	Location of the stereo camera with respect to the coordinate system	58
Figure 5.2	The results of grayscale images at the distance of 47.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G	60

Figure 5.3	The results of grayscale images at the distance of 45.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G	61
Figure 5.4	The results of grayscale images at the distance of 43.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G	62
Figure 5.5	The results of rectified images at the distance of 47.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G	65
Figure 5.6	The results of rectified images at the distance of 45.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G	66
Figure 5.7	The results of rectified images at the distance of 43.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G	67
Figure 5.8	Segmentation results for left images at the distance of 47.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G based on (i) Edge Thresholding, (ii) Morphology Operation, and (iii) Largest Area Selection	71
Figure 5.9	Segmentation results for right images at the distance of 47.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G based on (i) Edge Thresholding, (ii) Morphology Operation, and (iii) Largest Area Selection	72
Figure 5.10	Segmentation results for left images at the distance of 45.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G based on (i) Edge Thresholding, (ii) Morphology Operation, and (iii) Largest Area Selection	73
Figure 5.11	Segmentation results for right images at the distance of 45.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G based on (i) Edge Thresholding, (ii) Morphology Operation, and (iii) Largest Area Selection	74
Figure 5.12	Segmentation results for left images at the distance of 43.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G based on (i) Edge Thresholding, (ii) Morphology Operation, and (iii) Largest Area Selection	75
Figure 5.13	Segmentation results for right images at the distance of 43.9 cm for (a) point A, (b) point B, (c) point F, and (d) point G based on (i) Edge Thresholding, (ii) Morphology Operation, and (iii) Largest Area Selection	76
Figure 5.14	Absolute errors obtained for x coordinates based on the curve fitting and triangulation methods	81
Figure 5.15	Absolute errors obtained for y coordinates based on the curve fitting and triangulation methods	81

Figure 5.16	Absolute errors obtained for *z* coordinates based on the curve fitting and triangulation methods	82
Figure 5.17	Image description of the compensation technique	84
Figure 5.18	Robot movement to pick up an object automatically at position A, B and G at the distance of 47.9 cm with (i) Initial condition, (ii)Robot move toward to an object position and (iii) Robot successfully pick up an object.	88

PENGLIHATAN STEREO MENGGUNAKAN KAEDAH PAMPASAN UNTUK PENAMBAHBAIKAN KEJITUAN JARAK BAGI APLIKASI LENGAN ROBOT

ABSTRAK

Sistem penglihatan dua dimensi (2D) telah digunakan secara meluas dalam bidang automasi untuk meningkatkan produktiviti. Walau bagaimanapun, sistem penglihatan 2D tidak mampu untuk menentukan kedalaman objek. Oleh itu, penyelesaian yang paling sesuai kepada masalah ini adalah mengaplikasikan sistem penglihatan tiga dimensi (3D) dalam bidang automasi. Penglihatan stereo adalah antara cara yang paling berkesan untuk merealisasikan penglihatan 3D. Penyelidikan ini menggunakan sepasang kamera sebagai modul penglihatan stereo untuk mencari lokasi objek. Sistem penglihatan stereo yang melibatkan penentukuran kamera, pembetulan imej, pensegmenan imej, pengiraan sentroid dan ketempatan objek telah dijalankan. Proses pensegmenan imej terdiri daripada gabungan pengesanan pinggir, morfologi matematik dan pemilihan luas objek yang terbesar. Koordinat *x* dan *y* dikira dengan berasaskan koordinat sentroid objek dalam imej kiri dan koordinat *z* dikira berasaskan nilai ketidaksamaan yang diperolehi. Semua imej skala kelabu kiri dan kanan telah diselaraskan secara mendatar dengan berjayanya melalui proses pembetulan imej. Ralat ketidaksamaan telah dikurangkan selepas proses pensegmenan bayangan. Purata dan sisihan piawai ralat bagi koordinat *x* adalah -0.14 dan 0.19 cm. Purata ralat bagi koordinat *y* adalah -0.1 cm dan sisihan piawai ralat bagi koordinat *y* adalah 0.48 cm. Purata dan sisihan piawai ralat untuk koordinat *z* pada mulanya 0.37 dan 0.42 cm, telah dikurangkan kepada -0.12 dan 0.16 cm dengan

menggunakan teknik pampasan. Lengan robot dengan kebebasan 3-darjah mampu mengutip bola menggunakan lokasi tiga dimensi yang diperolehi daripada sistem penglihatan stereo.

STEREO VISION USING COMPENSATION TECHNIQUE FOR IMPROVEMENT OF DISTANCE MEASUREMENT ACCURACY FOR ROBOT ARM APPLICATIONS

ABSTRACT

Two-Dimensional (2D) vision systems have been widely applied to various automation fields in order to increase productivity. However, these 2D vision systems are not able to retrieve the third dimension (depth) of the object. Therefore, the most promising solution to this problem is to apply the Three-Dimensional (3D) vision system to the automation fields. Among the 3D vision range finding method, stereo vision is most effective way of realizing the 3D vision. This research uses a pair of cameras as a stereo vision module to locate the object. The stereo vision system, which consists of camera calibration, image rectification, image segmentation, centroid computation and object localization is implemented. Image segmentation is performed through a process of edge-based segmentation, mathematical morphology and largest area selection. The *x*, *y* coordinates are given by the centroid coordinates of the object in the left image and the *z* coordinates are computed based on the obtained disparity value. The grayscale left and right images are successfully aligned horizontally through the image rectification process. The disparity error is minimized after the image segmentation process. The *x* coordinates result for mean and standard deviation of error are -0.14 cm and 0.19cm, respectively. The mean of error for *y* coordinates is -0.10 cm and the standard deviation of error for *y* coordinates is 0.48 cm. The initial mean and standard deviation of error for *z* coordinates are 0.37 cm and 0.42 cm, respectively. However, after compensation

technique is performed, the mean and standard deviation of error for the *z* coordinates are reduced to -0.12 cm and 0.16 cm, respectively. The 3-degree of freedom robot arm can pick up a ball with the 3D location provided by the stereo vision system.

CHAPTER 1

INTRODUCTION

1.1 Research Background

Industrial robots are powerful, rapid, and precise manipulators which have been widely used in production since 1980s. Industrial robots can work repetitively for extended time periods, and thus improve productivity and stabilize product quality (Borangiu & Dumitrache, 2010). However, the industrial robot used for assembly operation can only handle work-pieces placed in preprogrammed locations and orientation. Automation cost increases because these devices require expensive work-piece feeding conveyor systems and frequent maintenance (Borangiu & Dumitrache, 2010; Hema, Paulraj, Ramachandran, & Sazali, 2007; Oh & Lee, 2007).

Applying the vision system to the automation field can solve this problem and make industrial robots flexible in performing tasks. The vision system is the best sensing device for positioning techniques in industrial applications (Poplawski & Sultan, 2007). The system provides a robot with workspace information useful for guiding and planning robot motion (Leu & Pherwani, 1989). Two-dimensional (2D) vision systems have been widely applied in industrial robot production for improved performance. The 2D systems enable robots to perform tasks efficiently by enabling the robots to recognize workspace deviations. However, 2D vision systems are insufficient because they only measure the work-piece *x, y* locations. Distance measures from the cameras to the work-pieces cannot be recovered by these 2D vision systems. Thus, there is a strong demand from manufacturers for a robust and accurate three-dimensional (3D) robot arm guidance system for handling applications

(Oh & Lee, 2007).

A similar problem to this situation is auto racking, where robots themselves feed the parts on the work cells. In auto racking, the parts are presented to the robot one at a time in different locations, depending on several factors such as the rack used to ship the part, among others. Unlike its predecessor, pick and place, auto racking requires the vision system to locate the singulated part for the robot in 3D space, rather than on 2D space such as a flat conveyor or tray (Hardin, 2008) .

An important type of 3D systems is range finding, that is, depth measurement. Vision-based range finding methods can be divided into two categories: active and passive. Active range finding emit some kind of energy toward the object, and then receive the reflected energy. The active range finding uses the ultrasonic wave, radio wave or microwave, visible light, infrared wave, as well as other form of waves to detect the range between the wave source and the object. Implementation of this approach includes triangulation and time-of-flight systems (Ballard & Brown, 1982; Idesawa, Yataeai, & Soma, 1977; Kanade, 1987; Leu & Pherwani, 1989; Nitzan, Barrouil, Cheeseman, & Smith, 1983; Rioux, Bechthold, Taylor, & Duggan, 1987; Will & Pennington, 1971). Laser-scanning-based vision systems can generate extremely accurate 3D surface maps of objects for robot guidance and inspection applications depending on the quality of the components used (Hardin, 2008). On the other hand, passive range finding only receive the energy from the objects. Occlusion cues, texture gradient, focusing, or stereo disparity determines the range (Horn, 1970; Leu & Pherwani, 1989; Panton, 1978; R & L, 1976; Rosenburg, Levine, & Zucker, 1978). With sub-millimeter and better accuracy obtained over large areas, machine

vision systems using the passive range finding method can also be highly precise (Hardin, 2008).

Stereo vision is one of the methods that can retrieve the depth information. It is an imaging technique for recover depth from camera images by comparing two views of the same scene (Brown, Burschka, & Hager, 2003; Chen & Lingling; Ganapathy & Oon-Ee; Li, Yan, & Liu). Stereo Vision is most effective way of realizing 3D vision and it has one advantage over other methods from extracting the relative depth information because it is passive and accurate (Chen, Zuo, & Zheng, 2009; Woodlill, Buck, Jurasek, Gordon, & Brown, 2007). This method is widely used in industrial automation and 3D machine vision applications for tasks such as bin picking, volume measurement, and 3D object location and identification (Bahadori, Iocchi, Leone, Nardi, & Scozzafava, 2007; Bertozzi & Broggi, 1998; Brown, et al., 2003; Li-Wei, Yuan-Hsiang, & Zhen-Zhong, 2010; Point Grey Research, 2010). Stereo vision technology has endowed robots with more flexible position control capabilities in tracking and grasping unknown objects in arbitrary locations (Wells, Venaille, & Torras, 1996).

1.2 Problem Statement

2D vision systems have been widely applied to various automation fields in order to increase productivity. However, these 2D vision systems can measure only x, y positions and rotation of products. In case the distance between camera and work-pieces varies arbitrary, 2D vision system is inadequate (Oh & Lee, 2007). The most promising solution to this problem is to apply the 3D vision system to the automation fields. The 3D vision that will be applied in this research is stereo vision system.

Stereo vision is the most effective way of realizing the 3D vision (Woodlill, et al., 2007).

The main task in stereo vision is to identify which pixels in the left image corresponds to the pixels in right image, which is used to estimate disparities (Calin, 2007). The disparity is negatively correlated with the distance from the cameras. The operation of finding the correct corresponding point of stereo images is found to be difficult due to some point points in one image may not appear in another image, or the size and spatial relationship may be different in the two images owing to the perspective distortion (Kho, 2009; Kuhl, 2005).

1.3 Scope of the Research

This research generally covers two major components. The first part of the study focuses on stereo vision. Stereo cameras will be used as an input to the stereo vision system. The captured images will then be processed using tools provided by Matlab® software. The *x*, *y* coordinates are given by the centroid coordinate of the object in the left image, and the *z* coordinates are computed based on the obtained disparity value. The computed object location will then be compared with the actual object location to verify measurement accuracy.

The second part of this research is the movement of the robot arm. Five servomotors and Bioloid® conjunction parts will be combined to construct the robot arm, which will then be interfaced with Matlab® through USB2Dynamixel. The finalized 3D object location obtained through the stereo vision system will be used to move the robot arm.

In this research, it is assumed that there is no movement of the object between the image capturing step and the robot movement to grasp the object. The work concentrates on the kinematic and stereo vision only, so the dynamic analysis is not a part of the research work.

1.4 Objectives of the Research

The objectives of this project are listed below:

(a) To develop the compensation technique for improving the accuracy of depth measurement in robot arm guided applications using stereo vision system.

1.5 Outline of the Thesis

This chapter presents the research background and the problem statement. The scope and research objectives are also discussed.

Chapter 2 presents the literature review. The development of the vision system is explained and their applications to various fields are reviewed.

Chapter 3 presents the theoretical background of this research. The detailed analyses of stereo vision system and robotic manipulator modeling are described.

Chapter 4 shows the methodology of the whole research process including how to obtain the 3D coordinates with the proposed method. A discussion on the inverse kinematic analysis for robot arm movement follows.

Chapter 5 discusses the results obtained from this research including the 3D coordinates and robot angle.

Chapter 6 highlights the conclusions and contributions of this research. Suggestions for future investigations are also provided.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

3D Vision System are divided into active range finding and passive range finding (Quintana, 2003). The main different of these systems or methods are; active range finding emit the energy toward the object, and then receive the reflected energy but the passive range finding only receive the energy from the objects (Dana H. Ballard, 1982; Kanade, 1987; Rioux, et al., 1987). For the active range finding methods, although the obtained range data is more accurate, but they dissipate more energy, and even some emitting wave is very harmful to human being. In military applications, the emitting wave can be detected by the opponent. The passive range finding methods are widely applied by the researchers (Faugeras & Toscani, 1986; Takahashi & Tomita, 1988) .The passive method is determined based on occlusion cues, texture gradient, focusing, or stereo disparity (Horn, 1970; M. Leu & R. Pherwani, 1989; Panton, 1978; Bajcsy & Lieberman, 1976; Rosenburg, et al., 1978). Stereo vision system has major one advantage over other methods from extracting the relative depth information because it is more accurate (Chen, et al., 2009; Woodlill, et al., 2007).

The following section presents a literature review on the vision in robotic applications. In addition, a few topics related to stereo vision are reviewed.

2.2 Stereo Vision System

Stereo vision is one of the best methods for retrieving depth information. This method has inspired numerous research into computer vision systems with two inputs exploiting the knowledge of their own relative geometry to derive depth information from the two views they receive (Gui & Tu, 2011; Hartley & Zisserman, 2000; Longfield & Chang, 2009; Sonka, Hlavac, & Boyle, 2008; Zhao & Zhao, 2009).

The most important parameters in stereo vision are the disparity of point pair and distance (depth) to the object. Disparity is the difference in object image locations as captured by stereo cameras. Once the disparity values are known, the world coordinates of a point can be determined using the triangulation method (Bleyer & Gelautz, 2005a, 2005b; Lee, Bae, & Kim, 2002). However, triangulation method measurement becomes less accurate if the camera is not well calibrated. Therefore, a better way to measure distance is using curve fitting method, which determines the appropriate equation for distance measurement (Abidin, 2007).

Disparity is inversely proportional to depth. Unfortunately, finding the correct disparity value for each image point is difficult due to image noise, untextured regions, and occlusions. Therefore, stereo matching remains a popular research subject (Okutomi, Katayama, & Oka, 2002).

2.3 Stereo Vision in Robotic Applications

Stereo vision has wide applications in industrial automation and 3D machine vision tasks such as bin picking, volume measurement, and 3D object location and identification (Bahadori, et al., 2007; Brown, et al., 2003; Li-Wei, et al.; Se & Jasiobedzki, 2008).

A stereo vision based on the automated bin picking system, which identifies the topmost object from a pile of occluded objects and computes its location is developed in (Hema, et al., 2007). The system focuses mainly on solving the segmentation and localization problems of partially occluded objects. Moreover, in (Oh & Lee, 2007), a general-purpose stereo vision system is presented which is easily applicable to various tasks with industrial robots on the basis of wide experiences on the 2D vision system for many automotive plants.

Additionally, the work for humanoid service robot HARO-I through vision guided homing which has active stereo vision and two modular arms is presented in (Jin & Xie, 2000). The work is based on the 3D vision reconstruction and geometric kinematics analysis. Besides that, a stereo vision system for autonomous vehicle navigation is presented in (Hassan, Hamzah, & Johar, 2010). In the research, area-based algorithm is selected to compute the disparity values. Then curve fitting tool is use for estimating the distance of each obstacle based on the obtained disparity values.

2.3.1 Review of Other Vision Technique in Robotic Applications

A technique to reconstruct the shape of the object and control its movement has been applied in some robotic applications (Yamashita, Kaneko, Matsushita, Miura, & Isogai, 2003). However, the technique in the system just measure the object's location, not reconstruct it. Besides that, the visual perception programming robot-arm system for carrying out flexible pick-and-place behavior has been developed as in (Sbnchez & Martinez, 2000). The system used an ultrasonic sensor to measure the depth of the object. However, the ultrasonic distance measurement is less accurate because the reading may get distorted due to shapes of solids to be

measured as reflection of ultrasonic wave is mostly dependent on shape of solid object which is treated as obstruction.

Furthermore, in Rosenberg, Levine, & Zucker (1978), a technique for computing the relative relationship of "in-front-of," "behind," and "equidistant" using heuristic evidence of occlusion in monocular color imagery is presented. The restraint of correct segmentation of the scene in the first place is the weakness of this approach which leads to incompatibility in robotic arm applications. Moreover, one of the passive methods is measuring texture gradients in the domain of natural outdoor scenes based on Fourier descriptors which are claimed to vary in a manner consistent with surface geometries in three dimensions (Bajcsy & Lieberman, 1976). However, the application of texture gradients approach is restricted to highly textured scenes.

2.4 Comparison of Camera Calibration Methods

Camera calibration process determining and relating the 3D position and orientation of the camera frame to a certain world coordinate. The purposes of this calibration process is for inferring 3D information from computer image coordinates or vice versa (Zhang & Chen, 2008).

The camera calibration technique proposed by Hall (1982) is based on an implicit linear camera calibration through 3x4 transformation matrix calculation, which relates the 3D point of the scene with their 2D image projecting points on the image plane (Hall, Tio, McPherson, & Sadjadi, 1982). The linear camera calibration technique proposed by Faugeras and Toscani (1986), extract the physical parameters of the camera from the 3x2 symbolic transformation matrix by equating it to the numeric matrix obtained by calibrating the camera with the technique of Hall.

The calibration techniques mentioned before use the least squares method to obtain a transformation matrix relating 3D points with their 2D projections. These techniques provide the advantage of a simple and rapid calibration. However, such techniques entail rough accuracy and cannot be used for lens distortion modeling. Moreover, extracting the parameters from the matrix is at times difficult because of the implicit calibration used (Quintana, 2003).

The linear method of Faugeras (1986) can be improved by including the radial lens distortion. Thus, a simple adaptation of the Faugeras linear method including radial lens distortion is implemented (Salvi, 1997; Salvi, Batlle, & Mouaddib, 1998).

When lens imperfection is included in the camera model, the calibrating technique becomes nonlinear. The camera parameters are then usually obtained by iteration with the constraint of minimizing a determined function, which is typically the distance between the imaged points and the modeled projections obtained by iteration. These iterating techniques provide the advantage of making almost any model calibrated. Moreover, increasing the iteration number up to convergence normally improves accuracy. However, these techniques require a good initial guess to guarantee convergence (Quintana, 2003).

The most popular calibration technique proposed by Tsai (1987) is based on a two-step technique which model only radial lens distortion. This camera calibration technique can be used with any camera. This calibration technique also avoids the problems encountered by other calibration methods such as large amounts of unknown per plane, the assumption that lens distortion is negligible, and known focal length (Chen, et al., 2009).

Two step techniques use linear optimization to first compute several parameters and then iteratively calculate for the remaining parameters. These techniques permit a rapid calibration; it reduces the number of iterations considerably. Moreover, convergence is nearly guaranteed because of the linear estimate obtained in the first step. Two-step techniques utilize the advantages of both linear and nonlinear camera calibration techniques (Quintana, 2003).

2.5 Stereo Matching

The main task of the stereo computation is to identify which pixel in the first image corresponds to the second image. This is known as the *Correspondence Problem*. The algorithms can roughly be divided into feature-based and area-based (Chowdhury & Bhuiyan, 2009; Henkel, 1997; Nasrabadi & Yi, 1989; Sonka, et al., 2008).

Area-based algorithms match windows in one image with windows in another image using cross-correlation. On the other hand, feature-based correspondence algorithms first extract the features to which the matching process are then applied (Kuhl, 2005; Nasrabadi & Yi, 1989; Sonka, et al., 2008).

The advantages of using the feature-matching algorithm over the area-based matching algorithm are as follows: a) feature-based methods are less ambiguous because of the small number of potential candidates for correspondence; b) the photometric image variation does not significantly influence the correspondence result; and c) a higher precision of disparity computation is provided because features can be detected to sub-pixel accuracy (Nasrabadi & Yi, 1989; Sonka, et al., 2008).

The operation of finding the correct corresponding point of stereo images is found to be difficult due to some points in one image may not appear in another image, or the size and spatial relationship may be different in the two images owe to perspective distortion (Kho, 2009; Kuhl, 2005). Therefore, a few matching constraints must be imposed in order to minimize the false matches. Below is a list of the some commonly used constraints (Klette, Koschan, & Schluns, 1996).

1) Similarity constraint: the matching pixels must have similar intensity values for the intensity-based approach, and the matching features must have similar attribute values for the feature-based approach.
2) Uniqueness constraint: a given pixel or feature from one image can match only one pixel or feature from the other image.
3) Continuity constraint: the disparity of the matches should vary smoothly almost everywhere over the image.

False matching can also be minimized through preprocessing steps such as the pre-reduction of noise with a low-pass filter, different illumination adjustments, or adjusting the white balance of each camera. However, camera calibration and using the epipolar constraint are the most effective preprocessing steps (Kuhl, 2005) .

2.6 Image Segmentation

Image segmentation is one of the most important steps leading to the analysis of processed image data. This process divides an image into parts corresponding to real world objects or regions contained in the image. One of the main segmentation problems is image data ambiguity, which is often accompanied by information noise. Segmentation methods can be divided into three groups according to their dominant features. First is the global knowledge about an image or its part; the knowledge is

usually represented by a histogram of image features. The second group is edge-based segmentations, it is one of the earliest segmentation approaches and still remains very important; It represents a large group of methods based on information about image edges. The third group is region-based segmentations, which are generally better for noisy images where borders are extremely difficult to detect. Different characteristics may be used in edge detection or region growing, such as brightness, texture, and velocity field, among others (Sonka, et al., 2008).

The most common problems of edge-based segmentation, caused by image noise in an image, are an edge presence in locations where there is no border, and no edge presence where a real border exists. Clearly, both cases negatively influence segmentation. Therefore, edge-based segmentation should be combined with other operations to minimize noise and improve segmentation.

2.7 Morphological Operations

Mathematical Morphology is a theory and technique for the analysis and processing of geometrical structures based on set theory. The two basic morphological operators are dilation and erosion. Morphological operations are mainly used in following purposes : a) image pre-processing (noise filtering, shape simplification); b) Enhancing object structure (skeletonizing, thinning, thickening, convex hull, object marking); c) Segmenting objects from the background; and d) Quantitative description of objects (area, perimeter, projections, Euler-Poincare characteristic) (Serra, 1982; Sonka, et al., 2008).

2.8 Summary

The stereo vision system is widely studied in the robot arm application. The Tsai camera calibration method is a popular method to determining and relating the 3D position and orientation of the camera frame to a certain world coordinate. The literature indicates that the featured-based stereo matching algorithm is more reliable and accurate compared to area-based matching algorithm in solving the correspondence problem. Edge-based segmentation technique and morphological operation are widely used in stereo computation.

CHAPTER 3

THEORETICAL BACKGROUND

3.1 Introduction

The basic principle of stereo vision usually involves camera calibration and establishing point correspondences between pairs of points from stereo images and 3D coordinates reconstruction. In this chapter, a theoretical background of a related topic to the stereo vision system is explained. This chapter also covers the theoretical aspects of robot manipulator operation.

3.2 Camera Calibration and Modelling

Camera calibration can be divided into two steps. The first is called camera modelling, which deals with the mathematical approximation of the physical and optical behavior of the sensor using a set of parameters. The second step deals with direct or iterative methods that estimate the parameter values. The model has two kinds of parameters: intrinsic and extrinsic. Basically, intrinsic parameters determine how light is projected through the lens onto the sensor image plane. Then the extrinsic parameters measure the camera position and orientation with respect to a world coordinate system. In turn, this coordinate provides metric information with respect to a user-fixed coordinate system instead of the camera coordinate system (Quintana, 2003).

3.2.1 Camera Model

Camera modeling approximates the internal geometry together with the camera position and orientation in the scene. Camera modeling is usually divided into four steps, detailed hereafter (see also Figure 3.1) (Quintana, 2003).

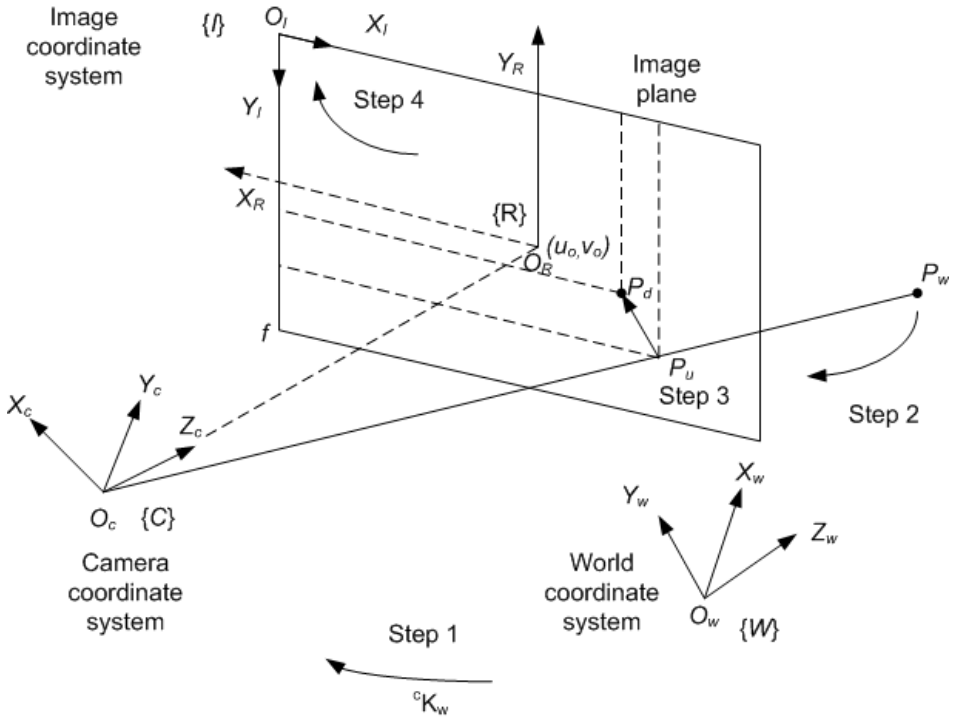

Figure 3.1 The geometric relation between a 3D object point and its 2D image projection (Quintana, 2003)

1) The first step consists of relating a point W_{P_w} from the world coordinate system to the camera coordinate system, obtaining C_{P_w}. A rotation matrix and a translation vector achieve this transformation.
2) Next, point C_{P_w} must be projected on the image plane and point C_{P_u} obtained using a projective transformation.

3) The lens distortion is modeled based on a disparity with the real projection. Then, point C_{P_u} is transformed to the real projection of C_{P_d} (which should coincide with the points captured by the camera).
4) Finally, another coordinate system transformation is needed to change from the camera metric coordinate system to the computer image coordinate system in pixels, obtaining I_{P_d}.

3.2.1.1 Camera Position and Orientation

The world coordinate system transforms to the camera coordinate system. This transformation is modeled using a translation vector and a rotation matrix, as shown in Equation (3.1).

$$C_{\boldsymbol{p}_w} = C_{\boldsymbol{R}_w} W_{P_w} + C_{\boldsymbol{T}_w} \tag{3.1}$$

where

$$C_{\boldsymbol{p}_w} = \begin{pmatrix} C_{X_w} \\ C_{Y_w} \\ C_{Z_w} \end{pmatrix} \text{ and } W_{P_w} = \begin{pmatrix} W_{X_w} \\ W_{Y_w} \\ W_{Z_w} \end{pmatrix}$$

Given a point W_{P_w} related to the world coordinate system and applying Equation (3.1), the point C_{P_w} in relation to the camera coordinate system is then obtained. Note that $C_{\boldsymbol{R}_w}$ expresses the orientation of the world coordinate system with respect to the axis of the camera coordinate system {C}. Moreover, C_{T_w} expresses the original position of the world coordinate system measured with respect to {C}(Quintana, 2003).

3.2.1.2 Perspective Projection

Consider that any optical sensor can be modeled as a pinhole camera (O. Faugeras, 1993), that is, the image plane is located at a distance *f* from the optical center O_C, and is parallel to the plane defined by the coordinate axes X_C and Y_C. Moreover, if an object point (C_{P_w}) related to the camera coordinate system is projected through the focal point (O_C), the optical ray intercepts the image plane at the 2D image point (C_{P_u}). Equation (3.2) shows this relation (Quintana, 2003).

$$C_{X_u} = f \frac{C_{X_w}}{C_{Z_w}} \qquad C_{Y_u} = f \frac{C_{Y_w}}{C_{Z_w}} \tag{3.2}$$

3.2.1.3 Lens Distortion

The third step models the lens distortion. Equations (3.3) transform the undistorted point C_{P_u} to the distorted point C_{P_d}, where δ_x and δ_y represent the involved distortion (Quintana, 2003).

$$C_{X_u} = C_{X_d} + \delta_x \qquad C_{Y_u} = C_{Y_d} + \delta_y \tag{3.3}$$

Tsai (1987) modeled the lens distortion in the same way but only considered radial distortion, as shown in Equations (3.4), in which δ_x and δ_y represent the radial distortion (Slama, Theurer, & Henriksen, 1980).

$$\delta_x = \delta_{xr} \qquad \delta_y = \delta_{yr} \tag{3.4}$$

The displacement given by the radial distortion (*dr*) can be expressed by Equations (3.5) considering only k_1, the first term of the radial distortion series proven to be sufficient for modeling radial distortion in most applications (Tsai, 1987).

$$\delta_{xr} = k_1 C_{X_d}(C_{X_d}^2 + C_{Y_d}^2) \qquad \delta_{yr} = k_1 C_{Y_d}(C_{X_d}^2 + C_{Y_d}^2) \tag{3.5}$$

3.2.1.4 Computer Image Frame

This final step expresses the C_{P_d} point in relation to the computer image plane in pixels {I}. This coordinate change can be made according to Tsai's camera model, which uses the following equations for the transformation:

$$I_{X_d} = -s_x d'_x{}^{-1} C_{X_d} + u_0 \qquad I_{Y_d} = -d_y{}^{-1} C_{Y_d} + v_0 \tag{3.6}$$

where (u_0, v_0) are the principal point components in pixels; s_x is the image scale factor; $d'_x = d_x \frac{N_{cx}}{N_{fx}}$; d_x is the center to center distance between adjacent sensor elements in the X direction; d_y is the center to center distance between adjacent sensor elements in the Y direction; N_{cx} is the number of sensor elements in the X direction; and N_{fx} is the number of pixels in an image row sampled by the computer (Quintana, 2003).

3.2.2 The Method of Tsai

The method of Tsai (Tsai, 1987) modeled the radial lens distortion but assumed several camera parameters, which are provided by manufacturers. This fact reduces calibrating parameters in the first step for the initial estimate. Moreover, despite the iteratively optimized parameters in the last step, the iteration number is considerably reduced using the calibration algorithm proposed by Tsai.

First, combining Equations (3.1), (3.2), (3.3), (3.4), and (3.5) obtains Equations (3.7).

$$C_{X_u} = C_{X_d} + C_{X_d k_1 r^2} = f \frac{r_{11}{}^W X_w + r_{12}{}^W Y_w + r_{13}{}^W Z_w + t_x}{r_{31}{}^W X_w + r_{32}{}^W Y_w + r_{33}{}^W Z_w + t_z}$$

$$C_{Y_u} = C_{Y_d} + C_{Y_d} k_1 r^2 = f \frac{r_{21}{}^W X_w + r_{22}{}^W Y_w + r_{23}{}^W Z_w + t_y}{r_{31}{}^W X_w + r_{32}{}^W Y_w + r_{33}{}^W Z_w + t_z}$$

$$r = \sqrt{C_{X_d}{}^2 + C_{Y_d}{}^2} \tag{3.7}$$

Once obtained in metric coordinates using Equation (3.6), $C_{X'_d}$ and $C_{Y'_d}$ can be expressed in pixels (I_{X_d} and I_{Y_d}, respectively), hence obtaining the following equations.

$$C_{X'_{di}} = -(I_{X_{di}} - u_0) d'_x \qquad C_{Y'_{di}} = -(I_{Y_{di}} - v_0) d_y \tag{3.8}$$

where,

$$C_{X'_{di}} = C_{X_{di}} s_x \qquad C_{Y'_{di}} = C_{Y_{di}} \tag{3.9}$$

Therefore, relating image point P_d (in metric coordinates) with object point P_w is essential. Figure 3.2 shows the radial distortion effects on the camera model. The segment $\overline{O_R P_d}$ is parallel to the segment $\overline{P_{oz} P_w}$. Considering this constraint establishes the following relationship:

$$\overline{O_R P_d} \,//\, \overline{P_{oz} P_w} \;\Rightarrow\; \overline{O_R P_d} \times \overline{P_{oz} P_w} = 0 \tag{3.10}$$

Using Equation (3.10), the following equations are obtained.

$$\overline{O_R P_d} \times \overline{P_{oz} P_w} = 0 \tag{3.11}$$

$$\left(C_{X_d}, C_{Y_d}\right) \times \left(C_{X_w}, C_{Y_w}\right) = 0 \tag{3.12}$$

$$C_{X_d} C_{Y_w} - C_{Y_d} C_{X_w} = 0 \tag{3.13}$$

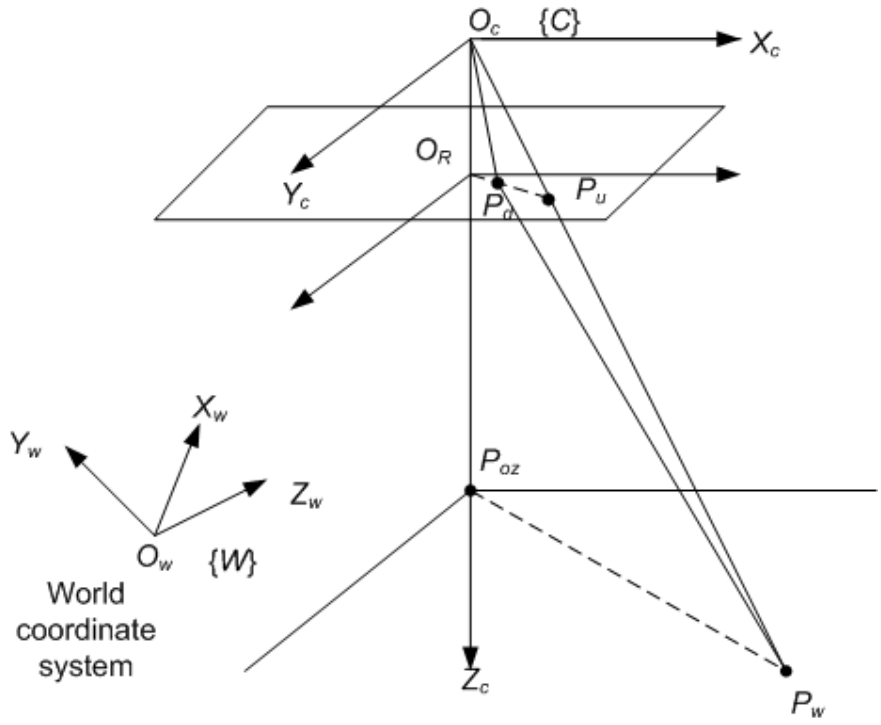

Figure 3.2 Illustration of the radial alignment constraint (Tsai, 1987)

Equation (3.13) can be arranged expressing the object point P_w, in relation to the world coordinate system instead of the camera coordinate system.

$$C_{X_d}(r_{21}{}^{W}X_w + r_{22}{}^{W}Y_w + r_{23}{}^{W}Z_w + t_y)$$

$$= C_{Y_d}(r_{11}{}^{W}X_w + r_{12}{}^{W}Y_w + r_{13}{}^{W}Z_w + t_x) \quad (3.14)$$

Using Equation (3.14) and arranging the terms,

$$C_{X_d} = C_{Y_d}W_{X_w}\frac{r_{11}}{t_y} + C_{Y_d}W_{Y_w}\frac{r_{12}}{t_y} + C_{Y_d}W_{Z_w}\frac{r_{13}}{t_y} + C_{Y_d}\frac{t_x}{t_y}$$

$$- C_{X_d}W_{X_w}\frac{r_{21}}{t_y} - C_{X_d}W_{Y_w}\frac{r_{22}}{t_y} + C_{X_d}W_{Z_w}\frac{r_{23}}{t_y} \quad (3.15)$$

Equation (3.15) can be computed for the n points obtained from Equations (3.8) by combining Equation (3.15) with Equations (3.9), obtaining,

$$C_{X'_{di}} = C_{Y'_{di}} W_{X_{wi}} \frac{s_x r_{11}}{t_y} + C_{Y'_{di}} W_{Y_{wi}} \frac{s_x r_{12}}{t_y} + C_{Y'_{di}} W_{Z_{wi}} \frac{s_x r_{13}}{t_y}$$

$$+ C_{Y'_{di}} \frac{s_x t_x}{t_y} - C_{X'_{di}} W_{X_{wi}} \frac{r_{21}}{t_y} - C_{X'_{di}} W_{Y_{wi}} \frac{r_{22}}{t_y} + C_{X'_{di}} W_{Z_{wi}} \frac{r_{23}}{t_y} \quad (3.16)$$

At this point, a system with n equations and seven unknowns is obtained. The system can be expressed in the following form:

$$\begin{pmatrix} C_{Y'_{di}} W_{X_{wi}} \\ C_{Y'_{di}} W_{Y_{wi}} \\ C_{Y'_{di}} W_{Z_{wi}} \\ C_{Y'_{di}} \\ C_{X'_{di}} W_{X_{wi}} \\ C_{X'_{di}} W_{Y_{wi}} \\ C_{X'_{di}} W_{Z_{wi}} \end{pmatrix}^T \begin{pmatrix} t_y^{-1} s_x r_{11} \\ t_y^{-1} s_x r_{12} \\ t_y^{-1} s_x r_{13} \\ t_y^{-1} s_x t_x \\ t_y^{-1} r_{21} \\ t_y^{-1} r_{22} \\ t_y^{-1} r_{23} \end{pmatrix} = C_{X'_{di}} \quad (3.17)$$

The seven unknown components of the vector are renamed to simplify the notation.

$$a_1 = t_y^{-1} s_x r_{11} \quad a_2 = t_y^{-1} s_x r_{12} \quad a_3 = t_y^{-1} s_x r_{13}$$

$$a_4 = t_y^{-1} s_x t_x \quad a_5 = t_y^{-1} r_{21} \quad a_6 = t_y^{-1} r_{22} \quad a_7 = t_y^{-1} r_{23} \quad (3.18)$$

Note that the a_i components can be easily computed using a least-squares technique. Therefore, the point of interest is to extract the camera calibrating parameters from these a_i components. First, t_y can be obtained using Equations (3.18) as follows:

$$t_y = \frac{\|r_2\|}{\|a_{5,6,7}\|} \quad (3.19)$$

Equation (3.19) is also simplified because the norm of vector r_2 is equal to the unity, obtaining the parameter t_y (Quintana, 2003).

$$|t_y| = \frac{1}{\sqrt{{a_5}^2 + {a_6}^2 + {a_7}^2}} \tag{3.20}$$

However, Equation (3.20) does not provide the sign of the t_y component and is thus insufficient. This sign is determined by taking a point (I_{X_d}, I_{Y_d}) located at the periphery of the image, far from the center, from the set of test points (its corresponding 3D point is maintained). Then the t_y sign is assumed to be positive, and the following equations are computed.

$$r_{11} = a_1 t_y / s_x \quad r_{12} = a_2 t_y / s_x \quad r_{13} = a_3 t_y / s_x \quad t_x = a_4 t_y$$

$$r_{21} = a_5 t_y \quad r_{22} = a_6 t_y \quad r_{23} = a_7 t_y \tag{3.21}$$

Using the corresponding 3D points ($W_{X_w}, W_{Y_w}, W_{Z_w}$), the 3D point on the image plane (without considering lens distortion) has a linear projection expressed using Equations (3.22).

$$C_{X_u} = r_{11} W_{X_W} + r_{12} W_{Y_W} + r_{13} W_{Z_w} + t_x$$

$$C_{Y_u} = r_{21} W_{X_W} + r_{22} W_{Y_W} + r_{23} W_{Z_w} + t_y \tag{3.22}$$

The t_y sign can be verified at this point. If both components of point (C_{X_u}, C_{Y_u}) have a sign equal to the components of point (I_{X_d}, I_{Y_d}), then the t_y sign is correctly assumed as positive; otherwise, consider the sign negative. The scale factor is the second parameter to be extracted (s_x). Note that arranging Equations (3.18) obtains the following equation:

$$s_x = \frac{\|a_{1,2,3}\| t_y}{\|r_1\|} \tag{3.23}$$

where the norm of r_1 is known as the unity, and the scale factor is always positive. Then, s_x is obtained using Equation (3.24).

$$s_x = \sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2}\,|t_y| \tag{3.24}$$

Furthermore, in relation to the camera coordinate system (C_{X_d}, C_{Y_d}), the 2D points can be computed from the same point with respect to the image coordinate system, that is, (I_{X_d}, I_{Y_d}), using Equations (3.9). Using Equations (3.21) can also help compute the r_1 and r_2 vectors of the rotation matrix $C_{\boldsymbol{R}_w}$, and the first element of translation vector C_{T_w}, i.e., t_x, can be calculated. Finally, the third orientation vector (r_3) can be computed using a cross product between r_1 and r_2 because of the orthogonality property (note also that the determinant of any rotation matrix is the unity, i.e., C_{R_w}= 1). At this point, the first three steps of Tsai's method are completed, as shown in Figure 3.3 (Quintana, 2003; Chen, et al., 2009).