DEPTH RECOVERY AND PARAMETER ANALYSIS USING SINGLE-LENS PRISM BASED STEREOVISION SYSTEM

KEE WEI LOON

(B.Eng., NATIONAL UNIVERSITY OF SINGAPORE)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2014

DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety.

I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Kee Wei Loon

27 November, 2014

ACKNOWLEDGEMENTS

I wish to express my gratitude and appreciation to my supervisor, A/Prof. Kah Bin Lim for his instructive guidance, insightful comments and constant personal encouragement throughout the course of my Ph.D. study. I benefit a lot from his critiques and comments. It is a great pleasure for me to pursue my graduate study under his supervision.

I gratefully acknowledge the financial support provided by the National University of Singapore (NUS) that make it possible for me to finish this study. My gratitude also goes to Mr. Yee, Mrs. Ooi, Ms. Tshin, and Miss Hamidah for their help on facility support in the laboratory so that my research could be completed smoothly.

For my colleagues: Zhao Mei Jun, Wang Daolei, Qian Bei Bei and Bai Yading, I am thankful for their constructive discussions and valuable advice on my research. It is also a true pleasure for me to meet many nice and wise colleagues in the Control and Mechatronics Laboratory, who made the past four years exciting and the experience worthwhile. I would like to thank the examiners for their reviewing, attending my oral qualification examination and giving many helpful advices for the future research.

Finally, I would like to thank my parents, and sister for their constant love and endless support through my student life. My gratefulness and appreciation cannot be expressed in words.

Table of Contents

DECLARATION I
ACKNOWLEDGEMENTS II
Table of Contents III
SUMMARY VI
LIST OF FIGURES VIII
LIST OF TABLES XIII
LIST OF SYMBOLS XIV
LIST OF ABBREVIATIONS XV
Chapter 1 Introduction 1
1.1 Problem Descriptions 2
1.2 Contributions 4
1.3 Outline of the thesis 6
Chapter 2 Literature review 7
2.1 Stereovision system 7
2.1.1 Conventional two camera system 8
2.1.2 Single-lens stereovision system 9
2.2 Stereo camera calibration 16
2.2.1 Conventional camera calibration methods 18
2.2.2 Virtual camera calibration technique 19
2.3 Stereo correspondence problem 21
2.3.1 Local method 22
2.3.2 Global method 25
2.3.3 Epipolar constraint 29
2.4 Parameter and quantization analysis 31
2.5 Summary 33
Chapter 3 Virtual Camera Calibration 34
3.1 Formation of virtual camera 34
3.2 Virtual camera calibration based on the proposed geometrical approach 37
3.2.1 Computation of the virtual cameras' optical centres 38
3.2.2 Computation of the virtual cameras' orientation 43
3.2.3 Computation of the virtual cameras' focal length 45
3.3 Experimentation and Discussion 46

3.3.1 Conventional calibration method analysis 47
3.3.2 Experimental results of the proposed geometrical approach 50
3.4 Summary 54
Chapter 4 Stereo Correspondence 55
4.1 Background of epipolar geometry constraint 55
4.2 Construction of virtual epipolar lines using geometrical approach 59
4.3 Experimentation and discussion 65
4.4 Summary 74
Chapter 5 Effects of Angle and Position of Bi-Prism 76
5.1 FOV of the Single-Lens Bi-Prism Stereovision System 77
5.2 Predicting the Type of FOV based on the Bi-prism Angle 79
5.2.1 Geometrical Analysis by Ray Tracing 80
5.2.2 Geometrical Analysis of Divergent System 85
5.2.3 Geometrical Analysis of Convergent System 88
5.3 Experiment 89
5.3.1 Experimental Results 91
5.3.2 Discussions 93
5.4 Effect of Translation of Bi-Prism on System's Field-Of-View 95
5.4.1 Effect of Translation In z-Direction 95
5.4.2 Effect of Translation in *x*-Direction 97
5.5 Experimental Results 101
5.6 Summary 103
Chapter 6 Parameter Analysis 105
6.1 Theoretical Analysis 106
6.1.1 Derivation of the Depth Equation 106
6.1.2 Relative Depth Error 110
6.2. Experiments 114
6.2.1 Experiment Results 115
6.2.2 Discussion 116
6.3 Study of Variable Parameters to Reduce Depth Error 118
6.3.1 Variable focal length, f 118
6.3.2 Variable bi-prism angle, $\boldsymbol{\alpha}$ 120
6.3.3 Variable $\boldsymbol{To}$ 124
6.4 Experiments 125

6.5 Summary 129
Chapter 7 Conclusions and Future Work 130
7.1 Contributions of the thesis 130
7.2 Future work 133
List of Publications 136
Bibliography 137
Appendices 151
A. Law of Refraction (Snell's Law) 151
B. Zhang's calibration algorithm 152
C. Mid-point theorem 153
D. Convergent System 154
E. Results of Set-up 3 and 4 155

SUMMARY

This thesis aims to study the depth recovery and parameter analysis of a single-lens bi-prism based stereovision system. The 2D image is captured by this system and can be split into two sub-images on the camera image plane, which are assumed to be captured by two virtual cameras simultaneously. A point in the 3D space would appear in different locations in each of the image planes, and the differences in positions between them are called the *disparities*. The depth information of the point can then be recovered by using the system setup parameters and the disparities. This system offers several advantages over the conventional system which uses two cameras, such as compactness, lower costs and ease in operation. In this research, the concept and formation of the virtual cameras are also introduced and parameters of the system are studied in detailed to improve the accuracy of the depth recovery.

A geometry-based approach has been proposed to calibrate the two virtual cameras generated by the system. The projection transformation matrices or the extrinsic parameters of the virtual cameras are computed by a unique geometrical ray sketching approach. This approach requires no complicated calibration process. Based on the calibrated virtual cameras, a virtual epipolar line approach is presented to solve the correspondence problem of the system. A specially designed experimental setup, with high precision stage was fabricated to conduct experiments. The results show that the proposed approach is effective and robust. By comparing the results of the proposed geometry-based approach to the results of conventional stereovision technique, the former approach produces better results.

Furthermore, the geometrical approach is used to predict the type of field of view (FOV) produced given a bi-prism angle, α. This is done by comparing two essential angles ϕ_2

and ϕ_4 defined during the theoretical development of our approach. The two main types of FOV generated by this system are divergent FOV and convergent FOV. By using the ray sketching approach, the geometry of each type of FOV can be theoretically estimated. Then, the effect of translation of bi-prism in the *z*- and *x*-axes on the system's FOV is determined using geometrical analysis. Experiments are conducted to verify the above predictions. While there are some degree of quantitative error between experimental results and theory, the general theoretical trends are largely supported by the results.

Finally, the parameter/error analysis of the single-lens bi-prism stereovision system in terms of the system parameters is studied in detailed. Theoretical equations are derived to estimate the error and the trend of error when the object distances increase. The relative depth error which is essential to design the system appropriately for practical usage is then formulated. It was found that the performance of the system is better for near range applications as compared to long range applications. Based on the findings, the possibility of manipulating the system parameters, named as variable parameter is then presented in order to reduce or maintain the error of the system for long range applications.

To summarize, the main contribution of this thesis is the development of a novel stereo vision technique. All the efforts are made to recover the depth of a 3D scene using the single-lens bi-prism based stereovision system and to improve the accuracy of the results.

LIST OF FIGURES

Figure 2.1: Modeling of two camera canonical stereovision system.....9

Figure 2.2 A single-lens stereovision system using a glass plate (Nishimoto and Shirai [37]) 10

Figure 2.3 A single-lens stereovision system using three mirrors (Teoh and Zhang [40]) 11

Figure 2.4 A single-lens stereovision system using two mirrors (Gosthasby and Gruver [42]) 11

Figure 2.5: Four stereovision systems using mirrors (a) two planar mirrors; (b) two ellipsoidal mirrors; (c) two hyperboloidal mirrors; (d) two paraboloidal mirrors (Nene and Nayar [44])12

Figure 2.6 Illustration of the bi-prism system proposed by Lee and Kweon [53].....14

Figure 2.7 Single-lens bi-prism stereovision system (Lim and Xiao [16]).....15

Figure 2.8 Virtual camera calibration of tri-prism system (Lim and Xiao [16]).....15

Figure 2.10 Illustrations of the coordinates systems.....17

Figure 2.11 Image captured using the bi-prism stereovision system, two black dots indicate the two unique pixels chosen for virtual camera modeling 20

Figure 2.12 Formation of virtual cameras using a bi-prism (top view).....20

Figure 2.13 Image captured by the system in non-ideal situation 21

Figure 2.14 (a) disparity space image using left-right axes and; (b) another using left-disparity axes 26

Figure 2.15 Definition of the epipolar plane.....30

Figure 2.16 The geometry of converging stereo with the epipolar line (solid) and the collinear scan-lines (dashed) after rectification 30

Figure 2.17 Depth error analysis of conventional stereovision 32

Figure 3.1 3-D schematic diagram of single-lens stereovision using a bi-prism.....35

Figure 3.2 An example of stereo-image pair taken by a CCD camera through a 6.4° bi-prism 35

Figure 3.3: Single-lens bi-prism stereovision system showing the virtual cameras and their FOVs 36

Figure 3.4 Computing the virtual camera's optical centre 40

Figure 3.5 Illustration of the incident and refracted angles 40

Figure 3.6 Coordinate system of frame A 42

Figure 3.7 Geometrical rays through bi-prism (all rays lie on the $XwZw$ plane) 44

Figure 3.8 Derivation of virtual camera focal length, fvc 45

Figure 3.9 System setup used in the experiment 47

Figure 3.10 Calibration board captured can be divided into two sub-images 48

Figure 3.11 Corner extraction of the calibration board for calibration 48

Figure 3.12 Extrinsic rig of the virtual cameras and the orientation of the calibration boards 49

Figure 3.13 Computing the optical centre using all the image points 51

Figure 3.14 Optical centre coordinates computed from all the pixels (512 x 384 pixels) 52

Figure 3.15 *x* coordinates of the computed optical centers, range: 9.1343-9.3204mm, mean = 9.2345mm, std = 0.0474mm 52

Figure 3.16 *y* coordinates of the computed optical centers, range: -0.047 -0.047mm, mean=0.00003mm, std = 0.0179mm 53

Figure 3.17 *z* coordinates of the computed optical centres, range: -1.0645 -2.0728mm, mean=0.4403mm, std =0.8766mm 53

Figure 4.1: Illustration of the epipolar geometry 56

Figure 4.2: The non-verged geometry of stereovision system 58

Figure 4.3: The geometry of verged stereo with the epipolar line (solid) and the geometry of non-verged stereo with epipolar line (dashed) 59

Figure 4.4 Construction of epipolar line on virtual camera 61

Figure 4.5 Illustration of the coordinate systems 63

Figure 4.6 Construction of epipolar lines using several points on $R3r$ 66

Figure 4.7 Epipolar lines and the first candidate points of several random points 67

Figure 4.8 Epipolar lines pass through their respective first candidate point 67

Figure 4.9 (a) and (b) Constructed epipolar lines based on the geometrical approach (Setup 1) 68

Figure 4.10 (a) and (b) Constructed epipolar line based on the conventional calibration approach (Setup 1) 69

Figure 4.11 20 pairs of Correspondence points (connected by blue lines) using different approaches (Setup 1) 70

Figure 4.12 Depth recovery errors using different methods (Setup 1) 72

Figure 4.13 Depth recovery errors using different methods (Setup 2) 73

Figure 5.1: Single-lens bi-prism stereovision system showing the virtual cameras and their FOVs 77

Figure 5.2: Two basic types of FOV: (a) divergent FOV, and (b) convergent FOV 78

Figure 5.3: Ray tracing of virtual bi-prism stereovision (only left virtual camera is shown) .. 80

Figure 5.4: Comparison of ϕ_2 and ϕ_4 against α for a fixed CCD camera image width (I=4.7616mm) 82

Figure 5.5: (a) Case 1: divergent system (b) Case 2: semi-divergent system (c) Case 3: convergent system 84

Figure 5.6: Detailed geometry of a divergent system 86

Figure 5.7: Detailed geometry of a convergent system 89

Figure 5.8: Experimental set-up 89

Figure 5.9 Interpretation of the captured image 90

Figure 5.10: Real scene captured using Set-up 1 configuration, the common FOVs are highlighted by the two white lines, the images were captured at a distance of (a) $z1 = 0.887m$ (b) $z2 = 1.075m$ (c) $z3 = 1.318m$ (d) $z4 = 1.821m$ 91

Figure 5.11: Graphical representation of the real scene captured by the system.....91

Figure 5.12: Comparison of experimental and theoretical FOV (a) Set-up 1 (b) Set-up 2.....93

Figure 5.13: Translation of bi-prism in the *z*-direction for a divergent system 95

Figure 5.14: Graphs showing rays 1 and 2 at different values of t_0 for setup 1.....96

Figure 5.15: Graphs showing rays 1 and 2 derived from experimental results at different t_0 for setup 1 96

Figure 5.16: Effect of *x*-axis translation of bi-prism on the system.....98

Figure 5.17: Ray tracing through the apex of the translated bi-prism 98

Figure 5.18: Ray tracing through the translated bi-prism half-planes 99

Figure 5.19: Geometrical analysis of *u* and *v* before and after *x*-axis translation.....100

Figure 5.20: Effect of increasing *d* on rays 1_l and 1_r 101

Figure 5.21: Graphs showing rays 1_l and 1_r derived from experimental results at different *d* for setup 1 102

Figure 5.22: Real scene captured using Set-up 1 configuration, the common FOVs are highlighted by the two yellow lines and the two sub-images are divided by the red line. The images are captured at varying d (a) $d = 0mm$ (b) $d = 4mm$ (c) $d = 8mm$ 102

Figure 6.1: Geometrical rays of the system 107

Figure 6.2 Experimental set-up.....114

Figure 6.3 Absolute depth errors with vs. actual depth (quantization error, ∂D≈1 pixel) ...115

Figure 6.4 Relative depth errors with vs. actual depth (quantization error, ∂D≈1 pixel) 115

Figure 6.5 Relative depth errors with vs. other parameters ($\partial f \approx 0.1$mm, $\partial \alpha \approx 0.001$rad, $\partial To \approx$ 1mm and $\partial n \approx 0.01$) 116

Figure 6.6 The estimated overall absolute depth and relative depth error 116

Figure 6.7 Relationship between the resolutions, field of view with focal length 119

Figure 6.8 2D schematic of the bi-prism geometry 120

Figure 6.9 Selection of the bi-prism size based on field of view of the camera 121

Figure 6.10 Non-convergence of $T+$ solution 122

Figure 6.11 $T-$ solution can be approximated to the real value of T 123

Figure 6.12 Value of T required to obtain the absolute error of 10mm-40mm 123

Figure 6.13 Value of α required to obtain the absolute error of 10mm-40mm 124

Figure 6.14 Absolute depth errors with different values of To 125

Figure 6.16 Absolute error of the system using bi-prism angle of 6.4°, To of 100mm with 4mm and 8mm focal lengths 126

Figure 6.17 Absolute error of the system using To of 100mm, focal lengths of 8mm with bi-prism angle of 6.4° and 21.6° 127

Figure 6.18 Absolute error of the system using focal lengths of 8mm and bi-prism angle of 6.4° with different values of To 128

Figure A1 Demonstration of the Snell's Law 151

Figure C1 Mid-point of two skew lines 153

Figure D1 Detailed geometry of a convergent system 154

Figure E1 Comparison of experimental and theoretical FOV for set-up 3 155

Figure E2 Comparison of experimental and theoretical FOV for set-up 4 156

LIST OF TABLES

Table 3.1 The values of parameters of the system used in the experiment 47

Table 4.1 Setup 1 67

Table 4.2 Setup 2 67

Table 4.3 Results comparison between the conventional calibration approaches and the proposed geometrical approach (Setup 1).....71

Table 4.4 Results comparison between the conventional calibration approaches and the proposed geometrical approach (Setup 2).....72

Table 5.1: Summary of the different cases in predicting a specific type of FOV.....84

Table 5.2: CCD cameras specifications 90

Table 5.3: Comparison of theoretical and experimental values of ray parameters.....92

Table 5.4: A summary of the effect of translation in both *z*- and *x*-axes on the FOV of a system 103

Table 6.1: The values of parameters of the system used in the experiment 114

Table 6.2 Real system parameters 122

Table 6.3 System Parameters.....125

LIST OF SYMBOLS

Baseline, i.e. the distance between the two camera optical centres:	λ
Effective real camera focal length:	f
Rotation matrix:	R
The angle of the bi-prism:	α
The center of the image plane:	O
The corner angle of the bi-prism:	δ
The depth of object in world coordinate system:	Z^w
The disparity of the corresponding points between the left and right image:	d
The distance between the apex of the bi-prism to the back plane of the bi-prism:	T
The distance between the real camera's optical centres to the apex of the bi-prism:	T_o
The epipole of left image:	e_l
The epipole of right image:	e_r
The extrinsic parameters:	M_{ext}
The intrinsic parameters:	M_{int}
The object point in world coordinate frame:	P_w
The point on the left image plane:	p_l
The point on the right image plane:	p_r
The refractive index of the prism glass material:	n
The sensor size of the real camera:	I
Translation vector:	T
World coordinates:	(X_w, Y_w, Z_w)

LIST OF ABBREVIATIONS

2D/2-D	Two-dimension
3D/3-D	Three-dimension
AD	Absolute intensity Differences
BP	Belief Propagation
CA	Cooperative Algorithms
CCD	Charge-Coupled Devices
CCS	Camera Coordinate System
DP	Dynamic Programming
DSI	Disparity Space Image
GC	Graph Cuts
LSSD	Locally scaled Sum of Squared Differences
LVC	Left Virtual Camera
LVCIP	Left Virtual Camera Image Plane
NCC	Normalized Cross Correlation
NN	Neural Network algorithm
RVC	Right Virtual Camera
RVCIP	Right Virtual Camera Image Plane
SAD	Sum of Absolute Differences
SSD	Sum of Squared Differences
SSSD	Sum of sums of absolute differences
SVD	Singular Value Decomposition
WCS	World Coordinate System

Chapter 1 Introduction

Stereovision is an area in computer vision which has drawn a great deal of attention in recent years. This is mainly due to its multitude of applications in robotics [1]-[2], medical devices [3]-[5], pattern recognition, artificial intelligence and many other fields. Apart from 3-D reconstruction, stereo imaging has been employed in engineering applications such as determining particle motion and velocity in stereo particle image velocimetry [6]-[8] and autonomous navigating vehicle [9]. More existing applications can be found in [10]-[15].

Stereovision refers to the ability to infer information of the 3-D structure of a scene from two or more images taken from different viewpoints. In general, the application of stereovision in 3-D scene recovery involves two main research issues. The first issue concerns a fundamental problem known as stereo correspondence. In a given stereovision image pair, it involves the search of the corresponding points in one image (the left image, say) in the other image (right image in this case). The problem becomes more difficult when some parts of the scene are occluded in one of the images. Thus, solving the stereo correspondence problem also involves in determining which of the image parts could not be matched. The second issue is the 3-D reconstruction which consists of the recovery of the 3-D depth of the scene. The ability of human eyes in 3-D perception is due to the computation of the positions differences between the correspondence image points which are known as disparity in brain. Therefore, if the geometry of the stereovision system is known and the stereo correspondence problem is solved, the disparities of all the image points (disparity map) can be reconstructed into the 3-D map of the captured scene.

Stereovision system usually employs two or multiple cameras to capture different views of a scene. A lot of efforts have been spent to develop a single-lens stereovision

system to replace the conventional two camera system. The advantages of single-lens stereovision system are obvious. As compared to conventional two or multiple camera stereovision systems, it has a more compact setup, lower cost, simpler implementation process, easier camera synchronization since only one camera is used, and also simultaneous image capturing, etc.

The focus of this thesis is on single-lens bi-prism stereovision system. Our research project employs the novel ideas of using a single camera in place of two or more cameras to achieve the stereovision effect and meanwhile to alleviate the operational problems of the above-mentioned conventional binocular, tri-ocular and multi-ocular stereovision system. The problems include difficulties in the synchronizing of image capturing, variations in the intrinsic parameters of the hardware used, etc. The solutions of these problems form the motivation of our earlier works in single-lens optical prism-based stereovision system as well as the concept of virtual camera in the year 2004 by Lim and Xiao [16]. By employing an optical prism, the direction of the light path from objects to the imaging sensor is changed and the different viewpoints of the object are thus generated. Such system is able to obtain multiple views of the same scene using a single camera in one image capturing step without synchronizations, offering a low-cost and compact stereovision solution. Continuous efforts have been made into this system in our research group, such as interpreting the concept of the virtual camera, enhancing the system modelling, solving the stereo correspondence problem and analysing the system error, etc.

1.1 Problem Descriptions

The projection of light rays onto the retina of our eyes will produce a pair of images which are inherently two dimensional. However, based on this image pair, we are able to interact with the 3-D surrounding in which we are in. This implies that one of the abilities of

the human visual system is to reconstruct the 3-D structure of the world from a 2-D image pair. Thus, algorithms are developed to duplicate this ability using stereovision system. In our works, the said desired motivation consists of the three important aspects, camera calibration, stereo correspondence, and parameter analysis.

For the single-lens system studied in this thesis, camera calibration is a process to recover all the intrinsic (focal length, sensor size and resolution) and extrinsic (position and orientation of the cameras) parameters of the virtual cameras. The accuracy of these parameters carries a great impact on reducing the error of the depth recovery. On the other hand, the complexity of the correspondence problem depends on the complexity of the scene. There are constraints (epipolar constraint [71]-[72]) that can help in reducing the number of false matches but there are still many unsolved problems in stereo correspondence especially for the single-lens prism based stereovision system. Besides, the study of light rays to compute the epipolar lines of the virtual cameras has not been covered in literature.

Furthermore, the study of field of view is an important aspect for this system as the choice of the system parameters (size and geometry of the prism) affects the overlapping region or common field of view of the virtual cameras. This is essential to make sure that the targeted object or scene is captured by the system in most of the applications.

Finally, note that the accuracy of the 3D depth recovery or reconstruction depends heavily on the system parameters. The relationship between the system parameters and the accuracy of depth recovery has received scant attention especially of this single-lens system, though it is of great practical importance. The system needed to be designed carefully to have accurate stereo correspondence and depth estimation. Thus, parameter analysis for the single-lens system will be studied in detailed in this thesis in order to improve and comprehend the accuracy of the depth recovery.

1.2 Contributions

Based on the earlier work in our lab by Lim and co-workers ([16]-[17], [19]-[34]) who modelled and modified the previous mirror based stereovision system to the current prism based stereovision system, contributions made in this thesis are presented as follows:

Virtual Camera Calibration

Virtual camera calibration which includes determining the extrinsic and intrinsic parameters of the virtual cameras is required. For this particular single-lens system, both the virtual cameras are formed by using a CCD camera with the aid of a bi-prism. Based on the virtual camera concept and geometrical calibration approach proposed by K.B. Lim and Y. Xiao [16], [17], we will propose a new geometrical approach to recover the basic properties of the virtual cameras such as optical centres, focal length and orientation. The virtual cameras will be modelled using the pinhole camera concept. Besides, the efficiency of the proposed method is verified by comparing it to the conventional methods such as Tsai [55] and Zhang [68] approaches.

Virtual Epipolar Line

To solve the stereo correspondence problem, virtual epipolar lines approach which will reduce the correspondence points searching time is proposed. This approach employs 3D geometrical analysis and makes use of the geometry of the virtual cameras. The main idea of the approach is to construct virtual epipolar geometry for the virtual cameras by using two unique points. Thus, with a known image point in one of the virtual image plane (let say left virtual camera), the candidates of the correspondence points in the other image plane (right virtual camera in this case) can be determined. Once the pair of correspondence points is found, the depth recovery could be achieved with simple geometry. Experiments have been carried out to study the effectiveness and accuracy of this method on single-lens prism based

stereovision system. The proposed method will also be compared with the conventional epipolar constraint using the fundamental matrix [35].

Field of View (FOV)

In machine vision, FOV is the part of the scene which is captured by a camera at a particular position in the 3D space. It is an important aspect in stereovision system as the choice of the system parameters affects the FOV of the camera. Objects outside the system FOV when the image is captured will not be recorded. Thus, the system should be carefully designed so that the object of interest is captured successfully. The FOV of the single-lens bi-prism stereovision system is affected by various parameters which include the corner angle of the bi-prism, the position and orientation of the bi-prism with respect to the camera and the material of the bi-prism.

In this study, the main objective is to study how the FOV of the system is affected by the bi-prism angle, as well as the position of the bi-prism with respect to the camera. The former will be studied in detail, encompassing both divergent and convergent systems while the focus of the latter will be on divergent systems only.

Parameter Analysis

The accuracy of the single-lens system is dependent on the system parameters such as focal length, angle of the bi-prism, refractive index of the bi-prism and distance between the camera and bi-prism. A mathematical equation is derived to estimate the range in terms of the system parameters which will be described in detail. The accuracy of the system will be studied in detail with respect to each of the parameters. The relative depth error which is essential to design the system appropriately for practical usage is then formulated. Furthermore, the concept of variable parameter that examines the possibility to improve the accuracy of the system for both short and long range by varying the values of the parameters is proposed.

An experimental setup was established and several experiments were carried out to test the effectiveness of the single-lens binocular stereovision systems and to verify the efficiency of the proposed methods. Results from the experiments are compared with the conventional approaches to confirm its accuracy and effectiveness. We believe that most of the work presented in this thesis, especially the virtual epipolar line and parameter analysis are novel and practically useful in science and industrial areas. Part of the content of this thesis has been published in [121]-[123].

1.3 Outline of the thesis

The outline of the thesis is structured as follows: Chapter 2 gives a review on the previous development of the single-lens stereovision system and the stereovision algorithms which includes camera calibration, stereo correspondence algorithm, epipolar geometry and parameter analysis. In Chapter 3, the geometrical method used for the virtual camera modelling is discussed and compared to conventional methods; The proposed virtual epipolar line technique to solve the stereo correspondence problem is described in Chapter 4 and the results are compared to the conventional approach as well; Chapter 5 describes the developments of the methodologies to predict the FOV of the system and its geometry given a specific bi-prism angle and examines the effect of a system's FOV under *z*- and *x*-axes translation of the bi-prism. A comprehensive study of the parameter analysis is presented in Chapter 6 and the conclusion is given in Chapter 7; Last but not the least, the future work is proposed in Chapter 8.

Chapter 2 Literature review

In this chapter, recent works pertaining to stereovision techniques are reviewed. They include the algorithms of calibration, stereo correspondence, depth recovery, and single-lens stereovision techniques. Section 2.1 shows various stereovision systems which include both two-camera and single-lens system developed earlier by researchers. Conventional camera calibration technique and stereo correspondence algorithms are presented in Section 2.2 and 2.3, respectively. Section 2.4 gives a review on the parameter and quantization analysis of the two-camera system while the final section, Section 2.5 summarizes the reviews done in this chapter.

2.1 Stereovision system

Conventionally, a stereovision system requires two or more cameras to capture the same scene in order to obtain disparities for depth recovery. By using a single camera with known intrinsic parameters, it is not possible to obtain the three-dimensional position of a point in 3-D space. This is because the mapping of a 3D scene onto a 2D image plane is essentially a many-to-one perspective transformation. As a result, in order to achieve 3D perception comparable to the human vision system, the same scene point has to be viewed from two or more different viewpoints. As long as this criterion is satisfied, we can achieve stereovision effect even by using a single camera. Thus, in the past few decades, there were various single-lens stereovision systems proposed to potentially replace the conventional two camera system with some significant advantages which will be covered in more detail in Section 2.1.2.

2.1.1 Conventional two camera system

A conventional stereovision system used in depth recovery employs two or more cameras to capture the images from different viewpoints. Figure 2.1 shows the classical stereovision system using two cameras. The coordinate systems are defined as follows:

(X_w, Y_w, Z_w) : World Coordinate System

(X_L, Y_L, Z_L) : Left Camera Coordinate System

(X_R, Y_R, Z_R) : Right Camera Coordinate System

(x_{lc}, y_{lc}) : Left Camera Pixel Coordinate System

(x_{rc}, y_{rc}) : Right Camera Pixel Coordinate System

The focal lengths of the two cameras are assumed to be same. They are translated by a baseline distance λ in the X_w direction. The optical axes of these cameras are parallel and perpendicular to the baseline connecting the image plane centres (in the same *XZ* plane). The coordinates of the scene point are shown below:

$$X_w = \frac{\frac{\lambda(x_l + x_r)}{2}}{x_l - x_r} \; ; \; Y_w = \frac{\frac{\lambda(y_l + y_r)}{2}}{x_l - x_r} \; ; \; Z_w = \frac{\lambda f}{x_l - x_r} \tag{2.1}$$

where λ is the length of the baseline connecting the two camera optical centres and f is the focal length of each camera. The value of $(x_l - x_r)$ is termed as disparity, which is the difference between the positions of a particular scene point appearing in the two image

planes. A more detailed explanation on the geometry of this setup can be found in Grewe and Kak [36].

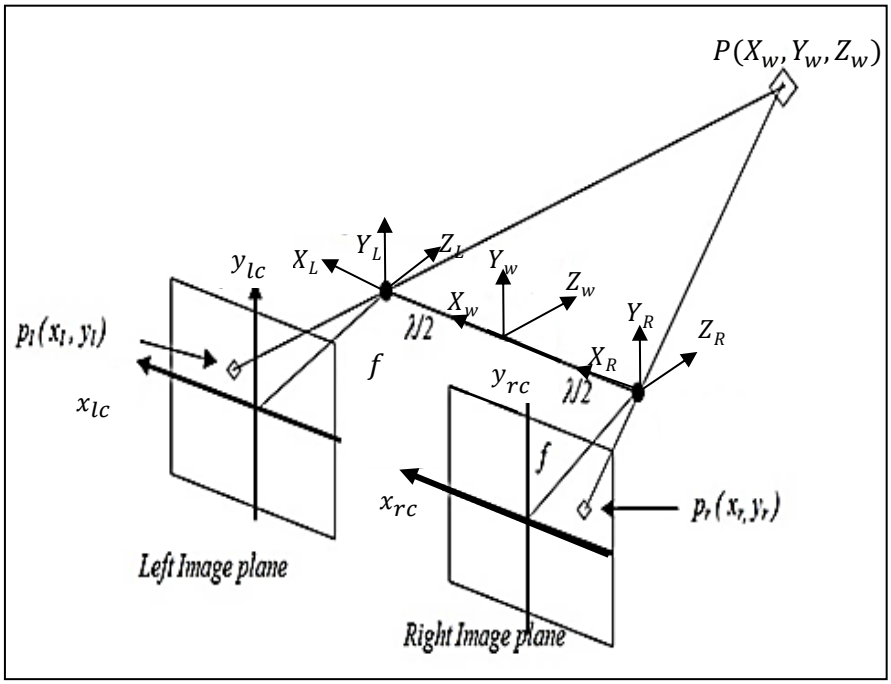


Figure 2.1: Modeling of two camera canonical stereovision system

In practice, the conventional stereovision systems have the advantages of a simpler setup and easier in implementation. However, the difficulty in synchronized capturing of the image pairs by the two cameras and the cost of the system make them less attractive. Therefore, single-lens stereovision systems are explored by researchers to solve these short-comings.

2.1.2 Single-lens stereovision system

In the past few decades, there were various single-lens stereovision systems proposed to potentially replace the conventional two camera system with some significant advantages such as lower hardware cost, compactness, and reduction in computational load.

Single-lens stereovision system with optical devices was first proposed by Nishimoto and Shirai [37]. They proposed using a glass plate which is positioned in front of a camera

and the glass plate is free to rotate. The rotation of the glass plate causes deviation of the camera's optical axis due to reflection which produces a pair of stereo images as shown in Figure 2.2. The main disadvantage of this method is the disparities between the image pairs are small and the system needs to capture the scene twice to obtain a pair stereo image.

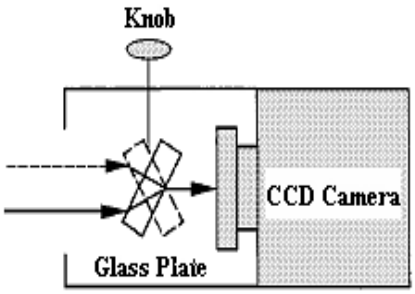


Figure 2.2 A single-lens stereovision system using a glass plate (Nishimoto and Shirai [37])

Gao and Ahuja [38] improved Nishimoto and Shirai [37] model into a multiple camera equivalent system. Instead of capturing one pair of stereo images, Gao and Ahuja [36] proposed to capture a sequence of images as the plate rotates which provide a large number of stereo pairs with larger disparities and field of view. Based on this system, Kimet al. [39] suggested a new distance measurement method using the idea that the corresponding pixel of an object point at a further distance away moves at a higher speed in a sequence of images.

The idea of the single-lens stereovision system with the aid of three mirrors was introduced by Teoh and Zhang [40]. Two of the mirrors are fixed at 45 degrees at the top and bottom and the third mirror rotating freely in the middle. Two shots will be taken with the third mirror aligned to be parallel to the fixed mirrors as shown in Figure 2.3.

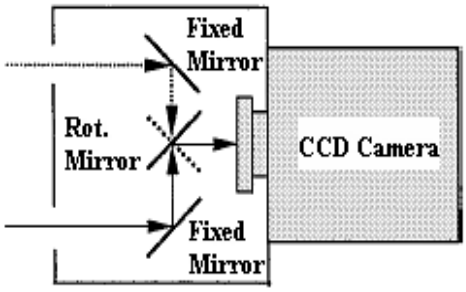


Figure 2.3 A single-lens stereovision system using three mirrors (Teoh and Zhang [40])

Francois et al. [41] further refined the concepts of stereovision from a single perspective to a mirror symmetric scene and concluded that a mirror symmetric scene is equivalent to observing a scene with two cameras and all the traditional analysis tools of the binocular stereovision can be applied. The main problem of the mirror based single-lens system is its applications are only limited to a static scene as the stereo image pairs are obtained with two separate shots. This problem was overcome by Gosthasby and Gruver [42] whose system captured image pairs by the reflection from mirrors as shown in Figure 2.4.

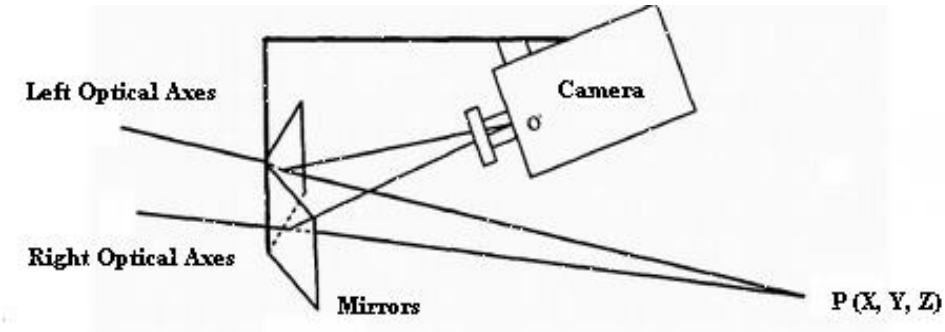


Figure 2.4 A single-lens stereovision system using two mirrors (Gosthasby and Gruver [42])

Inaba [43] later introduced a mirror based system which controls its field of view using a movable mirror. Subsequently, the mirror based system was further studied by Nene and Nayar [44]. Instead of using a flat mirror, they used hyperboloids and paraboloids reflecting surfaces which have a bigger field of views (see figure 2.5). In practice, such systems are difficult to be used as the projection of the scene by the curved mirrors is not

from a single viewpoint. In other words, this implies that the pinhole camera model cannot be used, thus making calibration and correspondence a more difficult task.

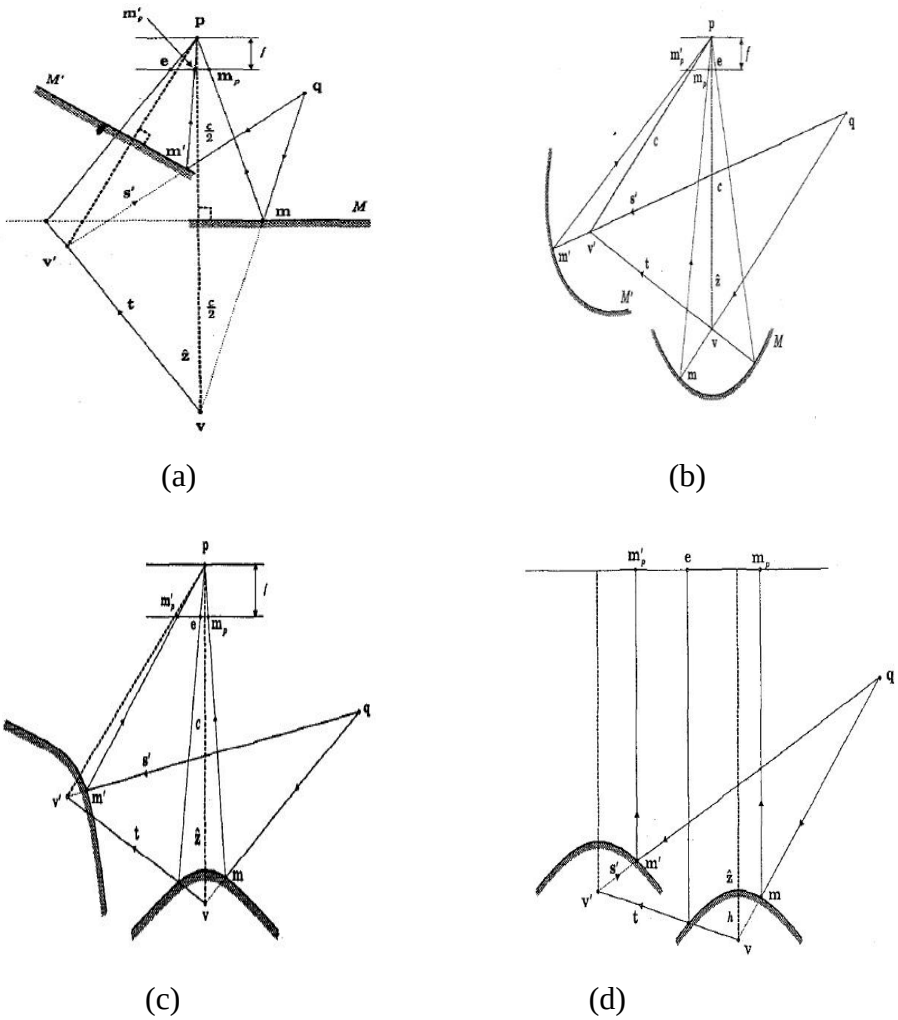


Figure 2.5: Four stereovision systems using mirrors (a) two planar mirrors; (b) two ellipsoidal mirrors; (c) two hyperboloidal mirrors; (d) two paraboloidal mirrors (Nene and Nayar [44])

There were also some efforts in developing single-lens stereovision system using known cues such as illuminations, known geometry of an object, etc. Segan et al. [45] proposed a system which used one camera and a light source to track user's hand in 3D space. He calibrated the light source and used the shadow of the hand projection as the cue for depth recovery. Moore and Hayes [46] presented a simple method of tracking the position and orientation of an object from a single camera by exploiting the perspective projection

model. Three coplanar points on the object are identified, which are the cues for the image pairs, and their distances from the camera lens are measured.

In the work by LeGrand and Luo [47], an estimation technique which retains the non-linear camera dynamics and provides an accurate 3-D estimation of the positions of the selected targets within the environment was presented. This method is applied in robot navigation where the robot continuously computes the centroid of the target and uses the estimation algorithm to calculate the target's position. This implies that the stereo information is generated from the motion information which is acquired through the movement-sensor attached to the robot.

Adelson and Wang [48] proposed a system called plenoptic camera which achieves single-lens stereovision. By analysing the optical structure of the captured object where the light striking the object is somewhat different compared to the light striking its adjacent region, they presented a method to infer the depth information of the object. There are more detailed studies on single-lens stereovision system with cues which can be found in [49]-[52].

Lee and Kweon [53] proposed a single-lens stereovision system using a bi-prism which is placed in front of a camera as shown in Figure 2.6. The advantages of using this system include potentially cost saving since only single camera is required; it is also more compact and has fewer system parameters. Stereo image pairs are captured on the left and right halves on the image plane of the camera due to refraction of the light rays through the prism. To solve the stereo correspondence problem of this system, they proposed the concept of virtual points. Any arbitrary points in 3-D space are projected into two virtual points with some deviation caused by the bi-prism which are determined by the refractive index and the angle of the bi-prism. They provided a simple mathematical model in obtaining the disparities of the virtual points and it works when the angle of the prism is sufficiently small. This is because during their derivations, they assumed that the two virtual points only have deviation

in the X-axis direction. In other words, they made an assumption that the two virtual cameras are coplanar. The error using this method will become significant when the angle of the prism becomes larger.

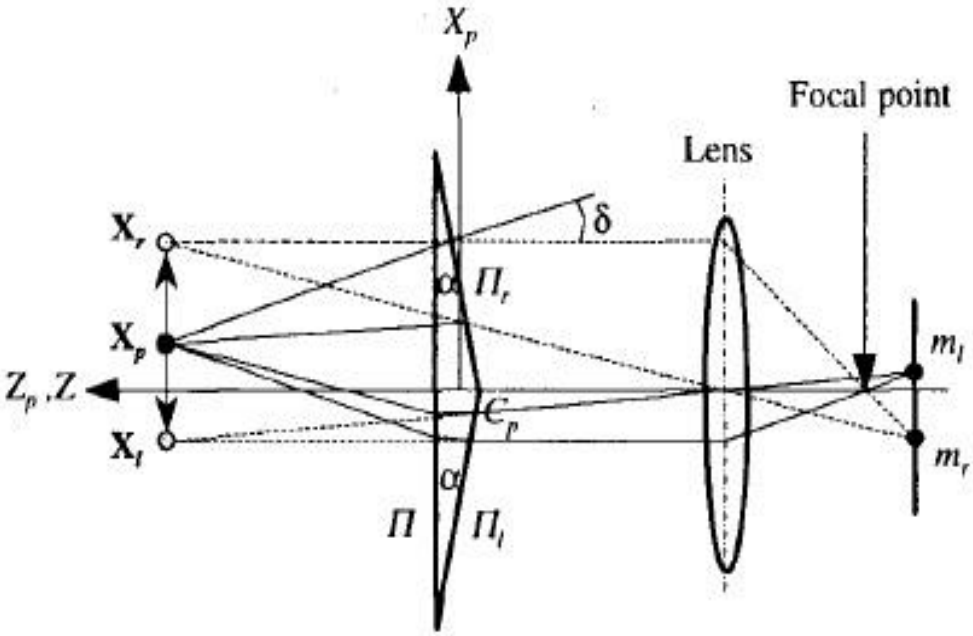


Figure 2.6 Illustration of the bi-prism system proposed by Lee and Kweon [53]

Lim and Xiao [16]-[17] improved the system and extended the bi-prism study to a multi faced prism. They proposed the concept of calibrating the virtual cameras which do not exist physically (more explanation in Section 2.2) as shown in Figure 2.7. By placing a two face prism in front of a CCD camera and the apex of the prism bisecting the CCD image plane into two halves, two sub-images of the same scene is captured on the left and right image plane of the camera. These images are taken to be equivalent to two images taken using two virtual camera systems with different orientation and position which produce some disparities. Besides, they even extended the concept to tri-faced and n-faced prism (see figure 2.8).

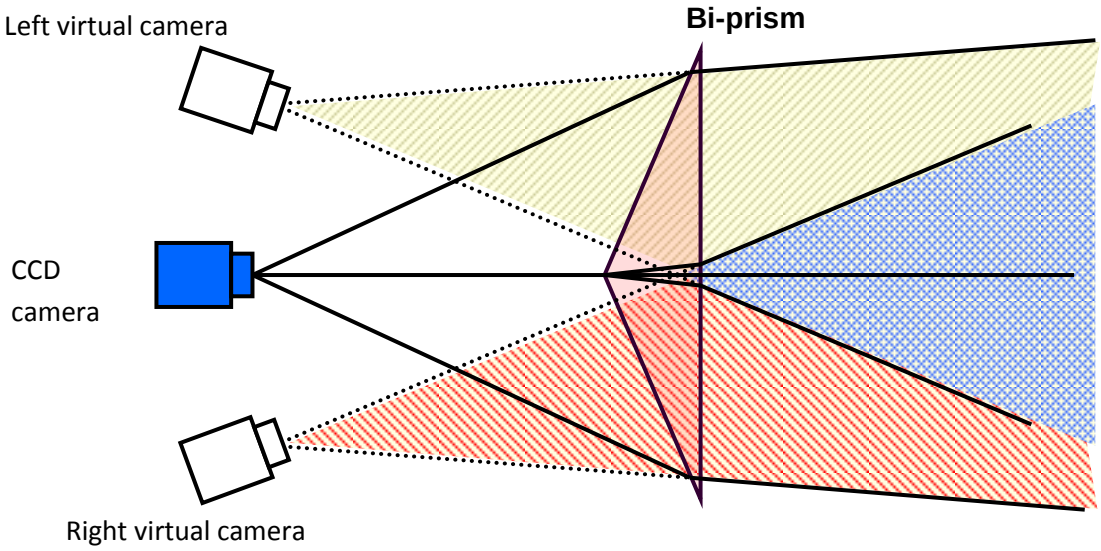


Figure 2.7 Single-lens bi-prism stereovision system (Lim and Xiao [16])

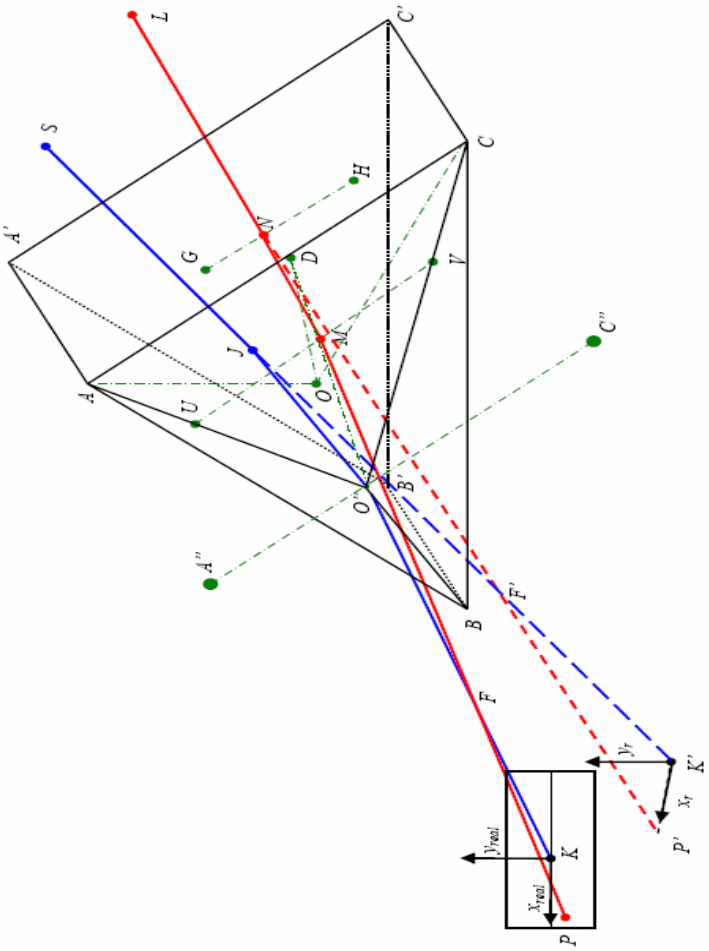


Figure 2.8 Virtual camera calibration of tri-prism system (Lim and Xiao [16])

Based on the virtual camera concept, Lim et al. [33]-[34] provided a comprehensive study on virtual camera rectification based on this system. Genovese et, al. [54] further studied the image correlation and distortion of the single-lens bi-prism based stereovision system.

Based on the single-lens prism based system developed by Lim et al. ([16], [17], [19]-[34]), as mentioned above, we focus on resolving stereovision problem such as virtual camera calibration, correspondence problem and parameter analysis of the single-lens bi-prism based stereovision system.

2.2 Stereo camera calibration

Camera calibration is defined as the formulation of the projection equations which link the known coordinates of a set of 3D points and their projections on the image plane pixel coordinates. This relationship is defined by the camera intrinsic parameters, such as camera focal length and lens distortion, and the extrinsic parameters, such as relative position and orientation of the cameras with respect to a predefined world coordinate system. Tsai [55] proposed a simple calibration process to determine the extrinsic and intrinsic parameters which link the 3-D world coordinates to the image plane pixel coordinates as shown in Figure 2.9. The illustrations of the relationship between the coordinate systems are shown in Figure 2.10.

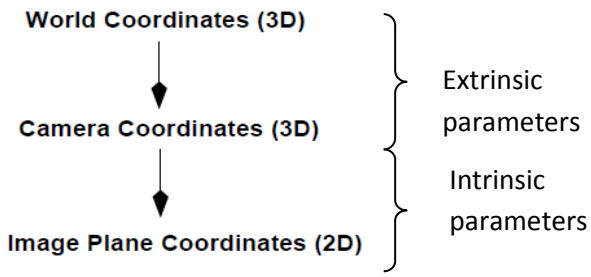


Figure 2.9 Transformation of 3-D world coordinates to camera image plane coordinates

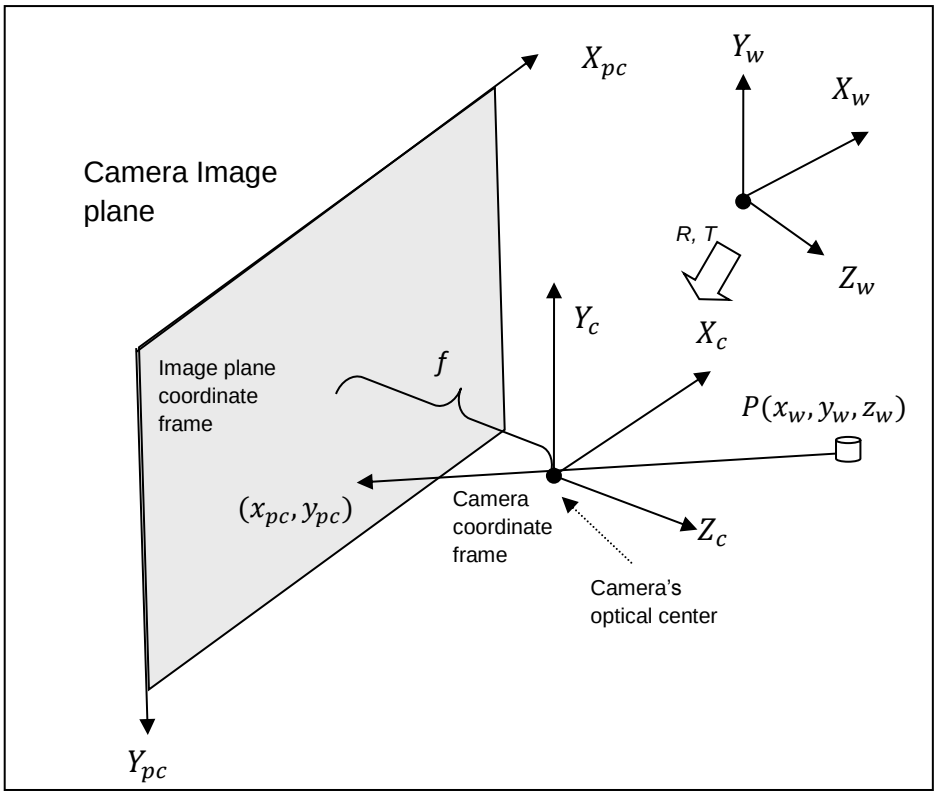


Figure 2.10 Illustrations of the coordinates systems

As shown in Figure 2.10, the relationship of a point P in the world coordinate frame and the camera image plane coordinate frame can be written as:

$$\begin{bmatrix} x_{pc} \\ y_{pc} \\ 1 \end{bmatrix} = M_{int} \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \\ 1 \end{bmatrix} \tag{2.2}$$

The coordinate systems are defined as follows:

(X_w, Y_w, Z_w) : World Coordinate System

(X_c, Y_c, Z_c) : Camera Coordinate System

(X_{pc}, Y_{pc}) : Image Plane Coordinate System

Where R and T are the extrinsic parameters (rotational and translational matrix, respectively) and M_{int} is the intrinsic parameters which include focal length, sensor size and distortion.

The objective of the calibration process is to recover the parameters in both extrinsic and intrinsic parameters. With the parameters known, we can easily recover the coordinates of a point $P\,(x_w, y_w, z_w)$ in the three dimensional world coordinates with the pixel coordinates of the corresponding points in the camera image plane $(x_{pc,} y_{pc})$. The literature review for conventional camera calibration methods and virtual camera calibration technique for single-lens bi-prism based stereovision system will be presented in Sections 2.2.1 and 2.2.2.

2.2.1 Conventional camera calibration methods

The accuracy of camera calibration to recover the extrinsic and intrinsic parameters will directly affect the performance of a stereovision system. Therefore, a great deal of efforts is spent to deal with this challenge. Based on the techniques used, camera calibration methods can be classified into 3 categories:

(1) *Linear transformation methods.* In this category, the objective equations are linearized from the relationship between the intrinsic and extrinsic parameters [56], [57]. Therefore, the parameters are recovered by solving these linear equations.

(2) *Direct non-linear minimization methods.* These methods use the interactive algorithms to minimize the residual errors of a set of equations, which can be achieved directly from the relationship between the intrinsic and extrinsic parameters. They are only used in the classical calibration techniques [58], [59].

(3) *Hybrid methods.* These methods make use of the advantages of those in the two previous categories. Generally, they comprise of two steps: the first step involves solving most of the camera parameters in linear equations; the second step employs a simple non-linear optimization to obtain the remaining parameters. These calibration techniques are able to

solve different camera models with different distortion models. Therefore, they are widely studied and used in recent works [60]–[64].

A concise introduction of stereovision can be found in the book by Trucco and Verri [35]. More explanations and discussions can be found in the books by Faugeras [65], Hartley et al.[66], and Sonka et al.[67]. Zhang [68] proposed a more flexible calibration technique which only requires the camera to observe a planar pattern which is captured at few different orientations. It allows either the camera or the planar pattern to be freely moved without knowing the motion. Comparing the Zhang's approach to the classical approach, the former is more flexible and easier to use.

To enhance the application of a stereovision system, calibration techniques are further improved to address the active stereovision system problem, which allows the independent movement for each of the cameras. This enables a wider effective field of view and reduces the occlusion problem. Kwon, Park and Kak [69] proposed a new method to estimate the locations and orientations of the pan and tilt axes for the cameras through a closed-form solution. By combining these axes with the homogeneous transformation relationships, they derived a set of calibration parameters which is valid over a large variation in the pan and tilt angles.

2.2.2 Virtual camera calibration technique

Lim and Xiao [16], [17] proposed a novel technique to calibrate the virtual cameras generated by the single-lens bi-prism based stereovision system. In their works, they proved that their proposed technique outperforms the classical methods for this particular single-lens system. We will briefly discuss their calibration technique in this chapter.

As shown in Figures 2.11 and 2.12, by choosing two unique image points: 1st point is the centre of the image plane which lies along the optical axis of the CCD camera and the apex of the bi-prism; 2nd point is the boundary point on the same scan line with the centre

point. By projecting the two points into 3D space, we will obtain two rays, Ray1 and Ray2 as shown in Figure 2.12. Both of these rays will be refracted twice through the bi-prism forming Ray12 and Ray22. The intersection point of the back projection rays of Ray12 and Ray22 to the left virtual camera indicates the position of the left virtual camera. $\omega_1'$is the field of view of the two virtual cameras generated by the biprism. $\phi_1, \phi_2, \phi_3, \phi_4, \phi_1{}', \phi_2{}', \phi_3{}', and\ \phi_4{}'$ are the series of incident and refracted angles of the two unique image points. In their derivations, they assumed that all the rays and points are in 2-dimensional ($X_w Z_w$ plane). This is valid since the two chosen points lie on the same scan line parallel to the X_w-axis of the world coordinate frame.

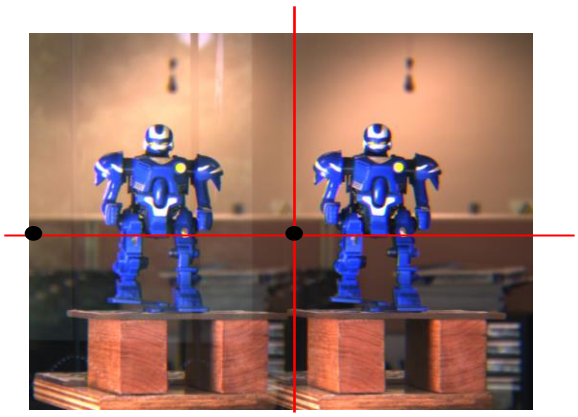


Figure 2.11 Image captured using the bi-prism stereovision system, two black dots indicate the two unique pixels chosen for virtual camera modeling

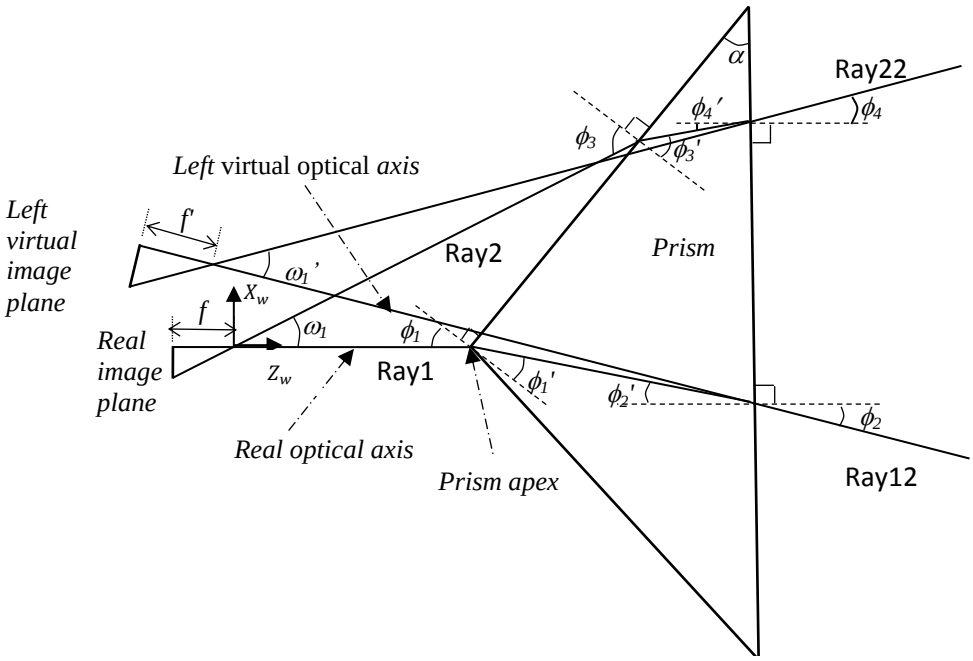


Figure 2.12 Formation of virtual cameras using a bi-prism (top view)

This technique is simple as all the geometrical mathematical derivations are in 2-dimensional. However, there are some limitations using this method. For example, when the field of view or the sensor size of the cameras are unknown, or when the apex of the prism is not placed along the optical axis of the CCD camera (bisecting the CCD image plane into two equal halve) as shown in Figure 2.13, both the unique image points required cannot be located.

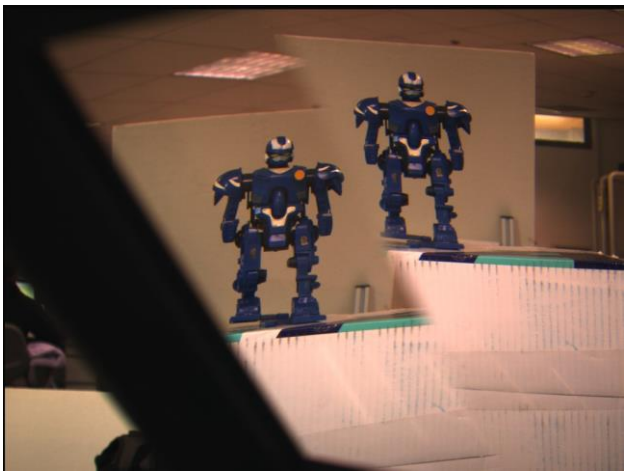


Figure 2.13 Image captured by the system in non-ideal situation

Using the earlier works of Lim [16] and [17] as foundation, we propose a better approach in modelling the virtual cameras of the single-lens prism based stereovision system (see Chapter 3). The proposed approach could be generalized to address the above-mentioned problem as shown in Figure 2.13 which could not be solved using the previous method.

2.3 Stereo correspondence problem

Stereo correspondence problem is always the key issue in stereovision. Given two or more images of the same scene, the correspondence problem is to find a set of image points in one image (the left image, say) which can be identified as the same image points in another image (the right image in this case). By obtaining the correspondence points and calculating the disparities (difference in positions on their respective image planes), the depth of the

points in 3D space can be recovered. The main purpose of the stereo correspondence algorithm is to find the best correspondence point accurately and effectively. It was heavily investigated as the general solution does not exist. This is because captured images are full of randomness, ambiguous matches due to occlusion, lack in texture, and variation in illuminations.

Brown et al. [70] described the detail of the taxonomy of stereo correspondence algorithm. It can be classified into two different approaches such as local approach where the corresponding matching process is locally applied to the pixel of interest and the other one is called global method, in which searching process works on the entire image. Local methods can be very efficient, but they are sensitive to ambiguous regions in images (e. g., occlusion regions). To further reduce the computational complexity, constraints from image plane geometry, epipolar constraint ([71] and [72]) is commonly exploited in solving the stereo correspondence problem. On the other hand, global methods can be less sensitive to these problems since global constraints provide additional support for regions which are difficult to be matched locally. However, these methods are more computationally expensive. In this section, our literature review on stereo correspondence will follow Brown's classification [70].

2.3.1 Local method

In general, local correspondence method can be divided into three categories: gradient method, feature based method, and block matching method.

(a) Gradient method

Gradient method (optical flow) can be applied to determine disparities between two images by formulating a differential equation relating motion and image brightness. It is

commonly applied in real time stereovision system. Assumption is made such that as the time varies, the image brightness (intensity) of points does not change as they move in the image. In other words, the change in brightness is entirely due to motion [73]. $E(x, y, t)$ is the image intensity at points (x, y) which is a continuous and differentiable function of space and time. If the image pattern is locally displaced by a distance (dx, dy) over a time period dt, the gradient method can be mathematically written as:

$$E(x, y, t) = E(x + dx, y + dy, t + dt) \tag{2.3}$$

$$\frac{dE}{dt} = 0 \tag{2.4}$$

$$\frac{dE}{dt} = E_x V_x + E_y V_y + E_t = \nabla E^T \left(V_x, V_y\right) + E_t = 0 \tag{2.5}$$

where ∇E and E_t are the spatial image intensity derivative and the temporal image intensity derivative, respectively. Among them, ∇E and E_t are known parameters which can be measured from the images while (V_x, V_y) are the unknown optical flow components $(\frac{dx}{dt}, \frac{dy}{dt})$ in the x and y directions. A smoothness term was later introduced by Horn and Schunck [73] in order to compute the optical flow for a sequence of images. Gradient method works very well when the 2D motion is “small” and the change of intensity is entirely due to motion.

(b) Feature based method

In contrast, feature based correspondence algorithms can deal with the sequence of images when the optical motion is “large”. Given a stereo image pair, they match dominant features in the left image to those in the right image. Feature based correspondence methods are insensitive to depth discontinuities and work very well when there is a presence of regions with uniform texture. Venkateswar and Chellappa [74] proposed the hierarchical of the method where the matching starts at the highest level of the hierarchy (surfaces) and

proceeds to the lowest ones (lines) because higher level features are easier to match due to fewer numbers and more distinct in form. Subsequently, segmentation matching was introduced by Todorovic and Ahuja [75] in order to identify the largest part in one image and its match in another image by measuring the maximum similarity measure defined in terms of geometric and photometric properties of regions (e.g., area, boundary, shape and colour). The ground-breaking study was proposed [76]-[79] to develop invariant features, which are invariant to image scale and rotation. Their results showed that the method is robust to match features correctly across a substantial range of affine distortion, noise, and change in illumination. In summary, features based methods can match some of the features or scenes accurately and quickly but perform poorly when feature extraction is not possible and only sparse depth map is formed.

(c) Block matching method

Block matching method or correlation based algorithm measures the similarity between two windows on the two images. The correspondence point is given by the window with the highest correlation number or the greatest similarity. Sum of Absolute Differences (SAD) is one of the simplest block matching methods which are calculated by subtracting pixels within a square neighborhood between the left image ($\boldsymbol{I_l}$) and the right image ($\boldsymbol{I_r}$) followed by the aggregation of absolute differences within the square window. The correspondence points are then given by the window that has the maximum similarity. In Sum of Squared Differences (SSD), the differences are squared and aggregated within a square window. This measure has a higher computational cost compared to SAD algorithm as it involves numerous multiplication operations. On the other hand, Normalized Cross Correlation is more complex compared to both SAD and SSD algorithms as it involves numerous multiplication, division and square root operations.

Although correlation-based algorithms are easier to implement and provide dense disparity map, it is very sensitive to the changes in illumination direction and has a poor effect on texture-less images. Some research works to achieve better and robust performance of the correlation-based algorithms can be found in [80] and [81]. The former provided a comprehensive description and comparison on different correlation algorithms, while, the latter presented the performance comparison of rank and census matching with those of correlation and difference metrics.

2.3.2 Global method

Global matching techniques such as graphs cuts, Loopy BP and tree-reweighted message passing, converts the stereo correspondence problem to energy minimization problem which yields high accuracy. It exploits non-local constraints in order to reduce sensitivity to local regions in the image that fails to match, due to occlusion, uniform texture, etc. These constraints make the computational cost of global matching significantly greater as compared to the local method. In general, the global energy function contains two components, such as smoothness term and energy part which are shown as follows:

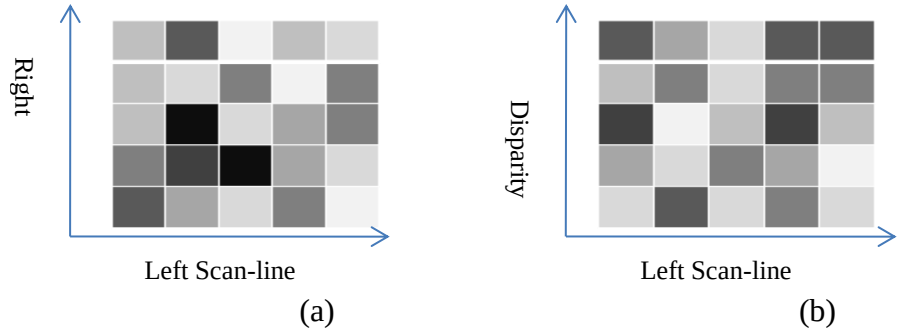
$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \tag{2.6}$$

$E_{data}(d)$ is the data term which measures the correctness of the disparity function d agrees with the stereo image pairs while $E_{smooth}(d)$ represents the smoothness assumptions made by the algorithm. λ is a weight of the smoothness term. Dynamic programming, graph cut and cooperative matching algorithms will be reviewed in this section.

(a) Dynamic programming

Dynamic programming (DP) is a global method that reduces the computational complexity of optimization problems by decomposing them into smaller and simpler sub-

problems [82]. A global cost function is computed in stages, with the transition between stages defined by a set of constraints. For stereo correspondence, the epipolar monotonic ordering constraint allows the global cost function to be determined as the minimum cost path through a disparity space image (DSI). The local cost functions for each point in the DSI could be defined using one of the area based methods. There are two ways to construct a DSI, which is shown in Figure 2.14.



(a) (b)

Figure 2.14 (a) disparity space image using left-right axes and; (b) another using left-disparity axes

As shown Figure 2.14, the intensities shown represent the respective costs of potential matches along the scan-lines, with lighter intensities having lower cost. Ohta and Kanade [78] and Cox et al. [79] defined the axes as the left and right scan-lines. For this case, dynamic programming is used to determine the minimum cost path from the lower left corner to the upper right corner of the DSI. The second method for constructing a DSI is to define the axes as the left scan-line and the disparity range (see Figure 2.14(b)), was done by Intille and Bobick [83]. In this case, dynamic programming is used to compute the minimum cost path from the first column to the last column (see Figure 2.9 (b)). With N pixels in a scan-line and a disparity range of D pixels, global optimization requires O (N^N) operations per scan-line, in addition to the time required to compute local cost functions. Birchfield and Tomasi [84] have reduced the complexity to O (ND log$D$) by pruning nodes when locally lower cost alternatives are available.

Beside computational efficiency, DP is designed to solve the occlusion problem which yields a much complicated problem as the energy function near an occlusion boundary is high. Belhumeur [85] proposed a two-stage approach to deal with this problem by replacing the matching cost at boundaries with a fixed occlusion cost. One of the advantages of dynamic programming is that it provides global support for local regions that lack texture which would be mismatched otherwise. These local regions present difficulty for a global search since the cost function in these regions is low. The main disadvantage of DP is that local errors may be propagated along a scan-line which corrupts other potential good matches. Another significant limitation of DP is its inability to integrate both horizontal and vertical continuity constraints. Thus, graph cut is proposed to tackle this problem while maintaining the framework of DP.

(b) Graph cut

As mentioned previously, graph cut is proposed to integrate horizontal and vertical continuity constraints effectively [86]. Zhao [87], and Thomos et al. [88] use the well-known preflow-push lift-to-front algorithm. The complexity of this algorithm is $O(N^2D^2 \log(ND))$, where N is the number of pixels in the image and D is the image resolution, which is significantly greater than that of dynamic programming algorithms. However, Roy and Cox [89] claimed that the average observed time is $O(N^{1.2}D^{1.3})$, which is much closer to that of dynamic programming. One limitation of this algorithm is that the classical implementations require significant memory resources, making this approach cumbersome for use with large and real time images. To handle this problem, Thomos et al. [88] developed a data structure that reduces the memory requirements by a factor of approximately four, making this algorithm more manageable for large data sets.

Furthermore, recent developments on graph cuts have focused on new graph architectures and energy minimization algorithms. Boykov and Kolmogorov [90] have developed an approximate Ford-Fulkerson style augmenting paths algorithm, which is much faster in practice than the standard push-re-label approaches. Kolmogorov and Zabih [91] propose a graph architecture in which the vertices represent pixel correspondences (rather than pixels themselves) and impose uniqueness constraints to handle occlusions. These recent graph cut methods have been shown to be among the best performers.

(c) Cooperative Algorithms

Cooperative optimization is the latest discovered optimization method to resolve the hard optimization problems [92]. It has been reported in [93] that cooperative optimization has achieved remarkable performances at solving a number of real-world *NP*-hard problems with the number of variables ranging from thousands to hundreds of thousands. The problems span several areas such as DNA image analysis [93], shape from shading [94] and stereo matching [95], proving its generality and power. In all these cases, its memory usage is efficient and fixed with simple, regular, and fully scalable operations. All these features make it suitable for parallel hardware implementations. The results also showed that it significantly outperformed the classic simulated annealing in finding the global optimal solutions.

Recently, Marr and Poggio [96] proposed a promising cooperative algorithm to compute matching scores locally using match windows. Nevertheless, they demonstrated the "global behaviour" by refining the correlation scores iteratively using the uniqueness and continuity constraints. Subsequently, Zhang and Kambhamettu [97] take advantage of image segmentation to compute the initial matching scores. Their proposed algorithm exploits the results of the segmentation in the choice of local support area and prevents the support area from overlapping a depth discontinuity.

Besides, Yong et. al. [98] improved the study of energy minimization problem so that it could be more robust to the change of lighting geometry, illumination and camera parameters between the left and right images. A very good review has been written by Szeliski et al.[99] who investigated three of the most recent methods mentioned above. They studied the trade-offs among the energy minimization algorithms and provided a software interface that allows researchers to easily switch between the optimization methods. Other than the aforementioned method, there have been a lot of improvement and development in this area which can be found in [100]-[103].

2.3.3 Epipolar constraint

An additional constraint using epipolar geometry [35] can be applied on the stereovision system to reduce the searching and computation time. Figure 2.15 shows the simplest two camera system such that the two image planes are coplanar. A point in the 3D space together with the two optical centres of the cameras forms a plane which is called epipolar plane. The intersection between the epipolar plane and the image planes is called epipolar line. In this case, due to the geometry of the two coplanar image planes, the epipolar line is a horizontal scan line. Thus, given a point P_l on the left image, the correspondence point P_r lies on the epipolar line (horizontal scan line) on the right image. By constructing the epipolar line, the stereo correspondence searching is reduced to a 1-dimensional search along the epipolar line [35].

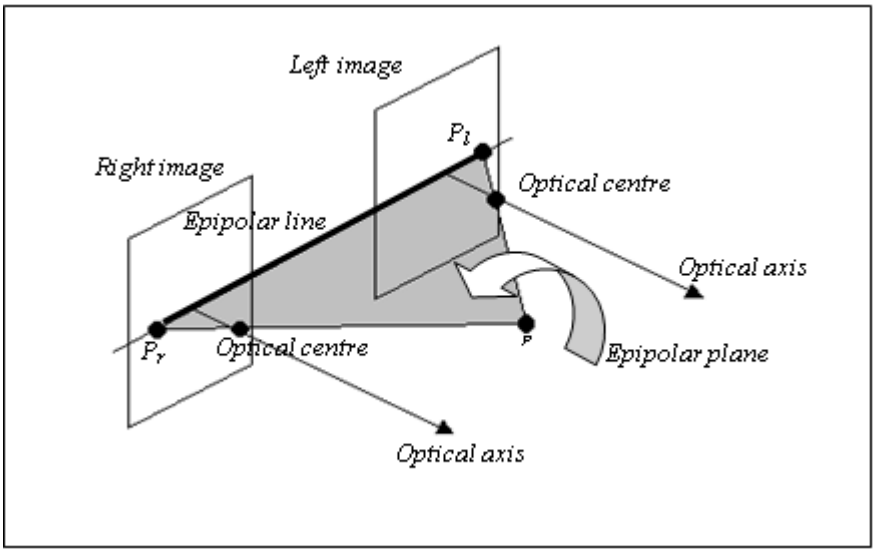


Figure 2.15 Definition of the epipolar plane

Another type of conventional converging stereovision system is shown in Figure 2.16. There are two epipolar lines, one in each of the two image planes (ep and $e'p'$). . In this configuration, an epipolar line is not along a horizontal scan-line, but inclined at an angle to it. The search of a corresponding point in the left image (say), is along the epipolar line $e'p'$ at the right image, and vice-versa. Searching the corresponding point on an inclined line is more tedious and time consuming as compared to the case of two coplanar image planes. However, the epipolar constraint still reduces the correspondence searching from whole the image to a 2-dimensional line.

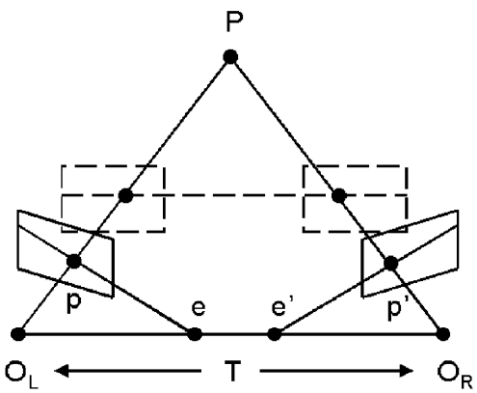


Figure 2.16 The geometry of converging stereo with the epipolar line (solid) and the collinear scan-lines (dashed) after rectification

2.4 Parameter and quantization analysis

The design of the conventional stereovision system is dependent on the system parameters such as focal lengths of the viewing cameras, the baseline between the cameras, distance of the viewing objects from the cameras and the interval of image sampling. It has been shown that even with perfect correspondence, the depth error still increase quadratically with the objects' depth due to the sampling interval of the hardware. The discrete nature of an imaging system causes the pixel coordinates to suffer quantization errors of up to $\pm\frac{1}{2}$ pixel. Thus, the disparity, $d = x_l - x_r$ contains error by as much as ± 1 pixel. The relationship between the stereovision system geometry and the accuracy of the depth has been studied in detailed [104]-[108] as it is of great practical importance.

Figure 2.17 shows the relationship of the depth error with respect to the baselines and object distances. Considering camera 1 and 2 capturing two objects at Z_1 and Z_2 ($Z_1 > Z_2$), it has been proven by Rodriguez [109] that the computed depth error at Z_1 (δZ_1) is larger than the depth error at Z_2 (δZ_2). They also showed a simple mathematical proof which indicate that the error increases quadratically with object distance. Consider two cameras with focal length f separated by distance b_1 and let d be the disparity in x coordinates of a pair of correspondence points, the depth equation Z is given by,

$$Z = \frac{\mathrm{b}_1 f}{d} \tag{2.7}$$

Thus, the depth error can be written in terms of the disparity error (δd) as:

$$\partial Z = \frac{\mathrm{b}_1 f}{d} - \frac{\mathrm{b}_1 f}{d+\partial d} = \frac{Z^2 \partial d}{\mathrm{b}_1 f + Z\partial d} \approx \frac{Z^2}{\mathrm{b}_1 f}\partial d \tag{2.8}$$

The last step is obtained by taking first order Taylor series approximation about $\partial d = 0$. Furthermore, comparing camera 1 and 2 with smaller baseline (b_1) with camera 1 and 3 with larger baseline (b_2), while keeping the object distance constant at Z_1, the computed depth error (δZ_1) using the former setup is higher than that of the latter setup (δZ_2).

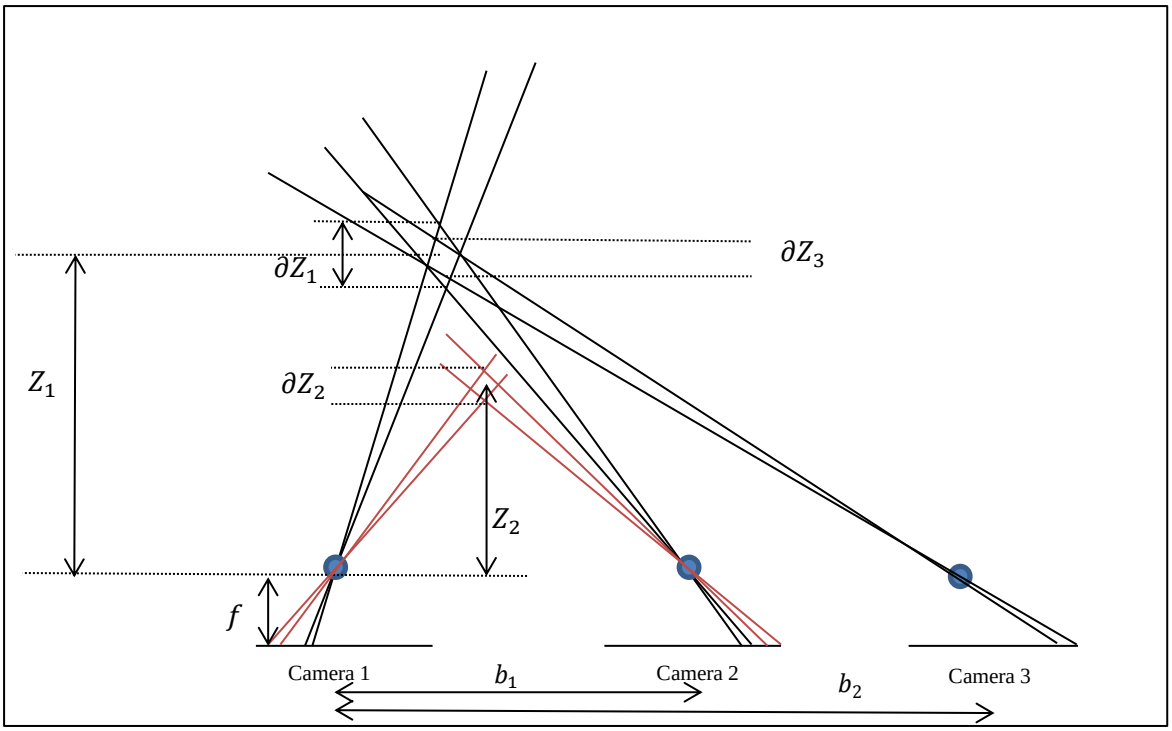


Figure 2.17 Depth error analysis of conventional stereovision

In other words, the accuracy of a stereovision system for long range is usually low while the accuracy in near range is much better. Besides, larger baseline tends to reduce the error of the system. The sources of error such as focal length error, quantization error and camera rotation error have been studied in [110]. Blostein et al. [106] proposed a mathematical model in terms of disparity value such that the probability of the percent range error is less than a certain value. It was found that the relative depth error is much better indicator as compared to the percent range error in designing a suitable stereovision system setup. Mcvey et al. [108], Verri et al. [110] and Matthies et al. [111] have provided a very detailed study on stereo quantization error. Based on their studies, Rodriquez et al. [109] derived the probability density function of the range estimation error and the expected value of the range error magnitude in terms of the design parameters of the stereo imaging setup. They gave experimental results to support the theoretical model proposed by them. Consequently, the study has been improved and extended in the recent years [112]-[113].

This single-lens bi-prism based stereovision system has been studied [16]-[17] in the past but none of them made any analysis on the quantization error and parameter analysis to estimate the range error. In this thesis, we conduct a comprehensive study on the quantization error and parameter analysis of the single-lens bi-prism stereovision system. Besides, based on the parameters of this system, we establish a new understanding of how the FOV of the system is affected by the bi-prism angle, as well as the position of the bi-prism with respect to the camera. The former will be looked at in detail, encompassing both divergent and convergent systems while the focus of the latter will be on divergent systems only.

2.5 Summary

In this chapter, we have reviewed various single-lens stereovision system and stereovision algorithms, which include camera calibration, stereo correspondence algorithms, epipolar geometry constraints, and parameter analysis. Based on the knowledge gained from the literatures, this thesis provides a comprehensive study on virtual camera calibration, stereo correspondence based on epipolar geometry and parameter analysis encompassing the single-lens bi-prism based stereovision system. The following chapters will describe these approaches in detail.

Chapter 3 Virtual Camera Calibration

To solve the stereo correspondence problem of the single-lens system, virtual camera calibration such as determining the extrinsic and intrinsic parameters of the virtual cameras is required. For this particular single-lens system, both the virtual cameras are formed by using a single charge coupled device (CCD) camera with the aid of a bi-prism. In this chapter, the concept of virtual cameras is introduced and then a geometrical approach to calibrate the virtual cameras is presented. The results are compared to the Zhang's [68] calibration method.

Section 3.1 and 3.2 will present the formation of the virtual camera and the concept of the proposed geometrical approach respectively. Experiment results will be shown in section 3.3 and summary is done in section 3.4.

3.1 Formation of virtual camera

The system presented in this thesis achieves stereovision effect with the aid of a CCD camera and a bi-prism positioned in front of the camera. This system, as schematically presented in Figure 3.1, is known as single-lens bi-prism stereovision system. As mentioned previously, this set-up is able to simultaneously capture two views of the same scene to form one stereo image pair. An example of a stereo image pair is shown in Figure 3.2.

This stereo image pair can be taken as images captured using two identically and symmetrically positioned virtual cameras that do not physically exist. The left (right) half of the image corresponds to the right (left) virtual camera image plane. In other words, this pair of stereo images is equivalent to the two images captured using a conventional two-camera system. Since this system uses the concept of virtual cameras, it is also referred to as virtual stereovision system. Figure 3.3 depicts a 2-D representation of this system.

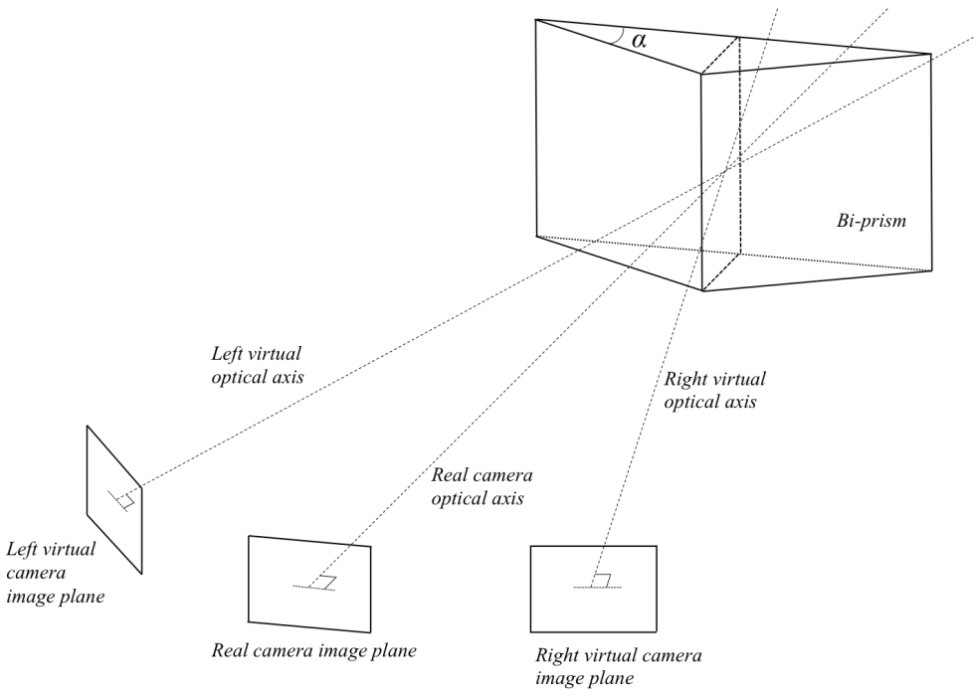


Figure 3.1 3-D schematic diagram of single-lens stereovision using a bi-prism

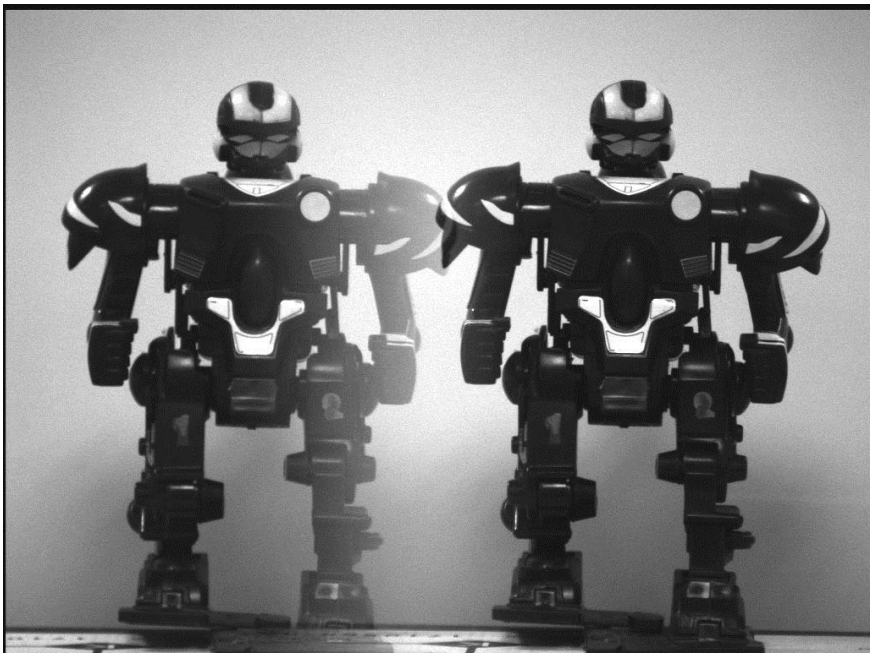


Figure 3.2 An example of stereo-image pair taken by a CCD camera through a 6.4° bi-prism

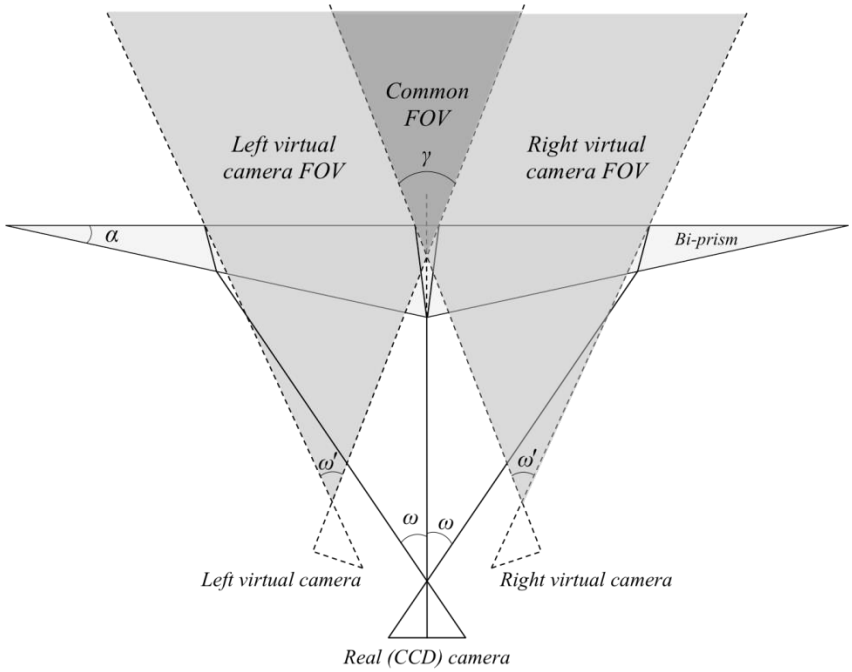


Figure 3.3: Single-lens bi-prism stereovision system showing the virtual cameras and their FOVs

As shown in Figure 3.3, there is an area of overlap region between the left and right virtual camera field-of-views (FOVs) known as the *common FOV*. This region defines an area in which an object can be 'captured' by *both* left and right virtual cameras and thus be fully projected on the real camera image plane. It is also a region in which an object must be placed in order for correspondence searching to be possible. As the viewpoints of these two virtual cameras are different, the two images captured will also be different. This difference can be used to determine the disparity of the points of interest located in the common view zone of the two virtual cameras. Henceforth, stereovision effect is achieved and thence the 3D reconstruction of a given scene is made possible with this setup.

Lim and Xiao [16]-[17] have first proposed a novel geometrical method to compute the position and orientation of the two virtual cameras. More details were discussed in Section 2.2.2. In short, their method is simple, fast, and all geometrical derivations are done in 2-dimensional. However, there are some limitations such as the field of view and sensor size of the camera have to be known (refer to section 2.2.2). Thus, in this chapter, a 3-dimensional geometrical approach is proposed to model the virtual cameras and solve the above mentioned problems.

3.2 Virtual camera calibration based on the proposed geometrical approach

In this section, a geometrical approach to recover the basic properties of the virtual cameras such as optical centres, focal length and orientation is proposed. The virtual cameras will be modelled using the pinhole camera concept. The following conditions must be fulfilled:

1) The real image plane of the camera has consistent properties, such as, pixels size, resolution and the sensitivity of the photoreceptors;
2) The bi-prism is exactly symmetrical with respect to its apex line;
3) The projection of the bi-prism apex line on the camera image plane will bisect the image plane equally and vertically, and;
4) The back plane of the bi-prism is parallel to the real image plane;

The detailed derivation will be presented in Section 3.2.1-3.2.3. The summary of the algorithm is shown below:

Step 1: **<u>Compute the virtual cameras' optical centres. (refer to Figure 3.4)</u>**

a) Choose two random image points, P_a and P_b on the right image plane of the CCD camera.

b) Back project the two points to the left virtual camera image plane by computing the 3D geometrical rays, $\overrightarrow{R1}_a, \overrightarrow{R2}_a, \overrightarrow{R3}_a, \overrightarrow{R1}_b, \overrightarrow{R2}_b \; and \; \overrightarrow{R3}_b$.

c) Determine the intersection point of the back projection rays of $\overrightarrow{R3}_a$ and $\overrightarrow{R3}_b$.

The intersection point of the back projection rays of $\overrightarrow{R3}_a$ and $\overrightarrow{R3}_b$ indicates the optical centre of the left virtual camera.

Step 2: **<u>Compute the orientations of the virtual cameras with respect to the world coordinate system (refer to Figure 3.7).</u>**

a) Identify one unique point (centre point of the CCD camera image plane), P_o of which its back projection ray of $\overrightarrow{R3}_o$ indicates the optical axis of the left virtual camera.

b) Compute the back-projection ray of $\overrightarrow{R3}_o$ which is normal to the virtual camera image plane.

The orientation of the left virtual camera can be found by computing the back-projection ray of $\overrightarrow{R3}_o$.

Step 3: **<u>Compute the focal length of the virtual cameras (refer to Figure 3.8).</u>**

a) By using the centre image point, P_o and one random image point, P_c on the CCD camera image plane, compute their back-projection rays of $\overrightarrow{R3}_o$ and $\overrightarrow{R3}_c$ to the left virtual camera.

b) Compute the angle of the back projection rays of $\overrightarrow{R3}_o$ and $\overrightarrow{R3}_c$.

c) By using the right angle triangle theorem, compute the focal length geometrically as shown in equations (3.16)-(3.17).

3.2.1 Computation of the virtual cameras' optical centres

As shown in Figure 3.4, in order to obtain the left virtual camera's optical centre, two random image points, $P_a(x_a, y_a, -f), P_b(x_b, y_b, -f)$ on the right image plane of the CCD camera are chosen and the geometrical rays of the image points must pass through the optical

centre of the CCD camera based on the pin-hole camera concept. These two rays which connect the two image points and the CCD camera's optical centre are refracted twice through the bi-prism forming a series of angles ($\phi 1_a$, $\phi 1'_a$, $\phi 2_a$, $\phi 2'_a$, $\phi 1_b$, $\phi 1'_b$, $\phi 2_b$, and $\phi 2'_b$) with their respective normal lines at the refraction points as shown in Figure 3.5. The back-projection of these refracted rays will intersect the image plane of the left virtual camera (LVC). The intersection point of these two back projection rays indicates the position of the left virtual camera's optical centre as shown in Figure 3.4. To derive the properties of the virtual cameras, the following parameters are assumed to be known:

α, angle of the bi-prism;

n, refractive index of the bi-prism;

T_o, distance between the real camera's optical centre to the apex of the bi-prism;

T, distance between the apex of the bi-prism to the back plane of the bi-prism;

I, sensor size of the real camera.

Definition of the coordinate system:

(X^{vc}, Y^{vc}, Z^{vc}), virtual camera coordinate system;

(X^{w}, Y^{w}, Z^{w}), world coordinate system.

These coordinate systems and parameters will be used throughout the chapter unless specified otherwise.

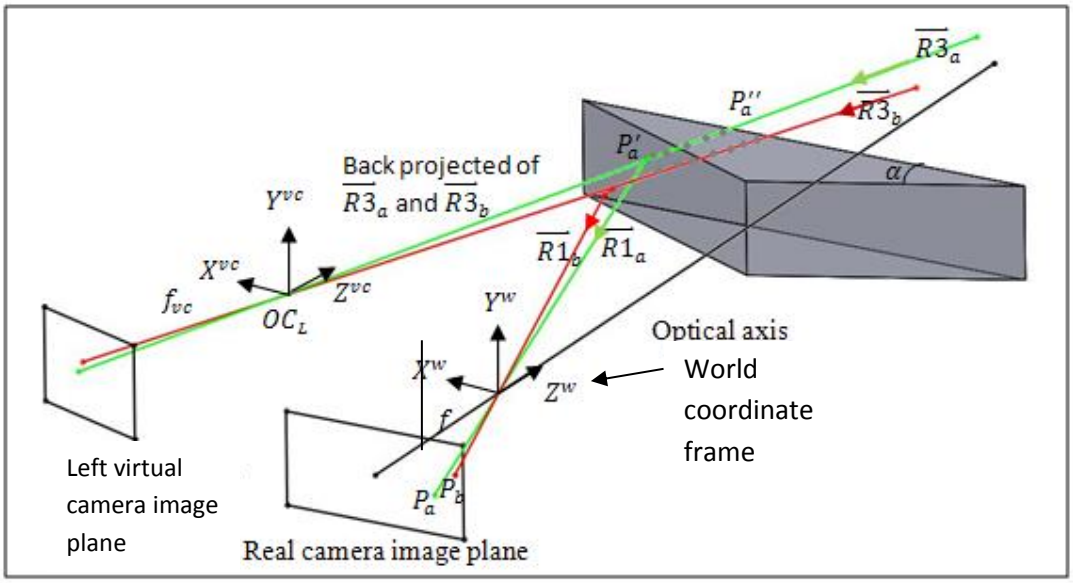


Figure 3.4 Computing the virtual camera's optical centre

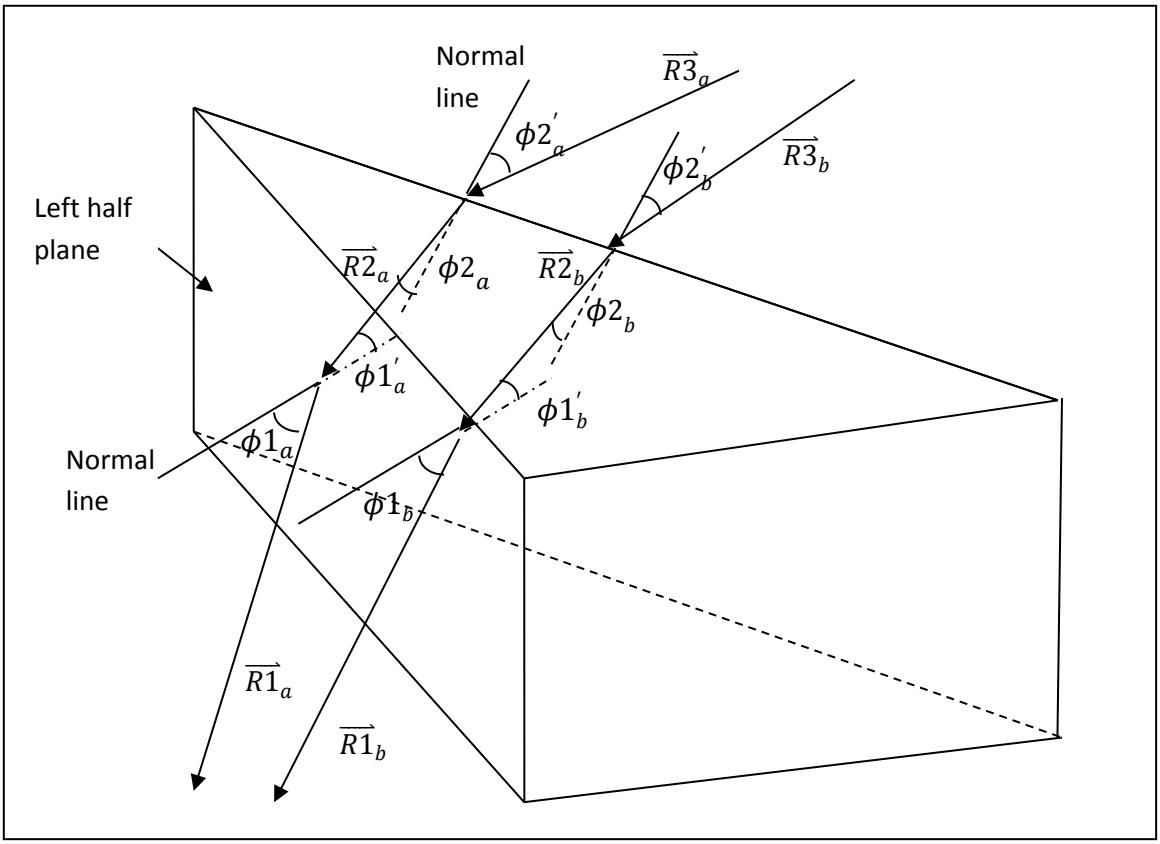


Figure 3.5 Illustration of the incident and refracted angles

All the vectors, points and planes are expressed in the world coordinate system; otherwise it will be mentioned in the text. According to the law of refraction (Appendix A), the following expression can be obtained:

$$n = \frac{\sin \phi 1_a}{\sin \phi 1'_a} = \frac{\sin \phi 2'_a}{\sin \phi 2_a} \tag{3.1}$$

The vector equation for $\overrightarrow{R1}_a$:

$$\overrightarrow{R1}_a = g(x_a\, i + y_a j - fk) \tag{3.2}$$

where g is an arbitrary scalar of the vector $\overrightarrow{R1}_a$.

The equation of the left half plane of the bi-prism is known which is:

$$X - Z\cot\alpha + T_o \cot\alpha = 0 \tag{3.3}$$

where X and Z are the coordinates of points on the bi-prism half plane.

The point P'_a shown in Figure 3.4 is the intersection point between the biprism half planes and the vector $\overrightarrow{R1}_a$:

$$P'_a = \left[\frac{-T_o x_a \cot\alpha}{f\cot\alpha + x_a}, \frac{-T_o y_a \cot\alpha}{f\cot\alpha + x_a}, \frac{T_o f \cot\alpha}{f\cot\alpha + x_a}\right] \tag{3.4}$$

The ray $\overrightarrow{R1}_a$ will be refracted twice through the bi-prism in 3D space. Thus, to simplify the derivation of the rays $\overrightarrow{R2}_a$ and $\overrightarrow{R3}_a$, we will use the coordinate system transformation to transform the rays to 2D. When the ray $\overrightarrow{R1}_a$ is refracted to the ray $\overrightarrow{R2}_a$, both the rays and the normal vector at the point of refraction are coplanar (plane A) based on the Law of Refraction. A frame A is assigned on plane A with y^A-axis normal to the plane and x^A-axis is the normal vector as shown in Figure 3.6, super script A denotes the frame A coordinate system and F^A is a 3 by 3 rotational matrix of frame A with respect to the world coordinate frame. When the ray $\overrightarrow{R1}_a$is expressed in the frame A, its y-component will be reduced to zero since the ray $\overrightarrow{R1}_a$ lies on plane A.

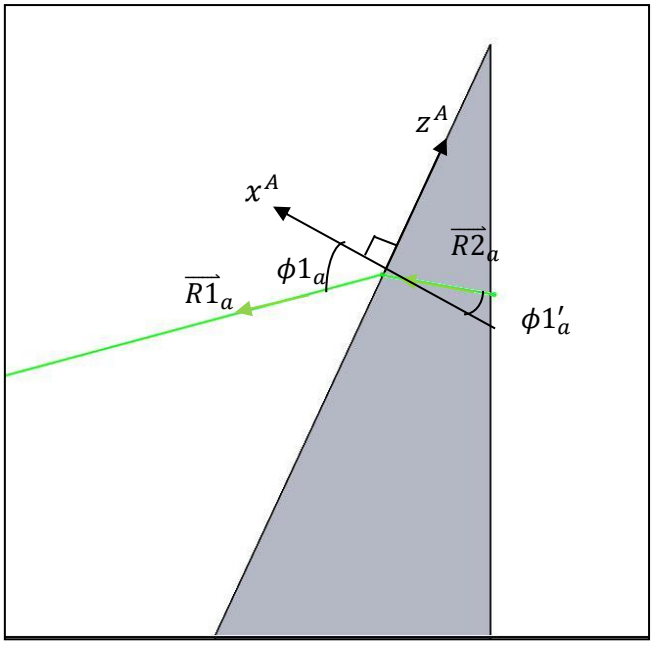


Figure 3.6 Coordinate system of frame A

$$\overrightarrow{R1}_a^A = F_A^T \overrightarrow{R1}_a \tag{3.5}$$

where $\overrightarrow{R1}_a^A$ is the direction vector of the ray $\overrightarrow{R1}_a$ expressed in frame A which can also be written as $[\cos\ \phi 1_a\, i + \sin\ \phi 1_a\, k]$ in vector form, $\phi 1_a$ is the incident angle of the ray $\overrightarrow{R1}_a$.

$$\phi 1_a = sin^{-1}\left(\widehat{R1}_a^A(3)\right) \tag{3.6}$$

The refracted angle:

$$\phi 1'_a = sin^{-1}\left(\frac{1}{n} sin\ \phi 1_a\right) \tag{3.7}$$

The direction vector of $\overrightarrow{R2}_a^A = [cos\, \phi 1'_a\, i + sin\, \phi 1'_a\, k]$ (3.8)

Expressed in world coordinate frame,

The direction vector of $\overrightarrow{R2}_a = F^A \overrightarrow{R2}_a^A$ (3.9)

When the ray $\overrightarrow{R2}_a$ is refracted to $\overrightarrow{R3}_a$, both the rays together with the normal vector are again coplanar. A frame B is assigned to the plane. By repeating the same derivation from (3.6)-(3.9), we can obtain the point P_a'' and the direction vector of the ray $\overrightarrow{R3}_a$. The full expression of the direction vector of $\overrightarrow{R3}_a$ after simplification is shown as below:

$$\text{The direction vector of } \overrightarrow{R3}_a = F^B \begin{bmatrix} -(1-n^2K^2)^{0.5} \\ 0 \\ nK \end{bmatrix} \tag{3.10}$$

where $K = F_{i,3}^{B\,T}\left(\left(1-\frac{1}{n^2}\left(F_{i,3}^{A\,T}\widehat{R1}_a\right)^2\right)^{0.5} F_{i,1}^{A} + \frac{1}{n}F_{i,3}^{A\,T}\widehat{R1}_a F_{i,3}^{A}\right)$ and F^B is a 3 by 3 rotational matrix of frame B with respect to the world coordinate frame.

The back extension of the ray $\overrightarrow{R3}_a$will pass through the optical centre and reach the image plane of the LVC as shown in Figure 3.4. By repeating the same procedure for point P_b, the back extension ray of $\overrightarrow{R3}_b$ will also go through the optical centre and reach the image plane of the LVC. The optical centre of the LVC (OC_L=$[OC_{Lx}, OC_{Ly}, OC_{Lz}]^T$) can be recovered by computing the intersecting point of the back extension rays of $\overrightarrow{R3}_a$and $\overrightarrow{R3}_b$. As the rays $\overrightarrow{R3}_a$ and $\overrightarrow{R3}_b$ are most likely two skew lines in 3D space, they may not intersect each other at a physical point. We will use the mid-point theorem (Appendix C) to compute the optical centre. The translation matrix, T_l can be obtained by computing the distance vector of the optical centre from the origin of the world coordinate system.

3.2.2 Computation of the virtual cameras' orientation

To obtain the orientation of the virtual camera, the optical axis of the virtual camera is required. By projecting point P_o (0,0,-*f*) which is the centre point of the real camera image plane that lies on the optical axis of the CCD camera into 3D space, we will obtain the ray $\overrightarrow{R1}_o$ as shown in Figure 3.7. The ray $\overrightarrow{R1}_o$ is refracted into two rays ($\overrightarrow{R2}_o$and $\overrightarrow{R2'_o}$) after going through the bi-prism apex.

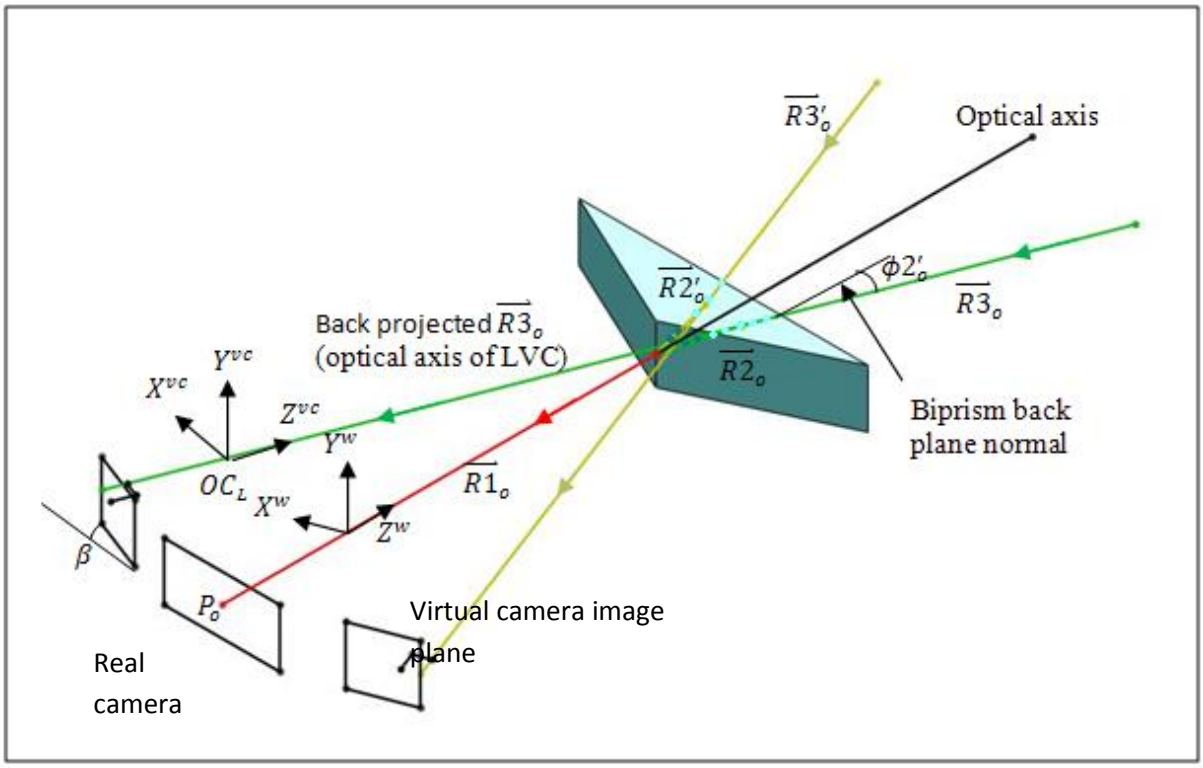


Figure 3.7 Geometrical rays through bi-prism (all rays lie on the X^wZ^w plane)

The ray $\overrightarrow{R2_o}$ which is the refracted ray on the right of the optical axis will be refracted again and becomes the ray $\overrightarrow{R3_o}$ after going through the biprism back plane. The back-extension ray of $\overrightarrow{R3_o}$ represents the optical axis of the LVC is shown in Figure 3.7. According to law of refraction, we have

$$n = \frac{\sin \phi 2'_o}{\sin \phi 2_o} = \frac{\sin \phi 1_o}{\sin \phi 1'_o} \tag{3.11}$$

where $\phi 1_o, \phi 1'_o, \phi 2_o$ *and* $\phi 2'_o$ are series of refraction angles on the bi-prism half and back planes. We can obtain the following relationship geometrically:

$$\beta = \phi 2'_o \tag{3.12}$$

β denotes the rotation angle of the LVC about the Y^w-axis (angle between the left virtual camera image plane (LVCIP) with the X^wZ^w plane). There is no rotation angle about the X^w and Z^w axis since the apex of the bi-prism is placed along the optical axis of the CCD camera. The left and right half planes of the bi-prism are symmetrical about the Y^wZ^w plane.

The orientation of the LVCIP can also be found by computing the ray $\overrightarrow{R3}_o$ (optical axis of LVC) as it is the normal vector of the LVCIP. The orientation of the LVCIP also represents its rotational matrix with respect to the world coordinate frame.

3.2.3 Computation of the virtual cameras' focal length

The focal length of the virtual cameras, f_{vc} should be the same as the focal length of the real CCD camera as the two virtual cameras are formed based on the CCD camera. They should contain the same intrinsic parameters assuming that the distortion through the bi-prism is negligible. However, f_{vc} will still be confirmed geometrically.

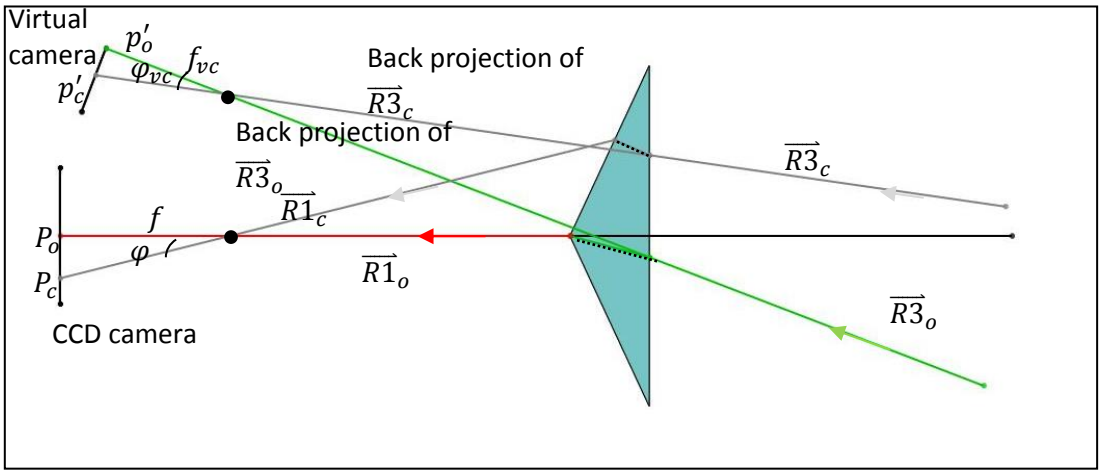


Figure 3.8 Derivation of virtual camera focal length, f_{vc}

As shown in Figure 3.8, the line joining the point P_o (the centre of the image plane) and the optical centre of the CCD camera will form the ray $\overrightarrow{R1}_o$ which will be refracted twice forming the ray $\overrightarrow{R3}_o$. The back projection ray of $\overrightarrow{R3}_o$ will go through the virtual camera optical centre (which have been recovered earlier) and intersect the LVCIP at a point, p'_o. Picking another arbitrary point P_c on the real CCD camera, the projection of P_c into 3D space will form the ray $\overrightarrow{R1}_c$ which will go through the optical centre of the CCD camera. The ray $\overrightarrow{R1}_c$ will undergo two refractions forming the ray $\overrightarrow{R3}_c$. The back projection ray of $\overrightarrow{R3}_c$ will go

through the LVC optical centre and intersect the LVCIP at a point, p_c'. Lim and Xiao [16]-[17] has demonstrated that the field of views of the virtual cameras are half of the field of view of the CCD camera. This is because the image plane of the CCD camera is divided into two sub-images which correspond to the two virtual cameras. In other words, all the pixels on the right half plane of the real CCD image plane correspond to the same pixels on the LVC. Thus,

$$\|\overrightarrow{P_oP_c}\| = \|\overrightarrow{p_o'p_c'}\| \tag{3.13}$$

$$cos\,\varphi = \frac{\overrightarrow{R1}_o \cdot \overrightarrow{R1}_c}{\|\overrightarrow{R1}_o\|\|\overrightarrow{R1}_c\|} \tag{3.14}$$

Since the optical centre of the CCD camera, the points P_o and P_c form a right angle triangle, we can obtain the following relationship:

$$f = \frac{\|\overrightarrow{P_oP_c}\|}{tan\,\varphi} \tag{3.15}$$

Consider the back projection rays on the virtual camera,

$$cos\,\varphi_{vc} = \frac{\overrightarrow{R3}_o \cdot \overrightarrow{R3}_c}{\|\overrightarrow{R3}_o'\|\|\overrightarrow{R3}_c'\|} \tag{3.16}$$

$$f_{vc} = \frac{\|\overrightarrow{p_o'p_c'}\|}{tan\,\varphi_{vc}} \tag{3.17}$$

In our experiment, the results (Section 3.3.2) show that $f \approx f_{vc}$.

3.3 Experimentation and Discussion

Figure 3.9 shows our system setup which is used to capture images. Experiment was performed with a CCD colour camera. The experiment was designed to test the correctness and accuracy of the proposed method. The first step was to make sure that the camera and bi-prism were positioned correctly. This was done by using a precise customized mechanical stand as shown in Figure 3.9. The accuracy of the mechanical stand is up to 0.01mm. Then, the virtual cameras are calibrated using both the conventional calibration method and the

proposed geometrical method. The values of the system parameters are shown in Table 3.1. $\alpha, T,$ and n are the prism properties which were provided by the manufacturer while T_o is recorded using the precise motorized stand. The rest of the parameters (f, C_x and C_y) are the real camera parameters which are provided by the manufacturer and verified using Matlab Calibration Toolbox [35] . All these parameters were used to recover the extrinsic and intrinsic parameters of the virtual cameras.

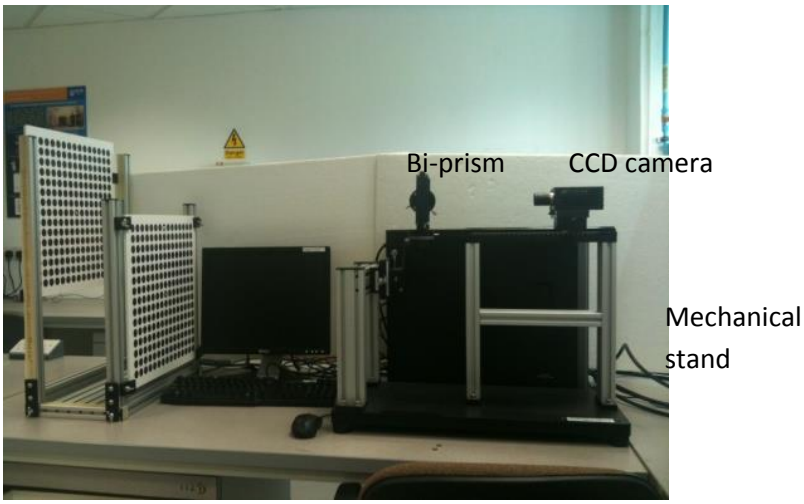


Figure 3.9 System setup used in the experiment

Table 3.1 The values of parameters of the system used in the experiment

α	n	f	T	T_o	C_x	C_y
6.4°	1.48	25 *mm*	5.6 *mm*	170 *mm*	512	384

3.3.1 Conventional calibration method analysis

In order to implement the Zhang [68] calibration method, all the images taken using this system have to be cut into two sub-images as shown in Figure 3.10. The calibration board is positioned 2m in front of the CCD camera. The optical centre of the CCD camera was chosen as the origin (0,0,0) throughout the entire experiment. The corner of the checkerboard on the left sub-image was extracted as shown in Figure 3.11. This was repeated for the right sub-image. In the experiment, we found that the most optimum number of images used for

calibration is 12 images. Thus, the procedures above were repeated for all the 12 images captured under different orientations and positions.

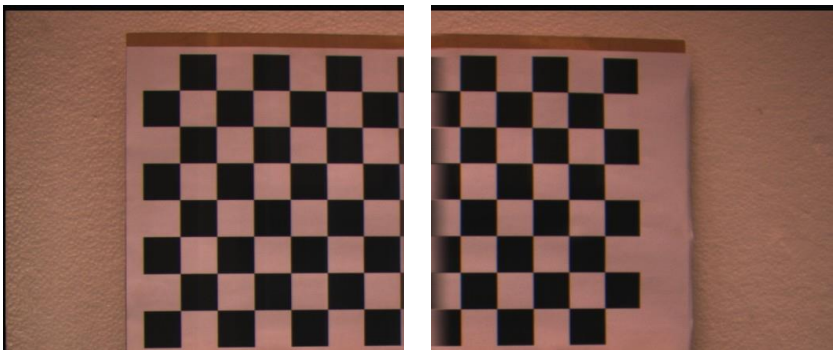


Figure 3.10 Calibration board captured can be divided into two sub-images

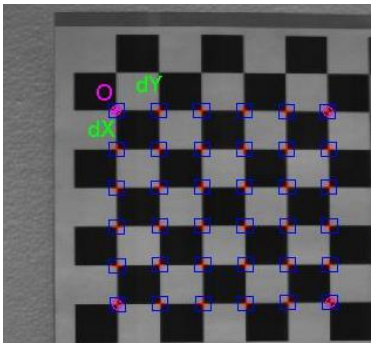


Figure 3.11 Corner extraction of the calibration board for calibration

Once the corners for all the images (left and right sub-images) are extracted, Zhang [68] calibration technique was used to recover both the extrinsic and intrinsic parameters of the left and right virtual cameras. The results were refined using the Levenberg-Marquardt non-linear squares algorithm [114]. The geometry of the virtual cameras is shown in Figure 3.12.

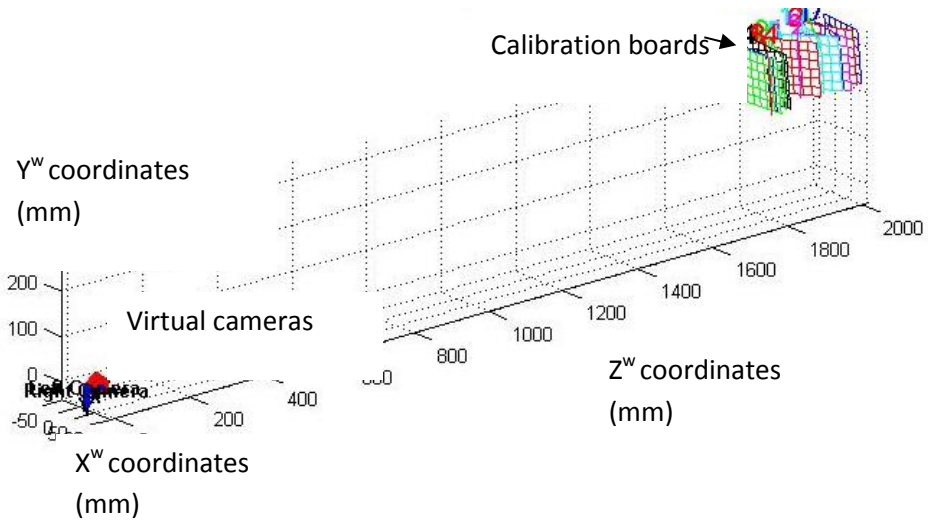


Figure 3.12 Extrinsic rig of the virtual cameras and the orientation of the calibration boards

After implementing Zhang's [68] calibration methods, the results for the extrinsic and intrinsic parameter of the virtual cameras are shown below:

Extrinsic parameters (position of the right virtual camera with respect to the left virtual camera):

Rotation angle about the principal axis, $R = \begin{bmatrix} 0.9557 & 0.0181 & -0.2937 \\ -0.0237 & 0.9996 & -0.0161 \\ 0.2933 & 0.0217 & 0.9558 \end{bmatrix}$

Translation vector, T_r = [-18.75111 -0.16810 -1.05143] ± [6.20638 3.39150 20.48864]

Focal length of the virtual camera, f_{vc} = 22mm.

As shown in the results above, the standard deviation of the results using the Zhang's method is quite high. This is because the Zhang's method is proposed for the conventional two camera system where the two cameras have larger baseline and rotational angles. On the other hand, the baseline and rotational angles of the virtual cameras of this single-lens system depend on the angle of the bi-prism (α). Smaller α which leads to smaller baseline and and larger field of view is preferred (which will be covered in detailed in Chapter 5). Thus, conventional approaches may not be accurate in calibrating the virtual cameras of the single-lens system.

3.3.2 Experimental results of the proposed geometrical approach

Based on the proposed geometrical approach in Section 3.2, the LVC and RVC are rotated about the Y^w axis at -0.0538 radian and 0.0538 radian which are symmetrical about the $Y^w Z^w$ plane. The optical centres of the LVC and RVC which are found using 2 random image points are (9.167mm, 0mm, -1.7066mm) and (-9.167mm, 0mm, -1.7066mm). The extrinsic parameters of the virtual cameras are:

$$R_R = \begin{bmatrix} cos\,0.0538 & 0 & sin\,0.0538 \\ 0 & 1 & 0 \\ -\,sin\,0.0538 & 0 & 0\,cos\,0.0538 \end{bmatrix}, T_R = \begin{bmatrix} -9.167 \\ 0 \\ -1.7066 \end{bmatrix}$$

$$R_L = \begin{bmatrix} cos\,0.0538 & 0 & -sin\,0.0538 \\ 0 & 1 & 0 \\ sin\,0.0538 & 0 & cos\,0.0538 \end{bmatrix}, T_L = \begin{bmatrix} 9.167 \\ 0 \\ -1.7066 \end{bmatrix}$$

$$f_{vc} = 24.4\ mm$$

where subscripts L and R denote left and right virtual cameras.

At this stage, it is not possible to compare the results between the conventional calibration approach and the proposed geometrical approach. This is because the calibration results are the extrinsic parameters (position and orientation) of the two virtual cameras that do not exist physically. We are unable to determine the exact position and orientation of the virtual cameras. In other words, the accuracy of the calibration results using both the methods is unknown. However, the results obtained in this chapter will be used to solve the stereo correspondence problem in the Chapter 4. More detailed comparison between the methods will be discussed in the next chapter.

Recalling the proposed geometrical approach in section 3.2, any two arbitrary image points can be used to compute the optical centre of the virtual camera. However, due to lens distortion and other source of errors, different image points chosen to compute the optical centre might affect the accuracy of the result. Thus, to further improve the result of the virtual

camera's optical centre, the intersection points of all the back extensions rays of the CCD camera image points with the optical axis were computed (see Figure 3.13). Based on the pin-hole camera concept, all the rays must intersect the optical axis at the same point which is also the optical centre of the virtual camera.

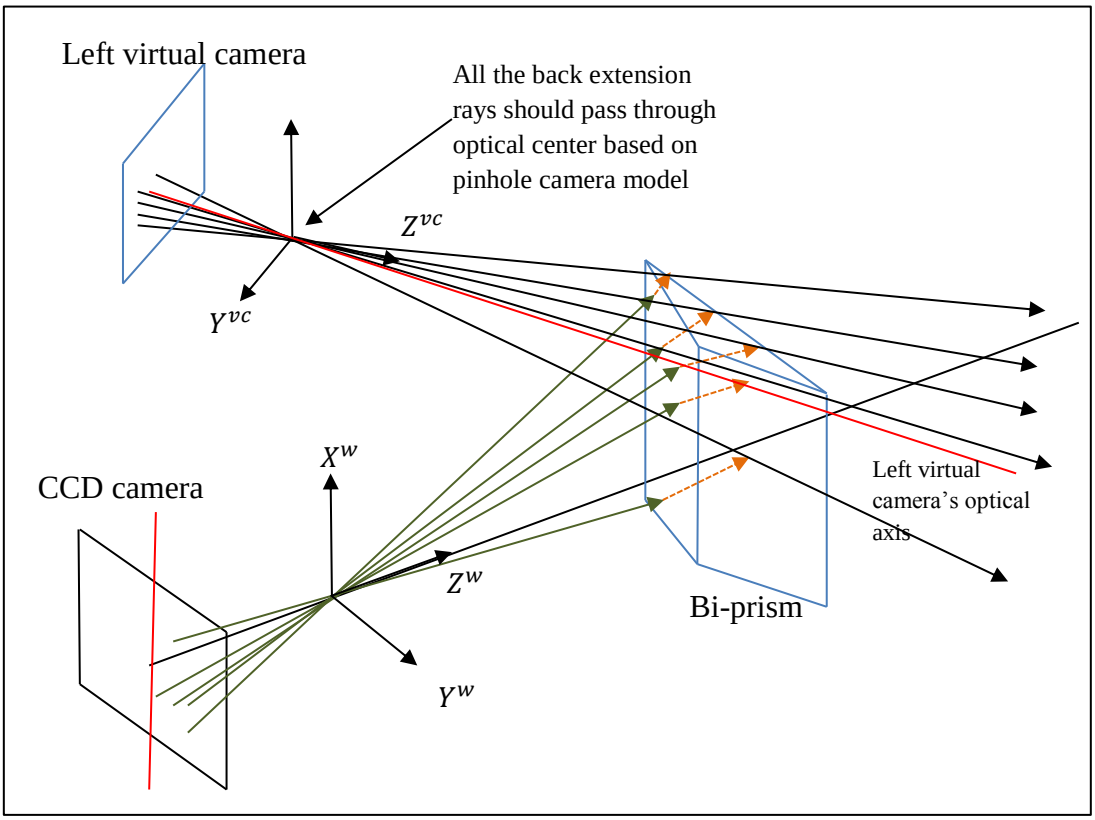


Figure 3.13 Computing the optical centre using all the image points

Figure 3.14 shows the coordinates of the optical centre computed using all the image points of the left virtual camera. There is some deviation of the computed optical centre if different image points are selected. To examine further, the *x*,*y* and *z* coordinates are plotted separately as shown in Figures 3.15-3.17. The *xy* plane of all the three figures corresponds to the right sub-image points on the CCD camera while the z-axis are the values of the x, y and z coordinates of the optical centres. The deviations became larger when the image points are further away from the centre. This could be due to quantization errors, distortion and mathematical errors which increase from the centre of the image. However, the deviations for all the image points are small and insignificant. Thus, we can conclude that all the rays are

converged to a single point (optical centre of the virtual camera). Geometrical approach can be used to approximate the parameters of the virtual cameras and the pin-hole camera theory can be applied on the virtual cameras as well. Besides, the accuracy of the optical centre computation can be improved by choosing the mean of the coordinates which are calculated using all the image points as mentioned above. However, to reduce the computation cost, we can choose two arbitrary points (Section 3.2) which are close to the centre of the image plane. The result is very close to the mean value calculated using all the image points.

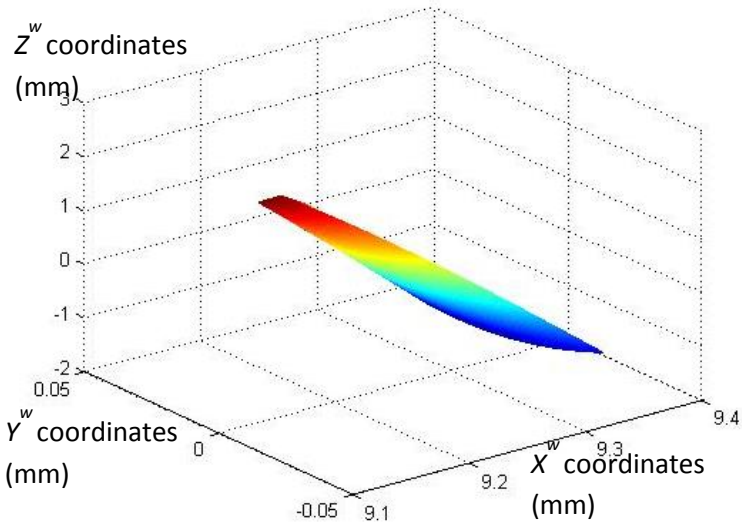


Figure 3.14 Optical centre coordinates computed from all the pixels (512 x 384 pixels)

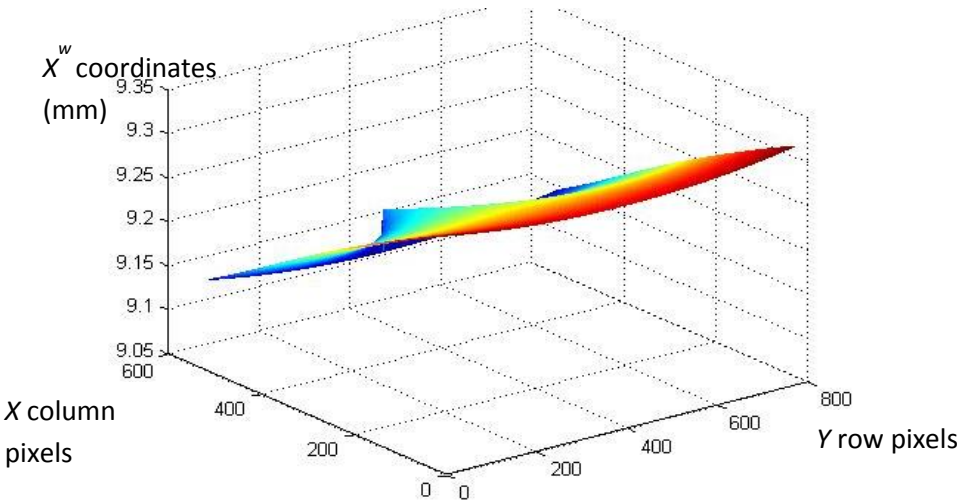


Figure 3.15 *x* coordinates of the computed optical centers, range: 9.1343-9.3204mm, mean = 9.2345mm, std = 0.0474mm

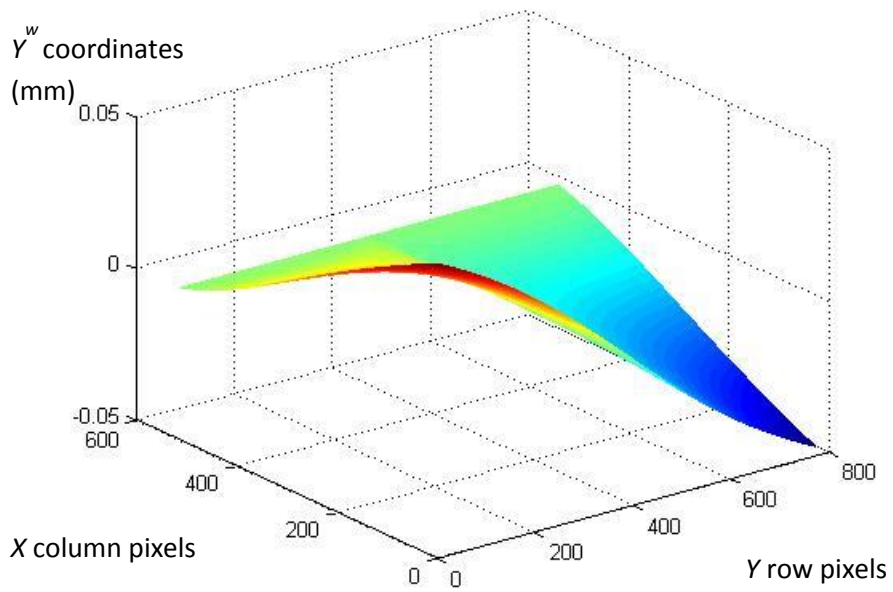


Figure 3.16 *y* coordinates of the computed optical centers, range: -0.047 -0.047mm, mean=0.00003mm, std = 0.0179mm

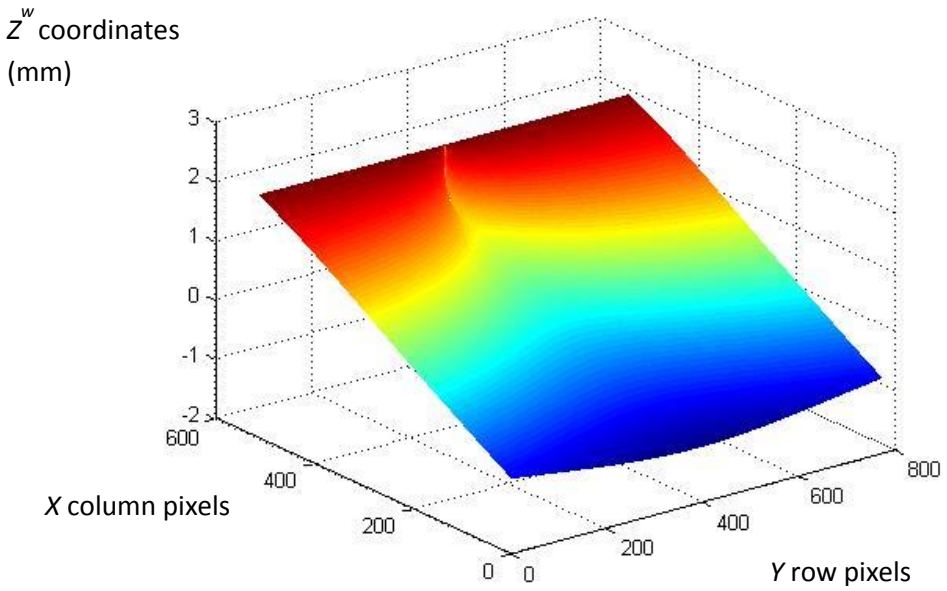


Figure 3.17 *z* coordinates of the computed optical centres, range: -1.0645 -2.0728mm, mean=0.4403mm, std =0.8766mm

3.4 Summary

Camera calibration is usually the first problem that needs to be solved in order to achieve the objectives of the stereovision system - 3D reconstruction or depth recovery. However, it is usually quite tedious and time consuming for real-time applications. Any changes in the camera, for example, the zooming, the calibration needs to be carried out again. Furthermore, accurate and consistent calibration parameters are hard to obtain due to the variation of the calibrated images. However, for the single-lens system discussed in this thesis, the two virtual cameras are generated from a CCD camera with the aid of a bi-prism. Thus, the two virtual cameras which are generated by one CCD camera have more consistent intrinsic parameters as compared to the conventional two-camera system. Besides, the geometrical method proposed in this chapter is developed based on the geometry of this unique system without the requirement of calibration images. Thus, the virtual cameras can be easily re-calibrated by going through the derived equations if there is any change in the system. Besides, the proposed method is also able to handle the scenario where the resolution and field of view of the system are unknown. The method proposed in [16]-[17] is to solve the calibration problem using a 2D geometrical approach in ideal situation: FOV and sensor size must be known and the back plane of the bi-prism must be parallel to the camera's image plane. All these limitations are overcome using the proposed 3D geometrical method with similar accuracy and speed.

In the next chapter, the approach of virtual epipolar line using geometrical analysis of ray sketching is proposed and discussed, which can tackle the stereo correspondence problem of the single-lens prism based stereovision system.

Chapter 4 Stereo Correspondence

In Chapter 3, we proposed a geometry-based approach to calibrate the virtual cameras which are generated using the single-lens prism based stereovision system. The second and also the most important aspect in stereovision is stereo correspondence. The final objective of stereo correspondence is to determine the disparity between the image pairs capturing the same scene. Disparity describes the difference in position of the correspondence image points on their respective image planes. The algorithms of stereo correspondence have been studied thoroughly in the past [115], they are classified into local approaches and global approaches, which have been discussed in Section 2.3. However, there are still many challenging problems needed to be solved such as the correspondence problem of the single-lens prism based stereovision system or more specifically the correspondence problem of the virtual cameras.

In this thesis, we propose a virtual epipolar line approach by making use of the geometry of the virtual cameras and epipolar geometry constraint to solve the correspondence problem of the single-lens system. Section 4.1 will present the background of an important concept in stereovision, namely epipolar geometry constraint. The proposed virtual epipolar line approach will be shown in Sections 4.2 while Section 4.3 will discuss and compare the experimental results of the proposed method and conventional approaches. Lastly, summary is presented in Section 4.4.

4.1 Background of epipolar geometry constraint

Epipolar geometry constraint which was introduced by Grew and Kak [36] is popularly used to ease the difficulties and reduce the mismatch of stereo correspondence.

There are three main concepts in epipolar geometry such as epipole, epipolar line and epipolar plane as shown in Figure 4.1.

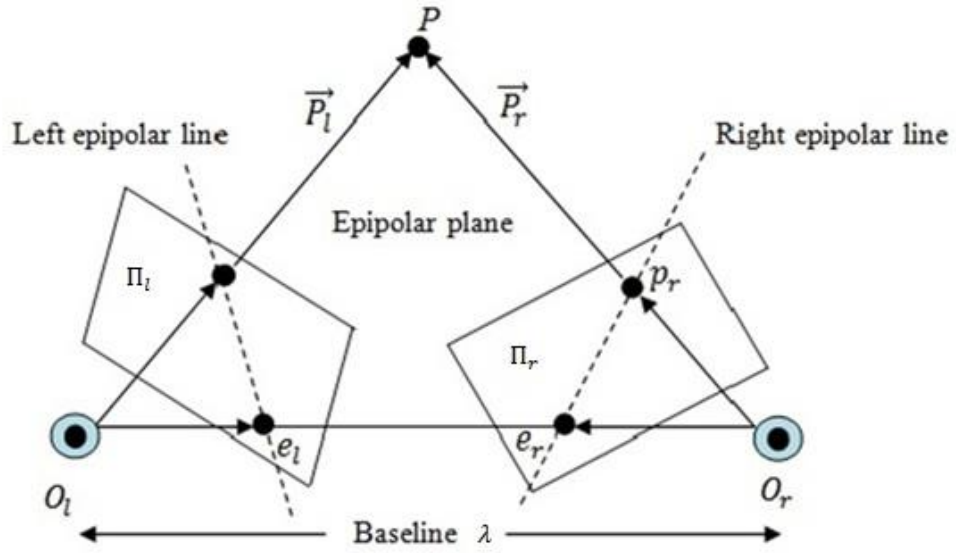


Figure 4.1: Illustration of the epipolar geometry

Figure 4.1 shows a conventional two camera system. Definitions of the symbols are shown below:

O_l and O_r,	the optical centres of the left and right cameras;
λ,	baseline or distance between the two optical centres;
Π_l and Π_r ,	the left and right camera image planes, respectively;
e_l and e_r,	the left and right epipoles, respectively.

For a 3D object point P, the points p_l and p_r are the two projection image points located on the left and right image planes, respectively. The intersection point between the line $\overrightarrow{O_l O_r}$ and the camera image planes is called *epipole* (e_l and e_r). The object point P and the two cameras optical centres O_l and O_r form a plane PO_lO_r , called *epipolar plane*. This epipolar plane intersects with the two image planes at a line, named as *epipolar line* (dash line as shown in Figure 4.1). The line e_lp_l is the left epipolar line and line e_rp_r is the right epipolar line. The concept of the epipolar geometry is defined as: for an arbitrary object point

P, given its left (right) projection image point p_l (p_r), its corresponding image point p_r (p_l) in the right (left) image must lie on the right (left) epipolar line $e_r p_r$ ($e_l p_l$). Therefore, once the epipolar geometry is established, the correspondence search will be reduced from the entire image to a two-dimensional epipolar line. Alternatively, the same information can be used to verify whether a matched candidate lies on the corresponding epipolar line. This is usually the most effective method to reject false matches.

In a conventional two camera system, there are mainly two types of stereovision system, namely the non-verged geometry and verged geometry due to the orientation of the cameras.

a) <u>The non-verged geometry of stereovision system</u>

In non-verged geometry of a stereovision system, the two image planes are coplanar and the baseline is parallel to the camera x-axis (see Figure 4.2). Under this configuration, epipolar geometry indicates that a point in space is projected to two positions on the same scan line in the left and right camera images. The different in position of the two projected points is called disparity and it only has x component $(x_r - x_l)$. Therefore, the stereo correspondence search for non-verged stereovision system is only a 1-D search along the horizontal epipolar line which is parallel to the x-axis. Due to its simplification, this configuration is commonly used to solve the stereo correspondence problem.

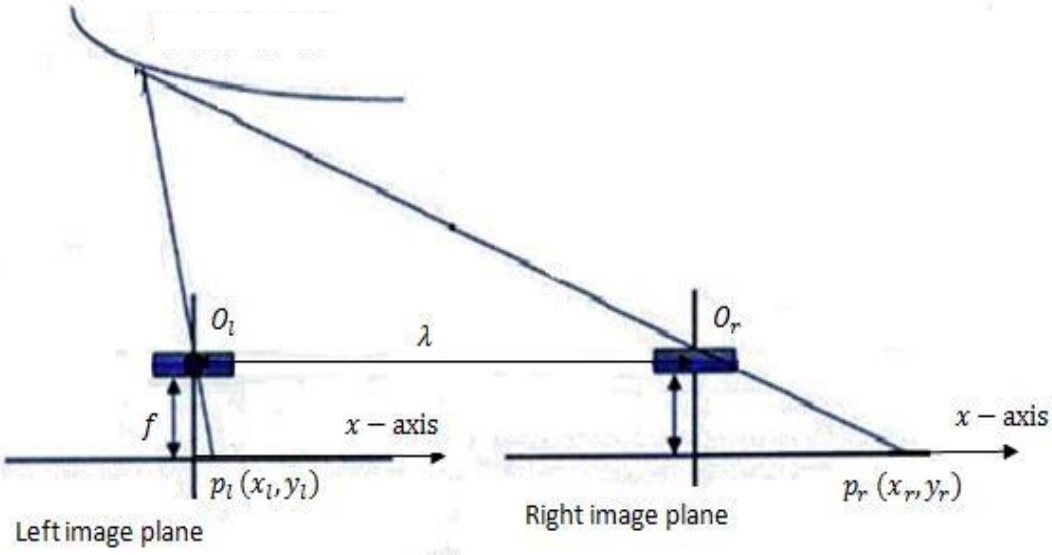


Figure 4.2: The non-verged geometry of stereovision system

b) The verged geometry of stereovision system

Although the non-verged geometry of stereovision system is preferred due to its simplicity and easy setup-up, it is impractical to establish this stereo configuration as the common field of view is very small. Due to this limitation, the geometry of the verged stereo shown in Fig. 4.3 is usually preferred. Under this configuration, the epipolar line is no longer a horizontal scan-line. The correspondence search is a 2-D search along a slanted epipolar line.

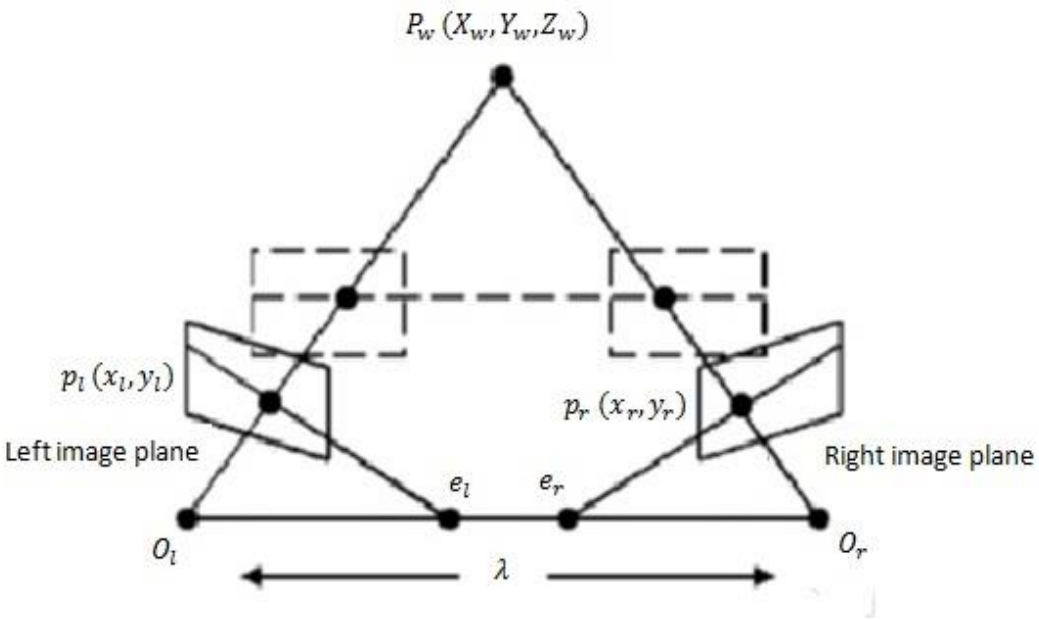


Figure 4.3: The geometry of verged stereo with the epipolar line (solid) and the geometry of non-verged stereo with epipolar line (dashed)

Epipolar geometry constraint is always utilised in order to have an efficient stereo correspondence algorithm. Trifocal Tensor, presented by Lamdan et al. [116] and Cheong [117] is an approach to determine the epipolar lines from three camera geometry. Besides, Boyer et al. [118] further improved the technique to compute the epipolar line from the relative orientation without camera calibration.

Chapter 3 has shown that the virtual cameras of the single-lens system are the type of verged geometry. We propose a virtual epipolar line approach based on the geometry of the virtual cameras to facilitate the correspondence search.

4.2 Construction of virtual epipolar lines using geometrical approach

In homogeneous coordinates, a point P^3 in 3D space is expressed in the form of (P_x, P_y, P_z, m), where m is equal to either a positive integer or zero (zero value means the point is at an infinite distance away). By definition, two parallel lines in homogeneous coordinates will intersect each other at infinity. This concept will be used when we construct

the virtual epipolar lines. To simplify the derivations, the conditions made in Section 3.2 are assumed to be true.

The coordinate systems shown in Figure 4.4 are defined as follows:

(X^w, Y^w, Z^w), world coordinate system;

$(X^{vcl}, Y^{vcl}, Z^{vcl})$, left virtual camera coordinate system;

$(X^{vcr}, Y^{vcr}, Z^{vcr})$, right virtual camera coordinate system.

The symbols used in derivations are defined as below:

(x_r, y_r), an arbitrary image points on the right image plane of the CCD camera;

f, focal length of the CCD camera;

n, refractive index of the bi-prism;

f_{vc}, focal length of the virtual camera;

β, rotational angle of the left virtual camera with respect to the world coordinate system;

OC_L/OC_R, optical centre of the left/right virtual camera;

As shown in Figure 4.4, given an arbitrary point $P_r(x_r, y_r, -f)$ on the right image plane of the real camera, we would like to find the correspondence point of P_r on the left image plane of the real camera. Since the right image plane of the real camera corresponds to the left virtual camera, P_r will be projected to the left virtual camera and the correspondence point is searched on the right virtual camera. Based on the pin-hole camera model, the incident ray ($\overrightarrow{R1_r}$) of the point P_r will go through the optical centre of the real camera and undergo two refractions to become $\overrightarrow{R3_r}$ through the bi-prism. The object point P^w in the 3D space can be any point lying along $\overrightarrow{R3_r}$ as shown in Figure 4.4. The back-extension ray of $\overrightarrow{R3_r}$ will go through the optical centre (OC_L) of the left virtual camera and intersect its image plane at a point p_r. To construct the epipolar line for the point p_r on the right virtual camera

image plane (RVCIP), two points along $\overrightarrow{R3}_r$ are chosen and projected to the RVCIP. The first selected point is the point at infinity; the second point is the optical centre of the left virtual camera (OC_L).

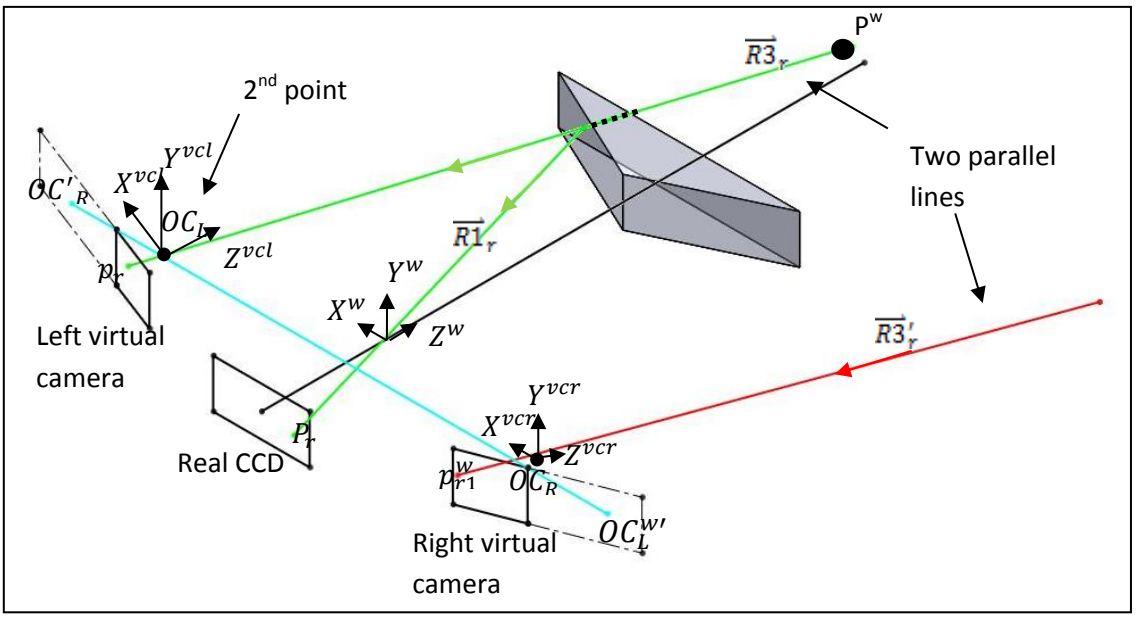


Figure 4.4 Construction of epipolar line on virtual camera

$$\text{The direction vector of } \overrightarrow{R1}_r = \begin{bmatrix} x_r \\ y_r \\ -f \end{bmatrix} \tag{4.1}$$

The unit vector of $\overrightarrow{R1}_r$,

$$\widehat{R1}_r = \frac{1}{||R1_r||} \begin{bmatrix} x_r \\ y_r \\ -f \end{bmatrix} \tag{4.2}$$

We will use the same coordinate system transformation method discussed in Section 3.2 to compute the rays $\overrightarrow{R2}_r$ and $\overrightarrow{R3}_r$. The general form of the direction vector of $\overrightarrow{R3}_r$ has been derived in Section 3.2 which is shown below:

$$\text{The direction vector of } \overrightarrow{R3}_r = F^B \begin{bmatrix} -(1-n^2K^2)^{0.5} \\ 0 \\ nK \end{bmatrix} \tag{4.3}$$

where $K = F_{i,3}^{B\,T}\left(\left(1-\frac{1}{n^2}\left(F_{i,3}^{A\,T}\widehat{R1}_r\right)^2\right)^{0.5} F_{i,1}^{A} + \frac{1}{n}F_{i,3}^{A\,T}\widehat{R1}_r F_{i,3}^{A}\right)$

Thus, under homogeneous coordinates,

$$\text{The direction vector of } \overrightarrow{R3}_r = F^B \begin{bmatrix} -(1-n^2K^2)^{0.5} \\ 0 \\ nK \\ 1 \end{bmatrix} \tag{4.4}$$

<u>a) Computation of the first candidate point for epipolar line construction:</u>

Projection of the point at infinity: Since parallel lines intersect at infinity in homogeneous coordinates, the correspondence ray $\overrightarrow{R3}'_r$ projected on the right virtual camera is parallel to the ray $\overrightarrow{R3}_r$ as shown in Figure 4.4.

$$\overrightarrow{R3}'_r = \begin{bmatrix} OC_{RX} \\ 0 \\ OC_{RZ} \\ 1 \end{bmatrix} + g\overrightarrow{R3}_r \tag{4.5}$$

where g is an arbitrary scalar of the vector $\overrightarrow{R3}'_r$ and $\begin{bmatrix} OC_{RX} \\ 0 \\ OC_{RZ} \\ 1 \end{bmatrix}$ is the coordinates of the right virtual camera's optical centre which have been recovered in Chapter 3.

The first candidate point (p_{r1}^{W}) for the epipolar line construction is the intersection point of the ray $\overrightarrow{R3}'_r$ and the RVCIP, where super script *w* denotes world coordinate frame, subscript *r1* denotes first candidate point for point p_r. Image plane of the right virtual camera expressed in the world coordinate frame:

$$x\sin\beta + z\cos\beta = OC_{Rx}\sin\beta + OC_{Rz}\cos\beta - f_{vc} \tag{4.6}$$

where *x* and *z* are the components of the points on the RVCIP, β is the rotation angle of the right virtual camera about the Y^w axis, OC_{Rx} and OC_{Rz} are the coordinates of its optical centre. All the coordinates (in world coordinate frame) of the image points on the RVCIP must satisfy equation (4.6).

By solving equations (4.5) and (4.6), we obtain

$$p_{r1}^{w} = \begin{bmatrix} -Tx \\ 0 \\ Tz \\ 1 \end{bmatrix} - f\overrightarrow{R3}_r' \frac{1}{cos\,\theta\,\overrightarrow{R3}_r'\,(3) + sin\,\theta\,\overrightarrow{R3}_r'\,(1)} \tag{4.7}$$

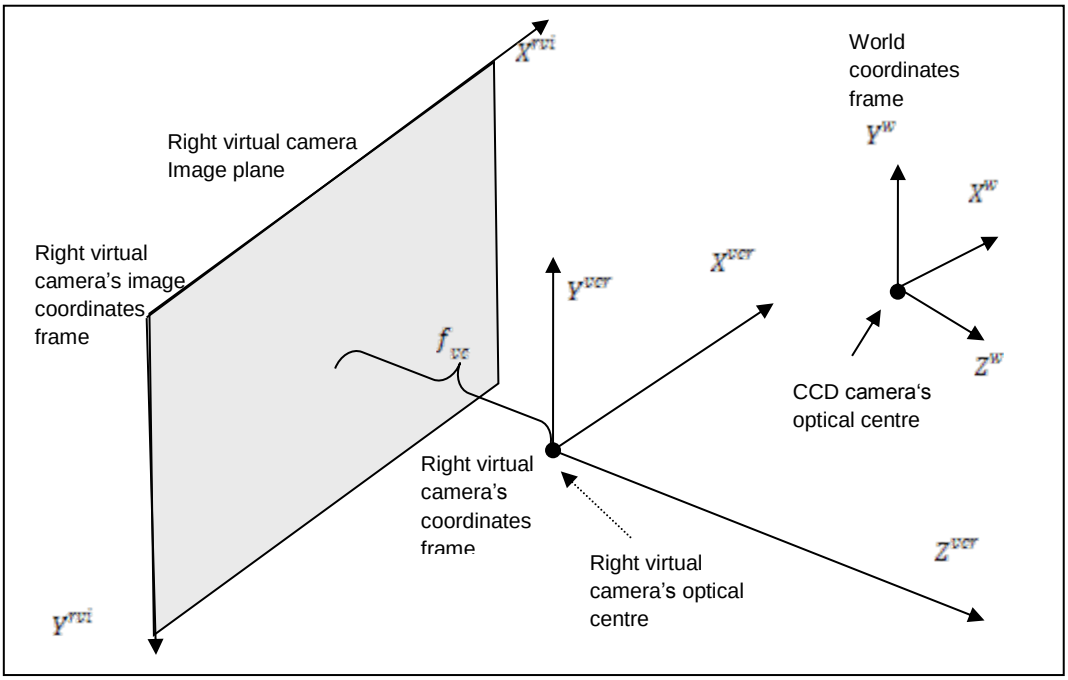


Figure 4.5 Illustration of the coordinate systems

p_{r1}^{w} is expressed in the world coordinate frame. It will be transformed to the virtual camera image coordinates. The illustration of the coordinate systems is shown in Figure 4.5. Transformation of the point, p_{r1}^{w} from the world coordinate frame to virtual camera coordinate frame can be written as:

$$p_{r1}^{vcr} = \begin{bmatrix} R_r & T_r \\ 0 & 1 \end{bmatrix} p_{r1}^{w} \tag{4.8}$$

where superscript *vcr* denotes the right virtual camera coordinate frame, R_r and T_r are the rotational and translational matrix of the right virtual camera with respect to the world coordinate frame.

By considering the intrinsic parameters of the right virtual camera, we transform the point from the virtual camera coordinates to virtual camera image coordinates.

$$p_{r1}^{rvi} = M_{int}\, p_{r1}^{vcr} \tag{4.9}$$

where superscript *rvi* denotes the right virtual camera image coordinates and M_{int} is the intrinsic parameter of the right virtual camera.

p_{r1}^{rvi} is the first candidate point for the virtual epipolar line construction.

b) Computation of the 2nd candidate point for epipolar line construction:

As shown in Figure 4.4, the second chosen point will always be the same which is the first point (OC_L, optical centre of the left virtual camera) on the ray $\overrightarrow{R3}_r$. The correspondence point of OC_L which is called $OC_L^{w\prime}$ on the right virtual camera is the intersection point of the line which connects OC_L and OC_R (optical centre of right virtual camera) with the RVCIP. Since the position and orientation of the virtual cameras have been recovered in the previous section, the $OC_L^{w\prime}$ can be computed easily. The vector of the line connecting the two virtual cameras' optical centre, $\vec{L}_{OC}$ is given by:

$$\vec{L}_{OC} = OC_{Rx}i + OC_{Rz}k + g((OC_{Rx} - OC_{Lx})i) \tag{4.10}$$

where g is an arbitrary scalar of the vector $\vec{L}_{OC}$. $OC_L^{w\prime}$, the intersection point between the RVCIP and the vector $\vec{L}_{OC}$ which can be obtained by solving equations (4.6) and (4.10). $OC_L^{w\prime}$ is also the epipole of the RVC and all the epipolar lines will pass through $OC_L^{w\prime}$. Thus, it can be taken as a common candidate point. The same procedures from (4.8)-(4.10) are repeated to convert $OC_L^{w\prime}$ from the world coordinates to the right virtual camera image coordinates, $OC_L^{rvi\prime}$.

Thus, the virtual epipolar line of p_r is the line connecting $OC_L^{rvi\,\prime}$ and p_{r1}^{rvi} (correspond to p_{r1}^{w} and $OC_L^{w\prime}$ under world coordinate system as shown in Figure 4.4). The correspondence search is reduced from whole image to a 2-dimentional search along this line.

This approach will be implemented on the system and the results will be shown in the next section.

4.3 Experimentation and discussion

The main objectives of the experiments are to verify the correctness of the proposed virtual epipolar line approach and study its accuracy and effectiveness as compared with the conventional methods. To achieve the former objective and test whether the experimental results agree with the theoretical derivations, several random points are selected and their respective epipolar lines are computed using several discrete points along the ray $\overrightarrow{R3}_r$ (see Figure 4.6). The first candidate point (projection of the point at infinity) which is used in our method is computed and plotted on the same image. If our proposed method is valid, the first candidate point should lie on the epipolar line.

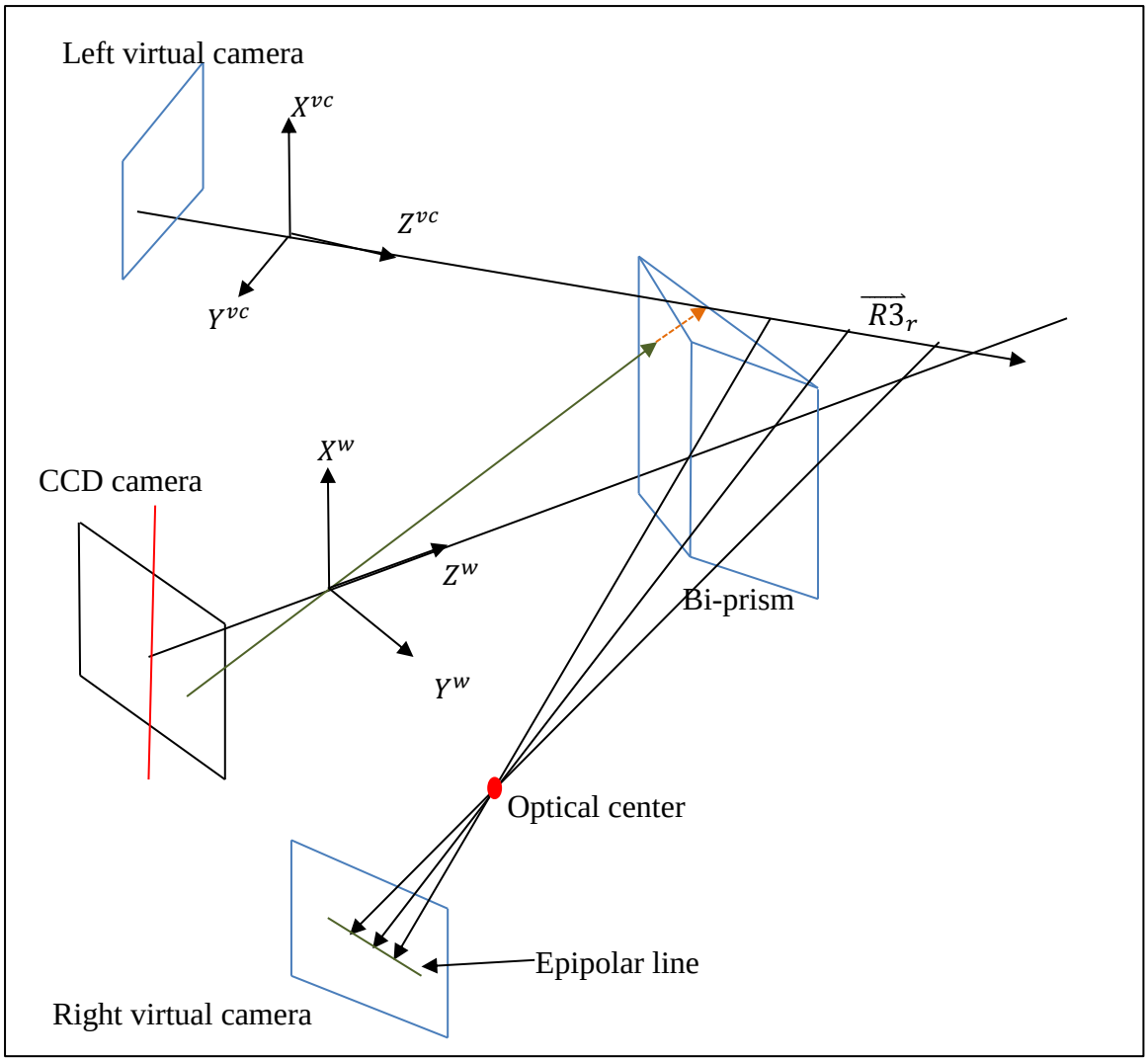


Figure 4.6 Construction of epipolar lines using several points on $\overrightarrow{R3}_r$

As shown in Figure 4.7, the first candidate point of our proposed method is computed and plotted on the image. The red crosses are the random points; blue lines are the computed epipolar lines and the white crosses are the first candidate points of the chosen random points respectively. Although the first candidate points are not within the image, the extension of the epipolar lines passes through their respective first candidate points as shown in Figure 4.8. In other words, it can be used as a candidate point to compute the epipolar line and our method is both mathematically and experimentally correct.

Figure 4.7 Epipolar lines and the first candidate points of several random points

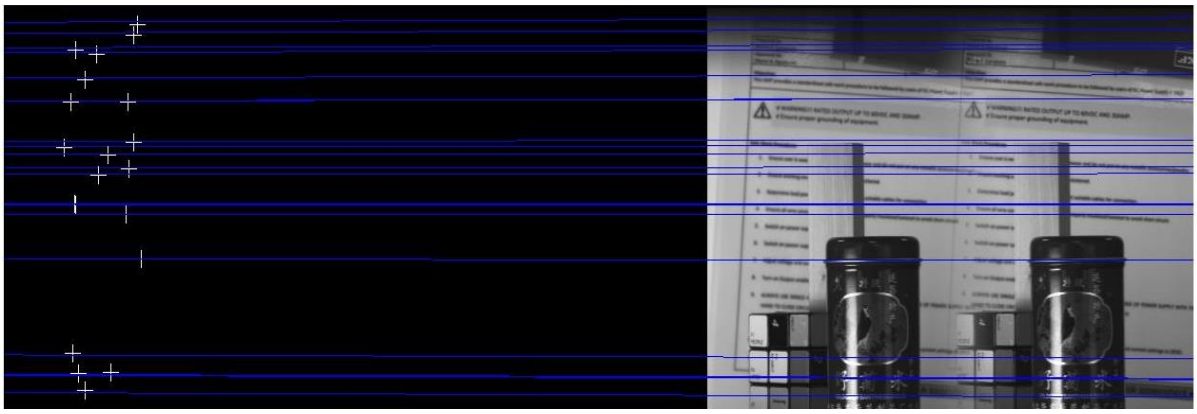


Figure 4.8 Epipolar lines pass through their respective first candidate point

To compare the results of our proposed virtual epipolar method with the conventional methods, two experiment setups are used and the parameters are shown below:

Table 4.1 Setup 1

α	n	f	T	T_o	C_x	C_y
6.4°	1.48	25 *mm*	5.6 *mm*	170 *mm*	512	384

Table 4.2 Setup 2

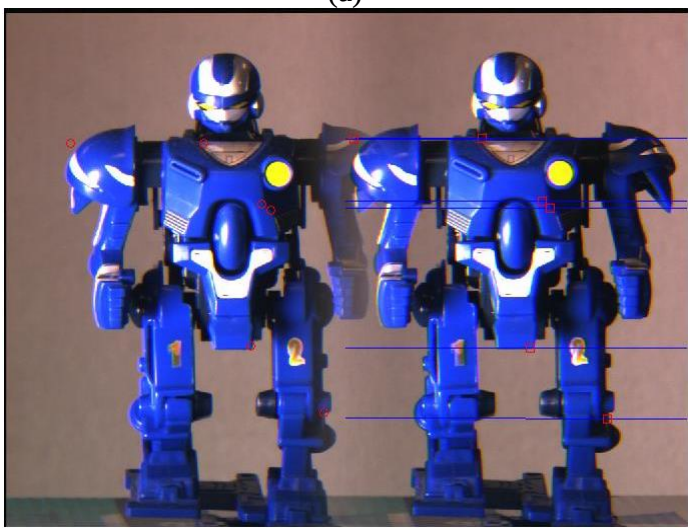
α	n	f	T	T_o	C_x	C_y
21.8°	1.48	8 *mm*	5.6 *mm*	170 *mm*	512	384

The calibration results of both the geometrical approach and Zhang's method [68] obtained in Chapter 3 will be used in the experiments. Several random points are selected and

their respective epipolar lines are constructed based on the proposed geometrical and conventional approach.

Figure 4.9(a) and (b) show the results of Setup 1, the square boxes are the correspondence points of the selected random pixels (circle dots). The correspondence points are searched along the constructed epipolar lines using the block-matching method [119]-[120] with window size of 35 pixels x 35 pixels.

(a)



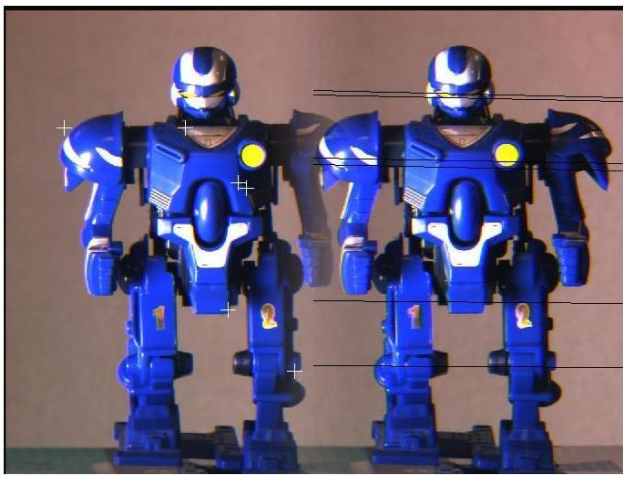
(b)

Figure 4.9 (a) and (b) Constructed epipolar lines based on the geometrical approach (Setup 1)

The epipolar lines and the correspondence points are also computed based on the Zhang's calibration approach [68]. The results using Setup 1 are shown in Figure 4.10(a) and

(b). As shown in the figures, Zhang's [68] method produces more wrong matches with less accurate epipolar lines as compared the proposed geometrical approach. More results and discussions will be shown later.

(a)

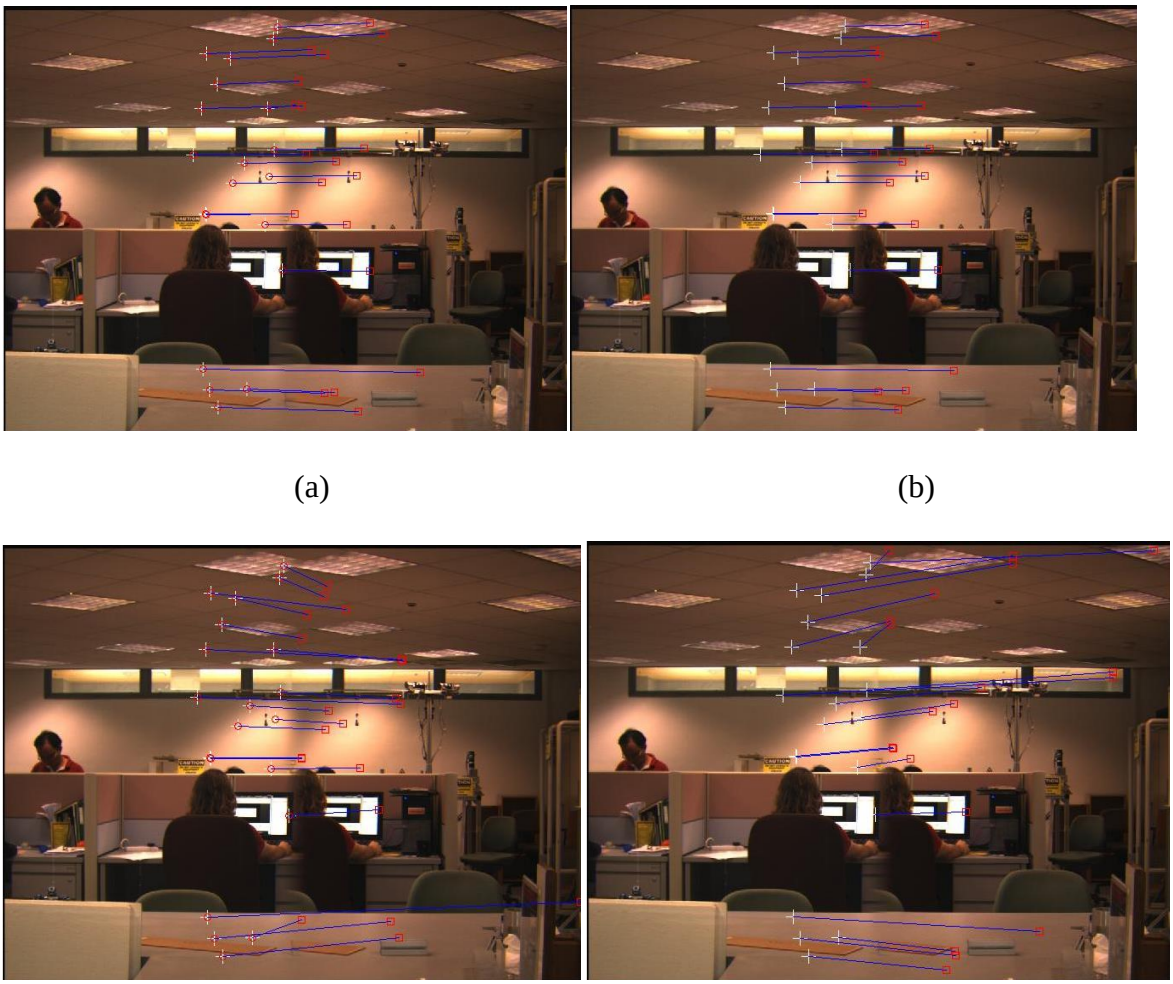


(b)

Figure 4.10 (a) and (b) Constructed epipolar line based on the conventional calibration approach (Setup 1)

Based on the results in Figures 4.9 and 4.10, we can deduce that the proposed geometrical approach outperforms the Zhang's calibration method [68]. The latter approach shows that its computed epipolar line and the correspondence points contains some errors. To show the efficiency of the proposed method, an image of our lab environment is captured

using this system. Twenty pairs of correspondence points are shown in Figure 4.11 for four different combinations of the approaches using Setup 1.



(a) (b)

(c) (d)

Figure 4.11 20 pairs of Correspondence points (connected by blue lines) using different approaches (Setup 1)

Illustration of Figure 4.11:

(a) Virtual camera calibration and epipolar line construction using geometrical approach.

(b) Virtual camera calibration using geometrical approach and epipolar line construction using fundamental matrix approach [35].

(c) Zhang's calibration and epipolar line construction using geometrical approach.

(d) Zhang's calibration and epipolar line construction using fundamental matrix approach.

The experimental results showed that the proposed method is able to solve the stereo correspondence problem effectively with acceptable accuracy. The proposed calibration approach showed better and significant results compared to the Zhang's method [68]. However, the proposed virtual epipolar line construction approach did not outperform the conventional fundamental matrix method [35]. The results for both the methods were very similar.

Furthermore, to compare the methods quantitatively, the correspondence results were used to compute the depths of the 3D points by using the mid-point theorem (Appendix C). The results obtained using the proposed geometrical approach were compared with Zhang's [68] and Tsai's method [55] as shown in Tables 4.3 and 4.4.

Each group in Tables 4.3 and 4.4 represented 50 object points with the same depth captured using the single-lens system with Setups 1 and 2 respectively. The actual depth was obtained using a précised laser sensor with accuracy up to 0.001mm. The trends of the errors are plotted in Figures 4.12 and 4.13 for both the setups.

Table 4.3 Comparison between the conventional calibration approaches and the proposed geometrical approach (Setup 1)

Group	Actual depth (mm)	Recovered depth (mm, Geometrical method)	Recovered depth (mm, Zhang's method)	Recovered depth (mm, Tsai's method)	Errors (%, geometrical method)	Errors (%, Zhang's method)	Errors (%, Tsai's method)
1	228	221	210	207.7	3.07%	7.89%	8.90%
2	312	309.2	292.3	287.1	0.90%	6.31%	7.98%
3	353	348.7	323.8	331.4	1.22%	8.27%	6.12%
4	418	409.6	383.6	386.5	2.01%	8.23%	7.54%
5	525	510.4	481.6	501.9	2.78%	8.27%	4.40%
6	625	616.9	568.8	601.9	1.30%	8.99%	3.70%
7	780	737.9	697.2	716.5	5.40%	10.62%	8.14%
8	936	894	825.2	877.6	4.49%	11.84%	6.24%
9	1192	1112.2	1038.4	1072.1	6.69%	12.89%	10.06%
10	1374	1254.9	1111.5	1075.5	8.67%	19.10%	21.72%
Avg					3.65%	10.24%	8.48%

Table 4.4 Results comparison between the conventional calibration approaches and the proposed geometrical approach (Setup 2)

Group	Actual depth (mm)	Recovered depth (mm, Geometrical method)	Recovered depth (mm, Zhang's method)	Recovered depth (mm, Tsai's method)	Errors (%, geometric al method)	Errors (%, Zhang's method)	Errors (%, Tsai's method)
1	228	224	217	216.3	1.75%	4.82%	5.13%
2	312	311.5	298.7	295.6	0.16%	4.26%	5.26%
3	353	351.3	333.6	330.9	0.48%	5.50%	6.26%
4	418	414	392.9	394.8	0.96%	6.00%	5.55%
5	525	515.2	491.6	501.5	1.87%	6.36%	4.48%
6	625	619.9	577.2	589.6	0.82%	7.65%	5.66%
7	780	757.3	718.9	719.5	2.91%	7.83%	7.76%
8	936	919.4	845.7	865.2	1.77%	9.65%	7.56%
9	1192	1152.7	1042.5	1072.6	3.30%	12.54%	10.02%
10	1374	1294.5	1123.1	1118.6	5.79%	18.26%	18.59%
Avg					1.98%	8.29%	7.63%

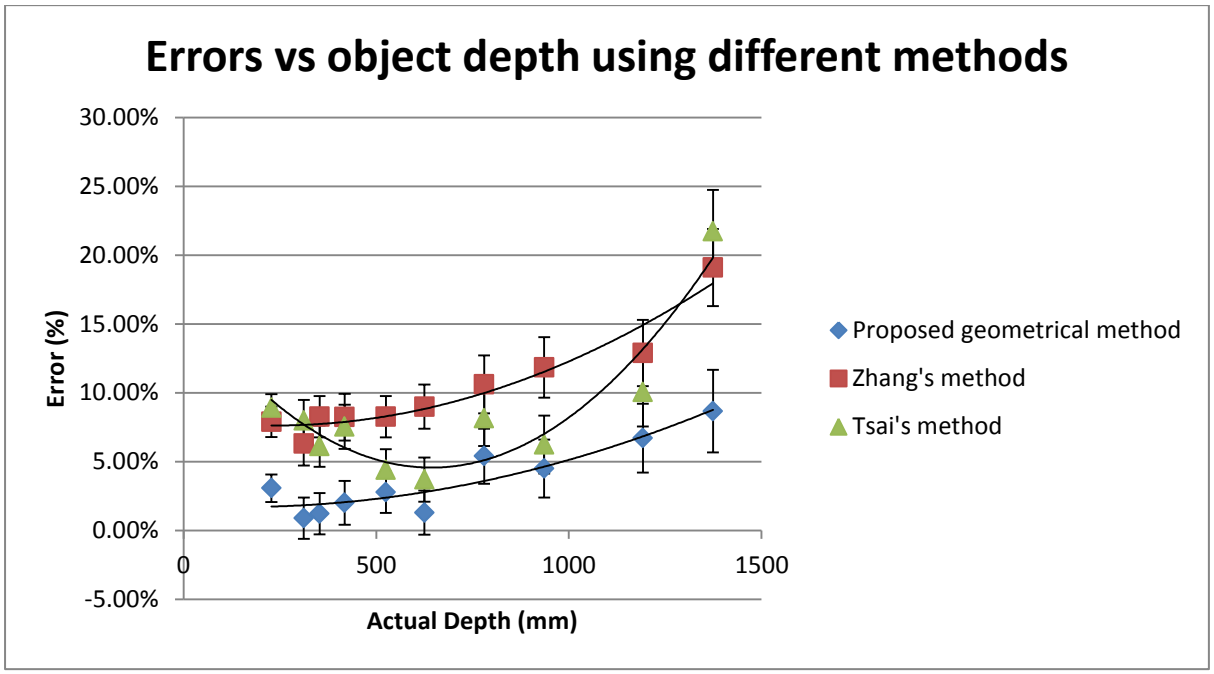


Figure 4.12 Depth recovery errors using different methods (Setup 1)

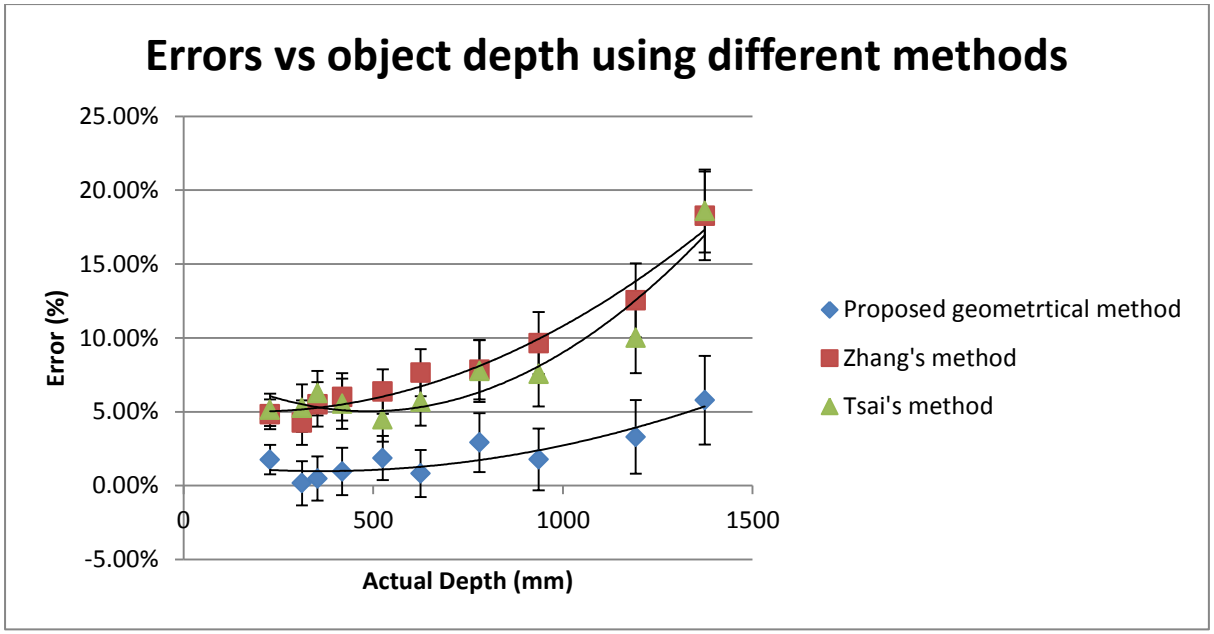


Figure 4.13 Depth recovery errors using different methods (Setup 2)

Comparing the results shown in Tables 4.3-4.4 and Figures 4.12-4.13, there are few observations and conclusions can be made:

1) The proposed geometrical method generally outperforms the conventional calibration algorithm in calibrating and solving the stereo correspondence problem for this particular system with better accuracy. Thus, the computed depths of the 3D points are more accurate by using the proposed method. In fact, the proposed approach only depends on the geometry of the system (position and orientation of the CCD camera and the bi-prism). All the extrinsic parameters and focal lengths of the virtual cameras can be recovered without capturing any images. On the other hand, the conventional approaches depend on the quality of the captured images (calibration boards). The accuracy also depends on the number of images and points used for calibration.

2) The conventional calibration approach developed by Zhang [38] and Tsai [55] has been proven to be robust to be applied in various stereovision systems. However, the proposed geometrical approach shows better performance in terms of accuracy

with shorter computation time and simpler geometrical derivation for this particular system.

3) Figure 4.12 shows an important observation in that the errors of all the three methods tend to be higher with the increasing object's depth. In other words, the single-lens system and the algorithms seem to perform better in close range as compared to long range. Since all the three algorithms exhibit the similar trend, this could be due to the hardware or sampling error of the system. We will study this area in more details in Chapter 6. For the case where the object is more than 1500mm away from the system, the depth recovery of the conventional methods is not reliable as the error is up to 20% while the error of the proposed geometrical method is around 8%.
4) The errors of the proposed method are mostly due to the assumptions made such as the perfect pin-hole camera model, intersection of parallel lines at vanishing point, zero distortions, perfect alignment of the bi-prism and the camera and perfect light refraction through the bi-prism. Thus, the accuracy of this method can be improved further by considering a more complete (and hence, more complicated) model which includes distortion on lens and bi-prism.

4.4 Summary

We have presented a virtual epipolar line method which is capable of solving the stereo correspondence and depth recovery problem for the single-lens bi-prism based stereovision system.

Given a right image point on the CCD camera, the expressions for both forward and refracted light rays in the three-dimensional space can be determined. Then, we can obtain the back projection ray of the second refracted ray to the left virtual camera. By picking two

unique points on this back projection ray, the virtual epipolar line can be constructed on the right virtual camera. This approach reduces the correspondence search from a whole image to 2-D slanted epipolar line. The experiment results prove the validity and favour our proposed method as compared to the conventional approaches. Moreover, by landing on the unique geometrical properties of the prism based imaging system, it can be easily generalized to multi-ocular stereovision systems.

Experimental results showed an important trend such that the errors of the depth recovery increase with the objects' depths. In other words, the further the object is placed from the system, the less accurate the system is. We will examine this observation in more detail in Chapter 6. In the next chapter, the effect of the most important parameter, α (the corner angle of the bi-prism) on the field of view of the single-lens system is examined.

Chapter 5 Effects of Angle and Position of Bi-Prism

As discussed in Chapter 2, many of the recent studies on single-lens bi-prism stereovision system are focused on correspondence searching using various techniques. In order to perform correspondence search, it is necessary for an object to fall within the field-of-view (FOV) of the system. It is also important that the correct system parameters are selected to produce a FOV type suitable for a particular application. While some applications may require a divergent FOV, others may be more feasible with a convergent FOV. The different types of FOV will be further discussed in Section 5.1. Developing new understanding in these areas of research would allow engineers to select the best parameters to produce the desired outcomes and optimize a particular system's performance in real life applications such as object detection, recognition and mobile robot navigation.

The FOV of the single-lens bi-prism stereovision system is affected by various parameters. These parameters include the corner angle of the bi-prism, the position and orientation of the bi-prism with respect to the camera and the material of the bi-prism.

In this chapter, the main objective is to establish a new understanding of how the FOV of the single-lens system is affected by the bi-prism angle, as well as the position of the bi-prism with respect to the camera. The former will be examined in detail, encompassing both divergent and convergent systems. This chapter is organized as follows: Section 5.1 provides a fundamental background on single-lens stereovision system using a bi-prism and different types of FOV. The methodologies to predict the FOV of the system and its geometry given a specific bi-prism angle is presented in Section 5.2. Section 5.3 shows the experiment results while Sections 5.4 and 5.5 examine the effect of the system's FOV under z- and x-axes translation of the bi-prism. Lastly, Section 5.6 recapitulates the chapter with a brief summary.

5.1 FOV of the Single-Lens Bi-Prism Stereovision System

This section provides a fundamental background of how single-lens bi-prism stereovision system works [16]-[17], including the required assumptions in modeling this set-up, the derivation of important geometrical parameters of this system and the convergence and divergence types of the system.

As Figure 5.1 shows, there is an overlapping area between the left and right virtual camera FOVs known as the common FOV. This region defines an area in which an object's image can be 'captured' by both left and right virtual cameras and thus be fully projected on the real camera image plane. It is also a region in which an object must be placed in order to achieve correspondence searching. This region is constrained by the left and right virtual optical axes making an angle of γ, known as the common FOV angle.

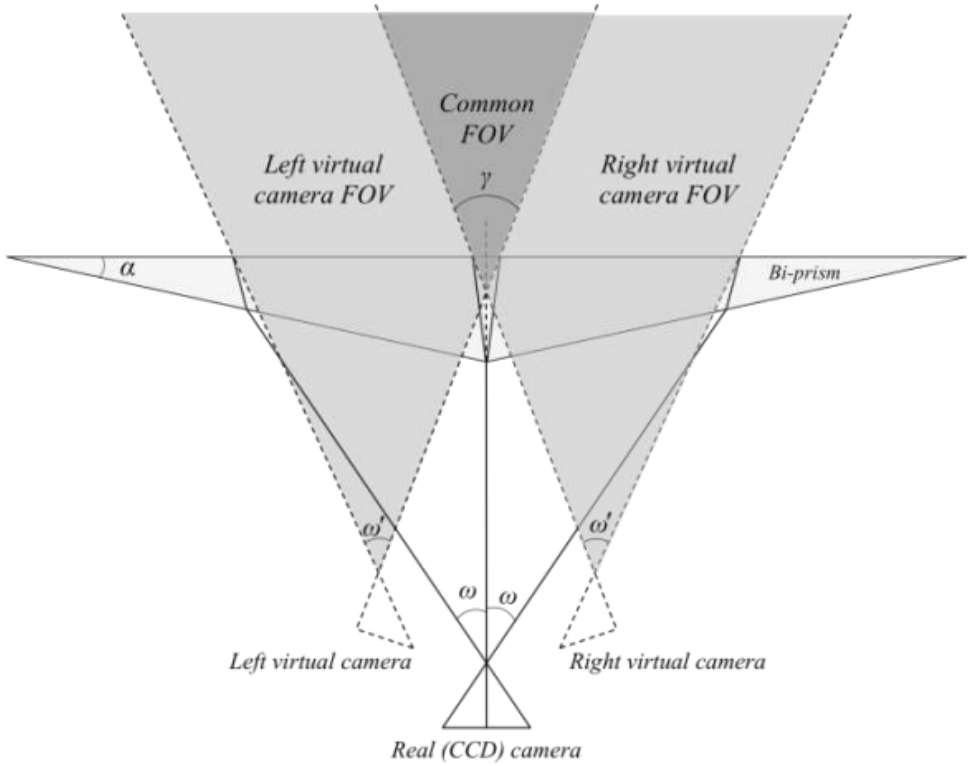


Figure 5.1: Single-lens bi-prism stereovision system showing the virtual cameras and their FOVs

Different application of single-lens bi-prism stereovision requires a different type of FOV. Knowing the type of FOV needed for a specific application will allow the engineer to design a set-up with system parameters that will produce that particular type of FOV.

Field-of-view (FOV) of the single-lens bi-prism stereovision system can be classified into two broad categories, convergent and divergent FOV, as shown in Figure 5.2.

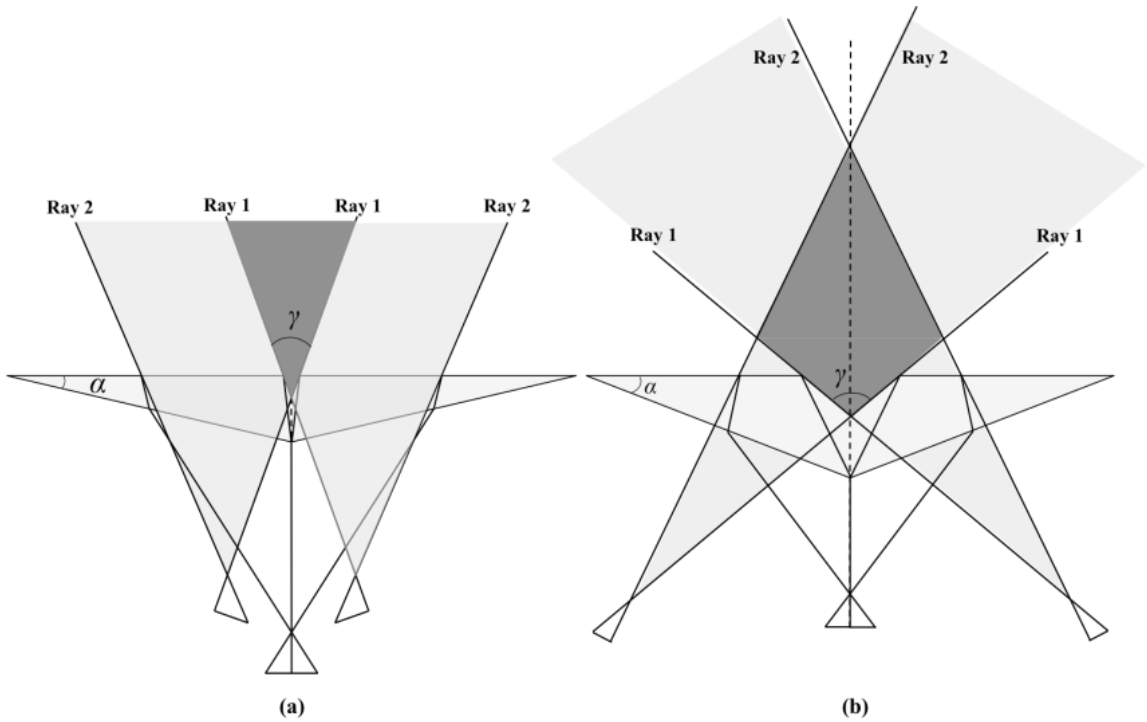


Figure 5.2: Two basic types of FOV: (a) divergent FOV, and (b) convergent FOV

In the context of single-lens bi-prism stereovision system, a divergent system is the one in which the FOV region is open and infinite, and bounded only by the left and right virtual optical axes, known as Ray 1 in Figure 5.2(a). In contrary, a convergent system is characterized by an enclosed and finite FOV region, and is fully bounded by Ray 1, as well as by left and right virtual camera FOV boundary lines, known as Ray 2. In our analyses, it is assumed that both Rays 1 and 2 are symmetrical about the real camera optical axis. These essential differences can be seen in Figure 5.2.

One of the factors to consider in designing a set-up for a specific application is whether the application requires an object to be placed at close-range or long-range with respect to the bi-prism. As convergent systems have a finite FOV, they would be more suitable for close-range applications such as indoor robot navigation and object detection. Conversely, divergent systems can be used for both close- and long-range applications, such as for obstacle detection in a dynamic scene. For a specific close-range application, a convergent system would be slightly more accurate in depth recovery than a divergent system due to a larger distance between the two virtual cameras known as baseline. This will be studied in detail in the next chapter. In modeling the single-lens system, the conditions specified in Section 3.2 are assumed to be true.

Under these assumptions, it can be said that the two virtual cameras have identical properties and are symmetrical about the real camera optical axis. It is also assumed that the following parameters are known, either by calibration from Chapter 3 or from technical data:

1. α, the corner angle of the bi-prism;
2. n, the refractive index of the bi-prism material;
3. T_0, the distance from the real camera optical center to the bi-prism apex;
4. T, the distance between the bi-prism apex and the back-plane;
5. f, the focal length of the real camera, and;
6. I, the real camera image width.

5.2 Predicting the Type of FOV based on the Bi-prism Angle

To predict the type of field-of-view (FOV) of a specific system, a simple methodology is presented which allows the geometry of each FOV to be defined using theoretical relationships.

5.2.1 Geometrical Analysis by Ray Tracing

The geometry of the single-lens bi-prism stereovision set-up by tracing the rays from the view zone to the image plane is shown in Figure 5.3. For simplicity, only the left virtual camera is shown. The same analysis applies to the right virtual camera as well.

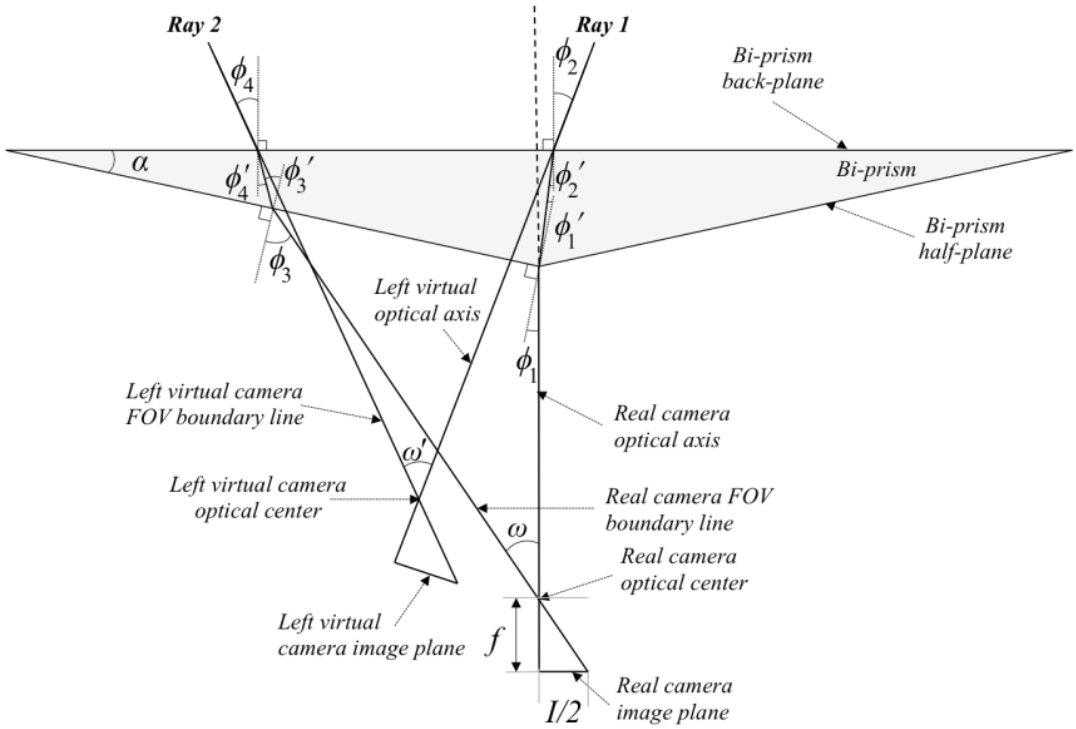


Figure 5.3: Ray tracing of virtual bi-prism stereovision (only left virtual camera is shown)

The left (and right) virtual optical axis is formed by back-extending the ray which makes an angle of ϕ_2 with the normal to the bi-prism back-plane. This ray, also known as Ray 1, is a result of refraction of the ray from the real camera optical axis at the bi-prism apex. Similarly, the left (and right) virtual camera FOV boundary line is formed by back-extending Ray 2, the ray that is at an angle of ϕ_4 to the normal of the bi-prism back-plane. Ray 2 is the refracted ray from the real camera FOV boundary line at the bi-prism half-plane surface. As shown in Figure 5.3, a series of angles ϕ_1, ϕ_2, ϕ_3, ϕ_4, ϕ_1', ϕ_2', ϕ_3', and ϕ_4' are formed at the points of refraction. According to the law of refraction,

$$n = \frac{\sin\phi_1}{\sin\phi_1'} = \frac{\sin\phi_2}{\sin\phi_2'} = \frac{\sin\phi_3}{\sin\phi_3'} = \frac{\sin\phi_4}{\sin\phi_4'} \tag{5.1}$$

The following relations can also be established from Figure 5.3 by geometrical analysis and can be expressed in terms of angle α, n, and ω :

$$\phi_1 = \alpha \tag{5.2}$$

$$\phi_1' = sin^{-1}(\frac{sin\,\alpha}{n}) \tag{5.3}$$

$$\phi_2' = \alpha - sin^{-1}(\frac{sin\,\alpha}{n}) \tag{5.4}$$

$$\phi_2 = sin^{-1}\left\{nsin\left[\alpha - sin^{-1}(\frac{sin\,\alpha}{n})\right]\right\} \tag{5.5}$$

$$\phi_3 = \omega + \alpha \tag{5.6}$$

$$\phi_3' = sin^{-1}\left[\frac{sin(\omega+\alpha)}{n}\right], \text{ where } \omega = tan^{-1}\frac{I}{2f} \tag{5.7}$$

$$\phi_4' = sin^{-1}\left[\frac{sin(\omega+\alpha)}{n}\right] - \alpha \tag{5.8}$$

$$\phi_4 = sin^{-1}\left\{nsin\left[sin^{-1}\left(\frac{sin(\omega+\alpha)}{n}\right) - \alpha\right]\right\} \tag{5.9}$$

Therefore, the common FOV angle γ (see Figure 5.2) which is twice of ϕ_2, is given by

$$\gamma = 2sin^{-1}\left\{nsin\left[\alpha - sin^{-1}(\frac{sin\,\alpha}{n})\right]\right\} \tag{5.10}$$

The type of FOV of this system can be predicted by simply comparing the values of ϕ_2 and ϕ_4, which are the angles that Rays 1 and 2 make with the normal of the bi-prism back-plane (see Figure 5.3). Rewriting the ϕ_4 expression,

$$\phi_4 = sin^{-1}\left\{n\,sin\left[sin^{-1}\left[\frac{sin\left(tan^{-1}\frac{I}{2f}+\alpha\right)}{n}\right] - \alpha\right]\right\} \tag{5.11}$$

Thus, given known values of I, f and n, different values of α will result in different combinations of ϕ_2 and ϕ_4, which then produce a specific type of FOV.

A dimensionless parameter, ε is proposed and defined as,

$$\varepsilon = \frac{\phi_4}{\phi_2} \tag{5.12}$$

This parameter is named as the *coefficient of field-of-view*. In subsequent discussions, this coefficient will be used to analyze various cases corresponding to different ranges of α. Figure 5.4 depicts a chart which compares the angles ϕ_2 and ϕ_4, for a fixed refractive index of 1.48 and CCD sensor width of 4.7616 mm.

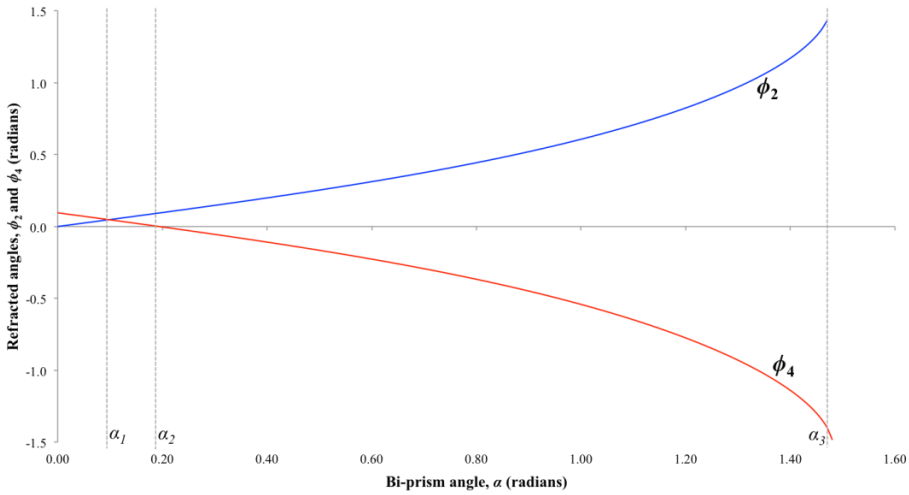


Figure 5.4: Comparison of ϕ_2 and ϕ_4 against α for a fixed CCD camera image width (I=4.7616mm)

From Figure 5.4, there are altogether three different cases corresponding to different ranges of α, and each of these results in a specific type of FOV.

<u>Case 1: $0 < \alpha \leq \alpha_1$</u>

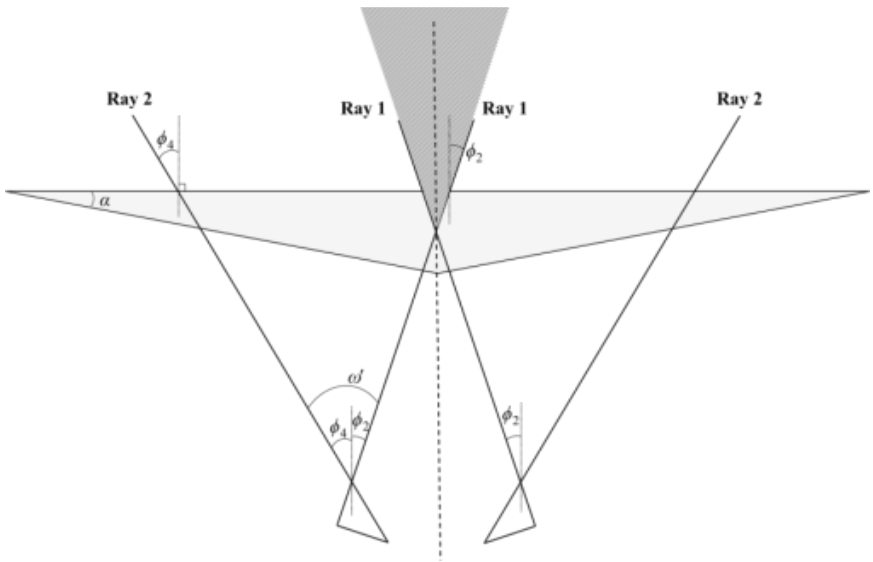
When α is less than α_1, it can be seen from Figure 5.4 that ϕ_4 is greater than ϕ_2 and both angles are positive. The coefficient of FOV, ε ranges from one to infinity, that is, $1 \leq \varepsilon < \infty$. The type of system produced is a divergent FOV, with Rays 1 and 2 diverging away from each other. This is depicted in Figure 5.5(a).

Case 2: $\alpha_1 < \alpha \leq \alpha_2$

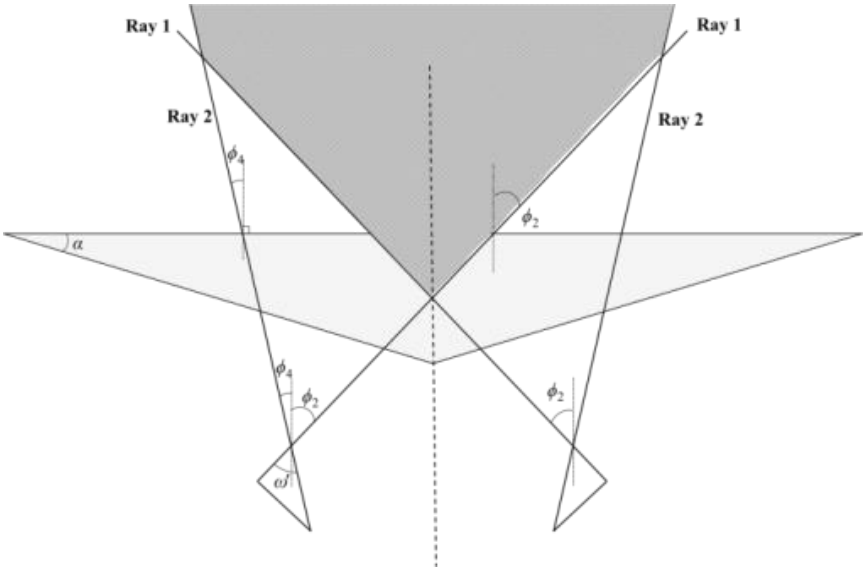
As α increases, ϕ_2 now becomes greater than ϕ_4. Ray 1 intersects Ray 2 to form a semi-divergent FOV, where the FOV remains open and infinite. The coefficient ε ranges between 0 and 1. This case is depicted in Figure 5.5(b).

Case 3: $\alpha_2 < \alpha < \alpha_3$

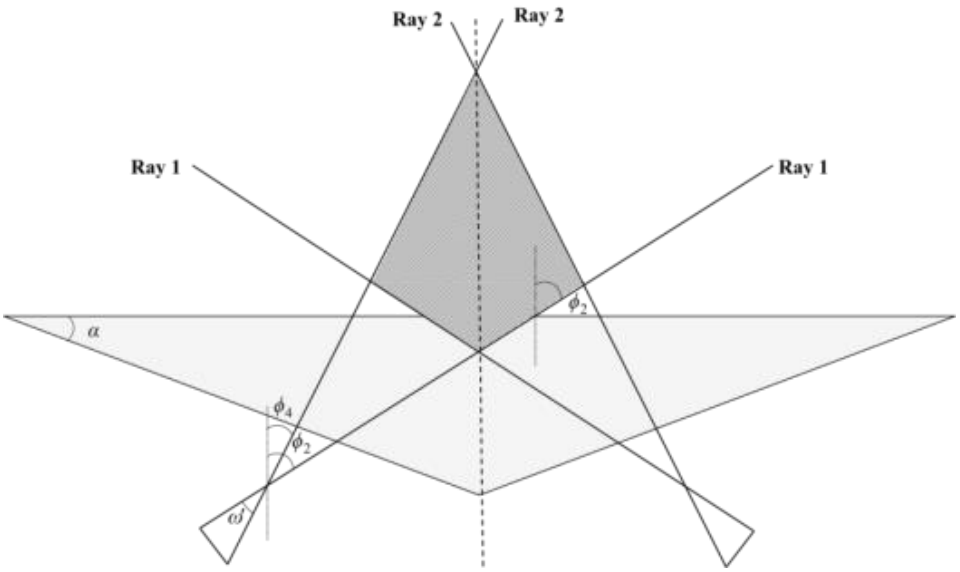
When α increases beyond α_2, both rays 1 and 2 intersect twice to enclose a finite FOV region. The coefficient ε becomes negative, that is, $\varepsilon < 0$. The resulting system is a convergent system, as shown in Figure 5.5(c). Table 5.1 summarizes all the cases.



(a)



(b)



(c)

Figure 5.5: (a) Case 1: divergent system (b) Case 2: semi-divergent system (c) Case 3: convergent system

Table 5.1: Summary of the different cases in predicting a specific type of FOV

Case 1	$0 < \alpha \leq \alpha_1$	$1 \leq \varepsilon < \infty$	Divergent system
Case 2	$\alpha_1 < \alpha \leq \alpha_2$	$0 \leq \varepsilon < 1$	Semi-divergent system
Case 3	$\alpha_2 < \alpha < \alpha_3$	$\varepsilon < 0$	Convergent system

Therefore, the type of FOV can be determined by simply finding the coefficient of FOV, ε (Equation 5.12). In the ensuing discussions, references may be made to include semi-divergent system under the general category of divergent system as well.

In practice, the boundary values of α_1 and α_2 are more relevant than α_3 as it may not be realistic to have a bi-prism with the corner angle approaching α_3. A bi-prism with α approaching α_3 produces a convergent system with a very small and finite FOV area. Application of such a system is limited.

Upon predicting the type of FOV given a specific bi-prism angle, the next important step is to derive parameters that define the FOV area. This will enable us to know the regions where the objects can be captured by the system. As shown earlier, a divergent FOV is mainly bounded by Ray 1 (see Figure 5.6) whereas a convergent FOV is bounded by both Rays 1 and 2 (see Figure 5.7). Therefore, it is possible to theoretically define the FOV by determining the linear equations of Rays 1 and 2.

5.2.2 Geometrical Analysis of Divergent System

Figure 5.6 shows a detailed geometry of a divergent system, with a *x*-*z* coordinate system affixed to the back-plane of the bi-prism to simplify the mathematical derivations.

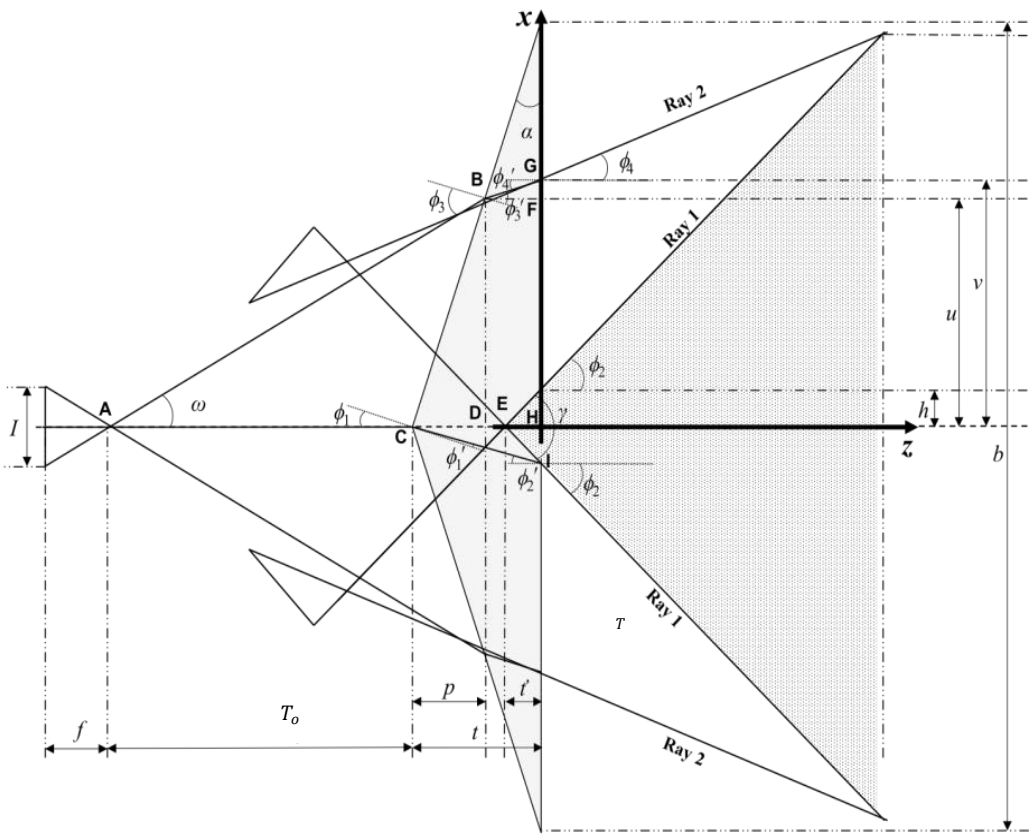


Figure 5.6: Detailed geometry of a divergent system

Referring to Figure 5.6, the linear equations of rays 1 and 2 can be found by determining these four parameters: h, ϕ_2, v and ϕ_4. h and v which are important in this study, also named as *coefficients of Ray 1* and *coefficient of Ray 2*, respectively, are the x-intercepts of Rays 1 and 2. ϕ_2 and ϕ_4 have been derived in the previous section (Equations 5.5 and 5.11).

Equation of Ray 1 is given by

$$x = z\,tan\,\phi_2 + h \tag{5.13}$$

Consider the right-angled triangle formed by the bi-prism half plane,

$$tan\,\alpha = \frac{T}{b/_2} \tag{5.14}$$

Next, consider Δ CHI.

$$tan\,\phi_2' = \frac{h}{T} \tag{5.15}$$

Substituting Equations 5.4 and 5.14 into Equation 5.15,

$$h = \frac{b}{2} tan\, \alpha\, tan \left[\alpha - \sin^{-1}\left(\frac{\sin \alpha}{n}\right)\right] \tag{5.16}$$

On the other hand, Ray 2 is defined by

$$x = z\, tan\, \phi_4 + v \tag{5.17}$$

Consider Δ ABD,

$$tan\, \omega = \frac{u}{T_o + p} \tag{5.18}$$

Consider ΔCBD,

$$tan\, \alpha = \frac{p}{u} \tag{5.19}$$

Solving Equations 5.18 and 5.19, we have

$$p = \frac{T_o\, tan\, \omega\, tan\, \alpha}{1 - tan\, \omega\, tan\, \alpha} \tag{5.20}$$

$$u = \frac{T_o\, tan\, \omega}{1 - tan\, \omega\, tan\, \alpha} \tag{5.21}$$

Now consider Δ BGF,

$$tan\, \phi_4' = \frac{v - u}{T - p} \tag{5.22}$$

Substituting Equation 5.20 and 5.21 into Equation 5.22,

$$v = \frac{T_o\, tan\, \omega}{1 - tan\, \omega\, tan\, \alpha} + \left(T - \frac{T_o\, tan\, \omega\, tan\, \alpha}{1 - tan\, \omega\, tan\, \alpha}\right) tan\, \phi_4' \tag{5.23}$$

Since $tan\, \omega = \frac{I}{2f}$, the above equation can be simplified to

$$v = \frac{T_o I}{2f - I\, tan\, \alpha} + \left(\frac{2fT - (T + T_o) I\, tan\, \alpha}{2f - I\, tan\, \alpha}\right) tan\, \phi_4' \tag{5.24}$$

Combining Equation 5.8 and 5.24, leads to

$$v = \frac{T_o I}{2f - I\, tan\, \alpha} + \left(\frac{2fT - (T + T_o) I\, tan\, \alpha}{2f - I\, tan\, \alpha}\right) tan \left[sin^{-1}\left[\frac{sin(\omega + \alpha)}{n}\right] - \alpha\right] \tag{5.25}$$

Therefore, given the known values of camera image width I, focal length f, refractive index n, back-plane length b, distance from the optical center to apex T_0 and bi-prism angle α, it is possible to determine the theoretical linear equations of Rays 1 and 2 for divergent system.

In addition, it can be deduced from the above equations that the camera image width I affects ϕ_4 and v but not ϕ_2 and h. In other words, the common FOV angle γ, which is twice of angle ϕ_2 would be independent of I. These theoretical predictions will be subsequently verified with experiments described in Section 5.3.

5.2.3 Geometrical Analysis of Convergent System

Similar geometrical analysis is done for a convergent system and the detailed geometry is shown in Figure 5.7. The equations of Rays 1 and 2 in the previous section for divergent system hold true for that of convergent system as well. The derivations of linear equations of rays 1 and 2 are omitted for the convergent case as they are similar as in the divergent case. The only difference between convergent and divergent FOV is that for the slope of Ray 2, the former case has a negative gradient (ϕ_4 is negative). By computing the Rays 1 and 2, the FOV of the system is fully recovered for both of the types. In the other words, objects which are placed inside the FOV can be captured and reconstructed by the single-lens system. The detailed derivations are shown in Appendix D.

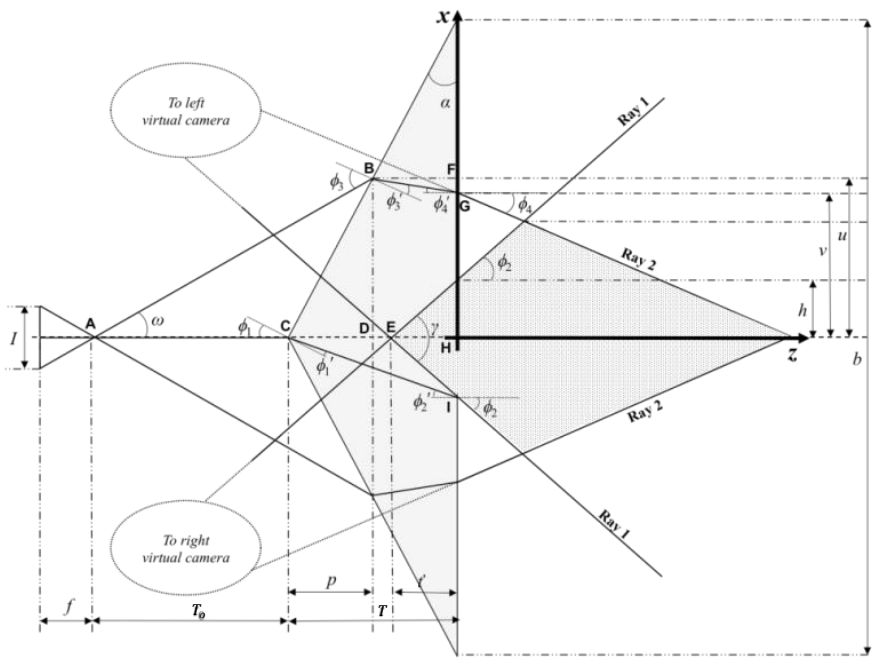


Figure 5.7: Detailed geometry of a convergent system

5.3 Experiment

The main objective of the experiment is to verify the derived theoretical predictions in the previous sections. Figure 5.8 shows the main experimental set-up comprising the bi-prism in a prism holder, the CCD camera, the laser sensor and the calibration board on their respective mechanical stands.

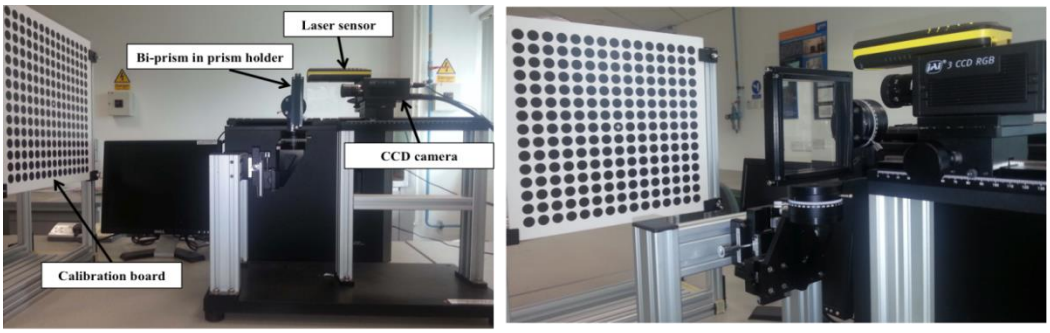


Figure 5.8: Experimental set-up

The specifications of the CCD cameras used in experiments are shown in Table 5.2. With these two CCD cameras and two bi-prisms with different corner angles, α (6.4° and 21.8°), there are altogether four different possible experimental set-ups: a) Set-up 1(α=6.4°,

I=4.7616mm) b) Set-up 2(α=21.8°, *I*=4.7616mm) c) Set-up 3(α=6.4°, *I*=6.3984mm) d) Set-up 4(α=21.8°, *I*=6.3984mm)

Table 5.2: CCD cameras' specifications

CCD cameras	Resolution (pixels)	Pixel Size (μm)	Sensor size (mm)
CV-M9CL	1024 x 768	4.65	4.7616
CB-140MCL	1376 x 1040	4.65	6.3984

The detailed experiment procedures are as follows:

1. The calibration board was placed at a distance along the *z*-axis (see Figure 5.9) and the distance was determined using the laser sensor.
2. The image was captured using the single-lens system.
3. The distance of the calibration board from the bi-prism was varied and steps 1 & 2 were repeated.
4. From the images, the points of intersection of Rays 1 and 2 with the calibration board were determined as A and B_l respectively (see Figure 5.9). The points were used to compute the Ray 1 and 2 parameters (h, ϕ_2, v and ϕ_4).

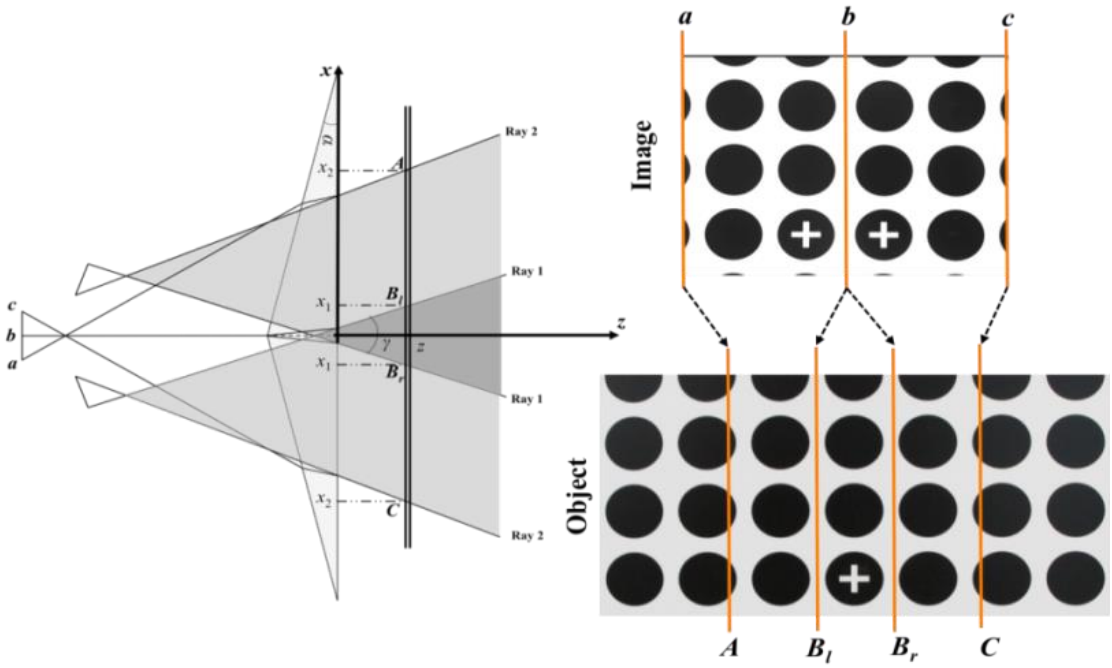


Figure 5.9 Interpretation of the captured image

5.3.1 Experimental Results

Figure 5.10 shows the images captured at different distances. Each circle of the calibration board has a diameter of 2cm and more circles covered means larger common FOV. Figure 5.11 shows a graphical representation of the scenes captured by the system. As the distances of the calibration board increase, the size of the common FOV increases.

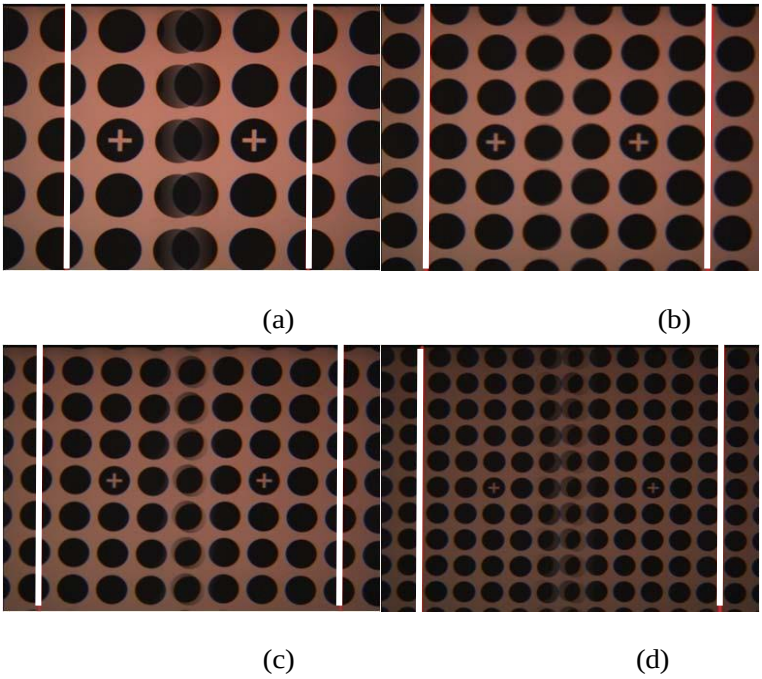


Figure 5.10: Real scene captured using Set-up 1 configuration, the common FOVs are highlighted by the two white lines, the images were captured at a distance of (a) $z_1 = 0.887m$ (b) $z_2 = 1.075m$ (c) $z_3 = 1.318m$ (d) $z_4 = 1.821m$

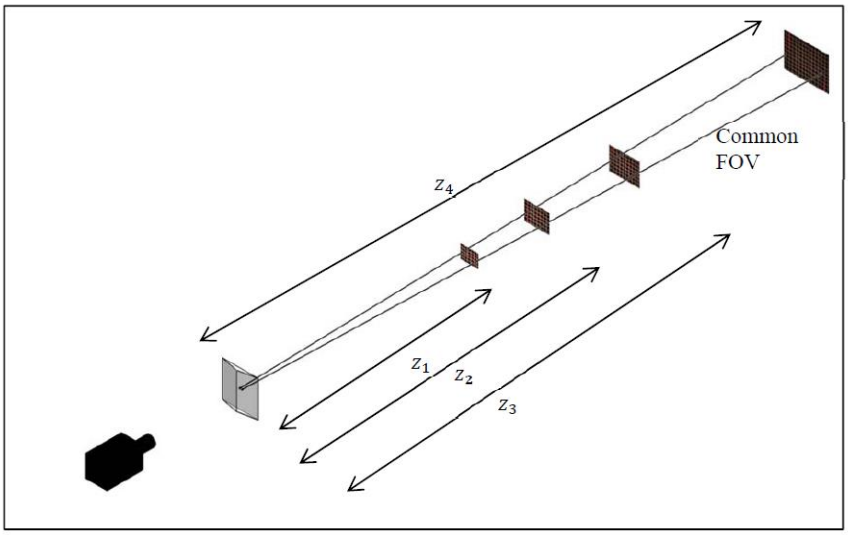


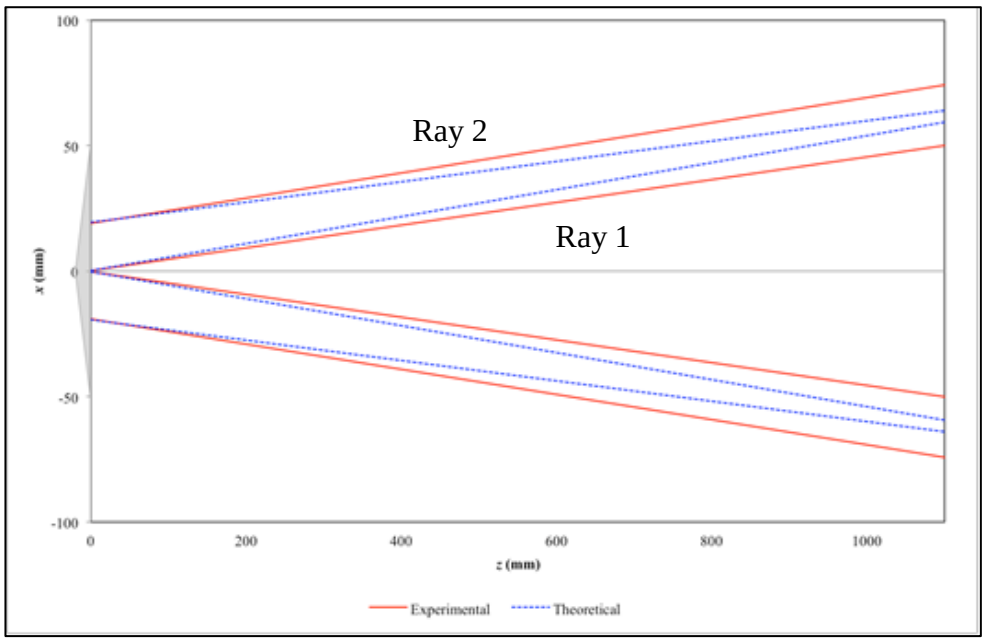
Figure 5.11: Graphical representation of the real scene captured by the system

For all the four set-ups, a quantitative comparison of theoretical and experimental values of Rays 1 and 2 parameters, which are, h, ϕ_2, v and ϕ_4, is shown in Table 5.3. The errors of Set-ups 2 and 4 (convergent systems) seem to be smaller as compared to the errors of Set-ups 1 and 3 (divergent systems). This suggests that the larger the angle, α of the convergent system, the better the accuracy is. More detailed discussion will be done in the next section.

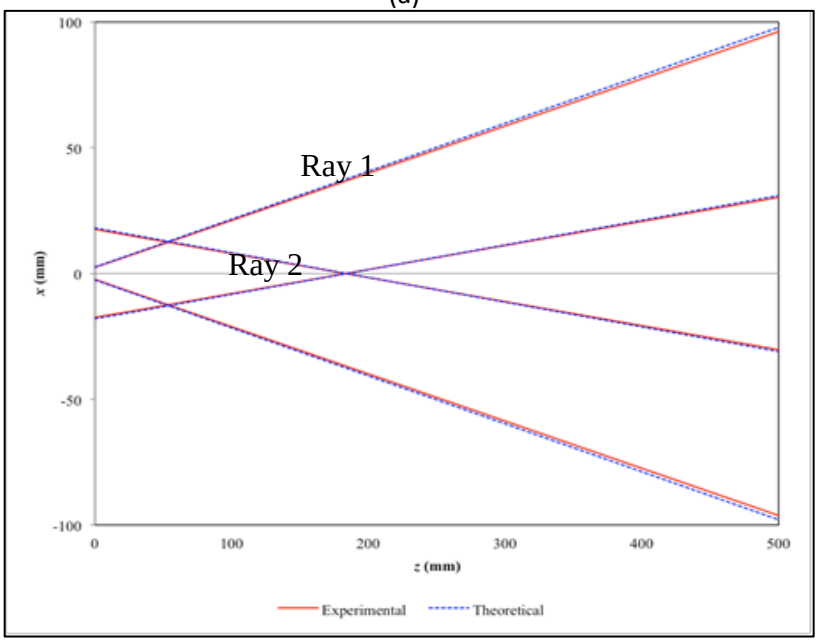
Table 5.3: Comparison of theoretical and experimental values of ray parameters

			Set-up 1	**Set-up 2**	**Set-up 3**	**Set-up 4**
Ray 1	**ϕ_2 (radians)**	**Theoretical**	0.05376	0.18834	0.05376	0.18834
		Experimental	0.04530	0.18554	0.04570	0.19613
		Error (%)	15.74	1.49	15.00	4.13
	h (mm)	**Theoretical**	0.20375	2.55045	0.20375	2.55045
		Experimental	0.21272	2.36224	0.20833	1.40703
		Error (%)	4.40	7.38	2.25	44.83
Ray 2	**ϕ_4 (radians)**	**Theoretical**	0.04053	-0.09785	0.07245	-0.06772
		Experimental	0.05008	-0.09537	0.07841	-0.07278
		Error (%)	23.57	3.64	8.23	7.47
	v (mm)	**Theoretical**	19.39945	18.07693	26.16954	25.30311
		Experimental	19.07949	17.53910	21.93493	20.54781
		Error (%)	1.65	2.98	16.18	8.79

Rays 1 and 2 (represent the border line of the FOV) are plotted for both the theoretical predictions and experimental results. Figure 5.12 shows a comparison between the experimental results and the theoretical predictions of Set-up 1 and 2. The theoretical FOVs presented in Figure 5.12 were obtained by using the theoretical equations derived in Section 5.2. Results of Set-up 3 and 4 are shown in Appendix E.



(a)



(b)

Figure 5.12: Comparison of experimental and theoretical FOV (a) Set-up 1 (b) Set-up 2

5.3.2 Discussions

As shown in Table 5.3 and Figure 5.12, experimental data deviates from the theoretical predictions to varying extent. Among the four parameters defining Rays 1 and 2, ϕ_2 and ϕ_4 are of greater importance. A small difference in the gradients of Rays 1 and 2 between theoretical and actual behaviors would multiply as z increases. This means the

greater the depth, the larger the error in predicting the width of the FOV. This would be undesirable if the system is to be used for long-range applications. Any error in angle ϕ_2 would be of most significance because it directly affects the common FOV angle γ. On the contrary, the *coefficients of Ray 1 and Ray 2* (*h* and *v*) have relatively lesser impact on the geometry of the FOV as the magnitudes of experimental and theoretical *h* and *v* differ by only a few millimeters.

Nevertheless, these deviations from theoretical values may be attributed to experimental errors. Interpretation of each image and determination of FOV region may not be perfectly accurate as it is entirely based on visual inspection. However, the experimental results exhibit the same trend as our theoretical predictions.

Lastly, it can be deduced from the experimental values in Table 5.3 that the common FOV angle γ is relatively unaffected by camera image width *I*. For example, using a bi-prism angle of 6.4°, the common FOV angle γ at I = 4.7616 mm is 5.191° while that at I = 6.3984 mm is 5.237°. The percentage difference of 0.89% between the two values of γ may be regarded as an insignificant difference. Therefore, the following conclusions can be made:

1. The common FOV of a particular system is not affected by the CCD image width *I*, as it has been theoretically postulated.
2. The results in this experiment affirm the coefficient, ε to predict the type of FOV given a specific bi-prism angle, as well as the use of theoretical linear equations of Rays 1 and 2 to estimate the geometry of the FOV. The accuracy of the FOV prediction can be improved by using a more accurate sensor in tracking the distances of the calibration and detecting the border lines of the common FOV.
3. The common FOV can be predicted mathematically.

5.4 Effect of Translation of Bi-Prism on System's Field-Of-View

In this section, we will extend our study on how the FOV of the single-lens bi-prism stereovision system is affected by the translation of the bi-prism with respect to the CCD camera in the z- and *x*-axes. The findings would be relevant in evaluating the feasibility of bi-prism translation in a specific set-up for a given industrial application.

5.4.1 Effect of Translation In z-Direction

Figure 5.13 shows the positions of the bi-prism before and after the translation of Δz, while keeping the positions of the camera and the calibration board fixed. By fixing a stationary *x*-*z* frame on the bi-prism, translating the bi-prism by Δz in the negative *z*-direction is equivalent to translating the camera in the positive *z*-direction by a distance of Δz. Since *f* is assumed to be constant and *T* is fixed for a specific bi-prism angle, translating the bi-prism in the *z*-direction is equivalent to varying the distance T_0.

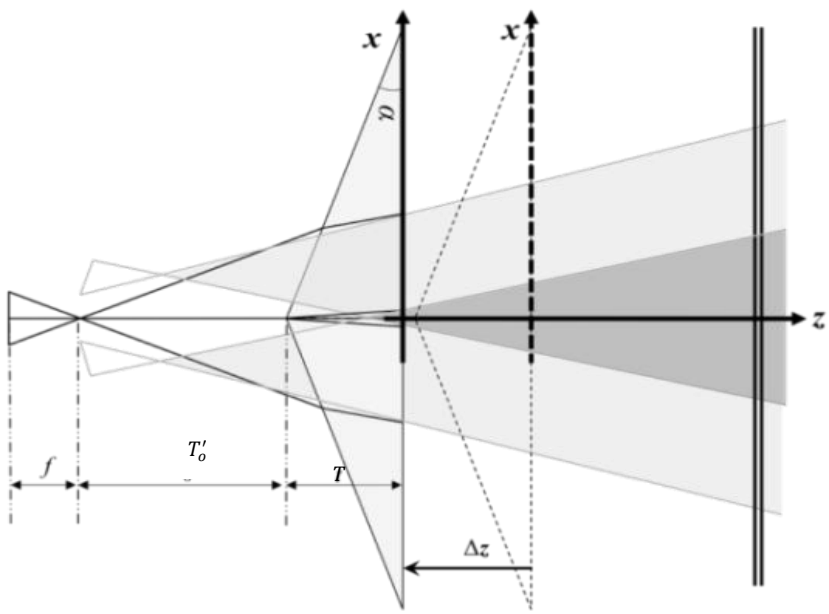


Figure 5.13: Translation of bi-prism in the *z*-direction for a divergent system

The ray parameters, ϕ_2, ϕ_4, *h* and *v*, have been derived in the previous section. Varying the parameter T_0 affects only *v*, which is the *x*-axis intercept of Ray 2. Figure 5.14 is

obtained using these theoretical linear equations for Rays 1 and 2 for different values of T_0. Therefore, it can be predicted that a translation of the bi-prism in the positive z-direction or an increase in distance T_0 has no effect on the Ray 1, but shifts the Ray 2 graph upwards. Common FOV angle γ is unaffected. These theories are validated with experiment results shown in Figure 5.15 using the similar experiment setup.

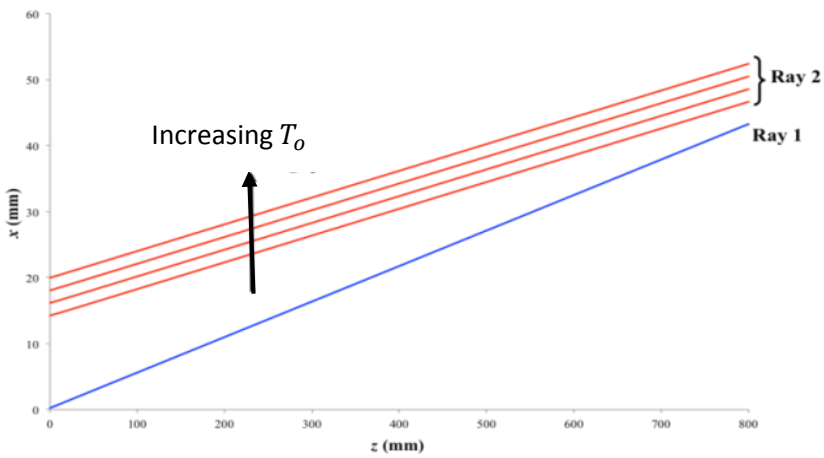


Figure 5.14: Graphs showing rays 1 and 2 at different values of t_0 for setup 1

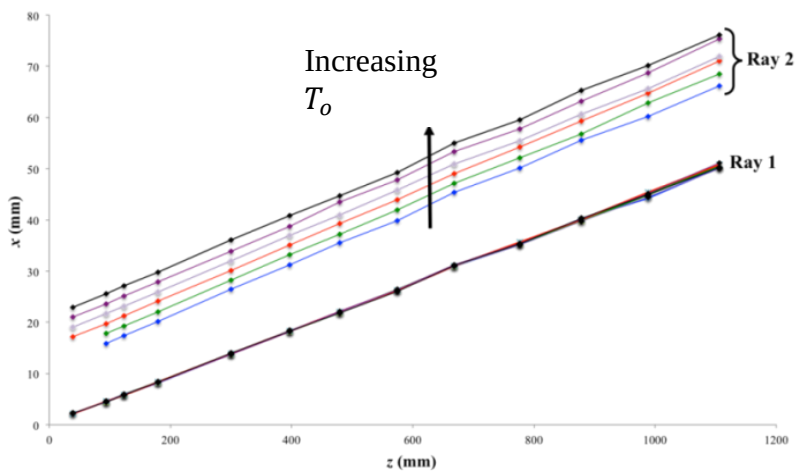


Figure 5.15: Graphs showing rays 1 and 2 derived from experimental results at different t_0 for setup 1

By comparing Figure 5.15 with the theoretical prediction in Figure 5.14, the actual behavior of Rays 1 and 2 in response to a z-axis translation of bi-prism conforms closely to

what was postulated. Both experiment and theory agree that the translation of bi-prism in the positive *z*-axis has no effect on ray 1, but it shifts ray 2 upwards.

5.4.2 Effect of Translation in *x*-Direction

In this section, we will explore the effect of *x*-axis translation of the bi-prism on the system's FOV by theoretical relationships and experiments.

Previously, under the assumption that the optical axis bisects the bi-prism exactly through the apex, refraction at the apex resulted in two symmetrical refracted rays inside the bi-prism due to the same incident angles upon the bi-prism half-planes at the apex. These rays inside the bi-prism refract again at the back-plane surface, producing another two symmetrical rays, which has been termed as Ray 1. Ray 1 makes an angle of ϕ_2 with the normal to the back-plane and is assumed to be symmetrical about the optical axis.

This observation no longer holds true when the bi-prism is translated in the *x*-direction. Due to this translation, the ray incident upon the two half-planes has two different incident angles. As a result, the two refracted ray from the apex are not symmetrical and are to be distinctly known as Ray 1_l and Ray 1_r as shown in Figure 5.16.

By geometrical inspection, it can also be observed that this translation has no effect on the angle between the real camera FOV boundary line and the normal to the half-plane, as well as the subsequent refracted angles. That is, the angles ϕ_3, ϕ_3', ϕ_4' and ϕ_4 are the same as the case without translation. Therefore, it will be assumed that the refracted ray from the bi-prism back-plane, known as Ray 2_l and Ray 2_r, may be approximated to Ray 2 as in the un-translated case. In this analysis, we introduce a series of new parameters to aid our understanding of refraction at the apex under *x*-axis translation of the bi-prism. These parameters are I_d, ω_0, μ_1, μ_1', μ_2, μ_2', μ_3, μ_3', μ_4, μ_4', u', v', h_l and h_r (see Figure 5.17-5.19)

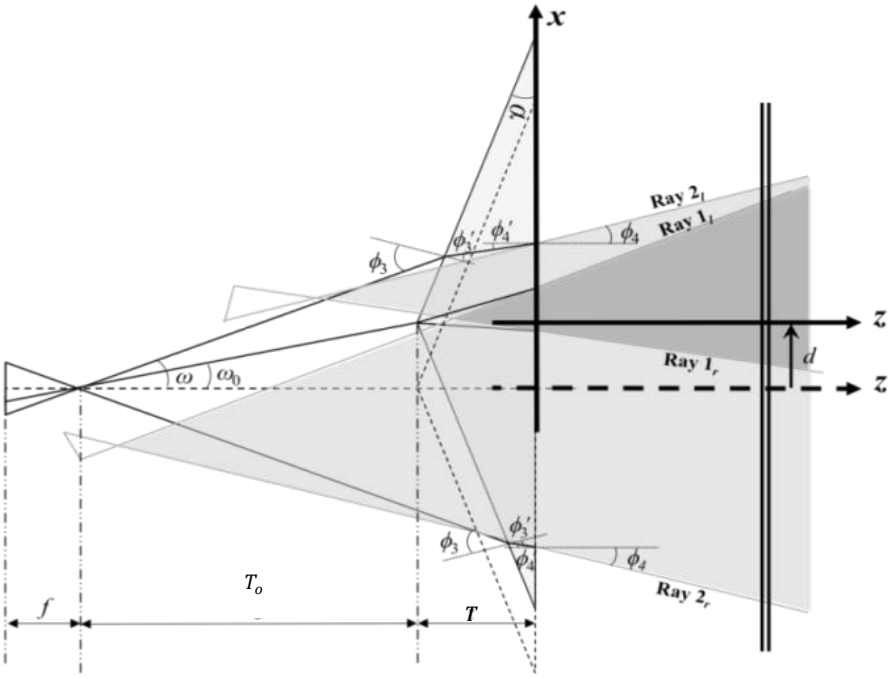


Figure 5.16: Effect of *x*-axis translation of bi-prism on the system

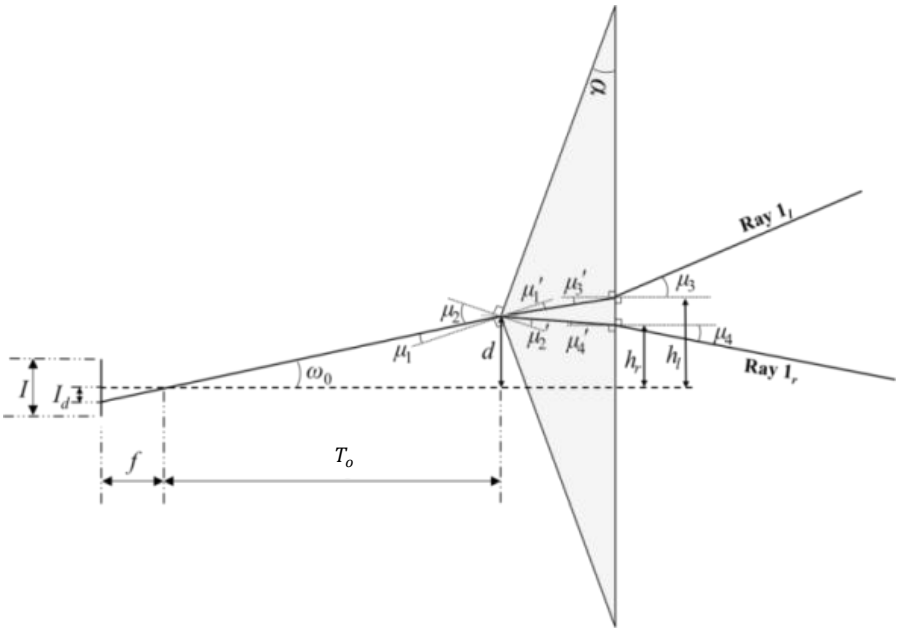


Figure 5.17: Ray tracing through the apex of the translated bi-prism

From Figure 5.17, by using law of refraction

$$n = \frac{\sin \mu_1}{\sin \mu_1'} = \frac{\sin \mu_2}{\sin \mu_2'} = \frac{\sin \mu_3}{\sin \mu_3'} = \frac{\sin \mu_4}{\sin \mu_4'} \tag{5.26}$$

The following relations can be established:

$$tan\,\omega_0 = \frac{d}{T_o} \tag{5.27}$$

$$\mu_1 = \alpha - \omega_0 \tag{5.28}$$

$$\mu_2 = \alpha + \omega_0 \tag{5.29}$$

$$\mu_3' = \alpha - \mu_1' \tag{5.30}$$

$$\mu_4' = \mu_2' - \alpha \tag{5.31}$$

$$h_l = d + T\,tan\,\mu_4' \tag{5.32}$$

$$h_r = d + T\,tan\,\mu_3' \tag{5.33}$$

The ray tracing through the translated bi-prism half-plane is shown in Figure 5.18.

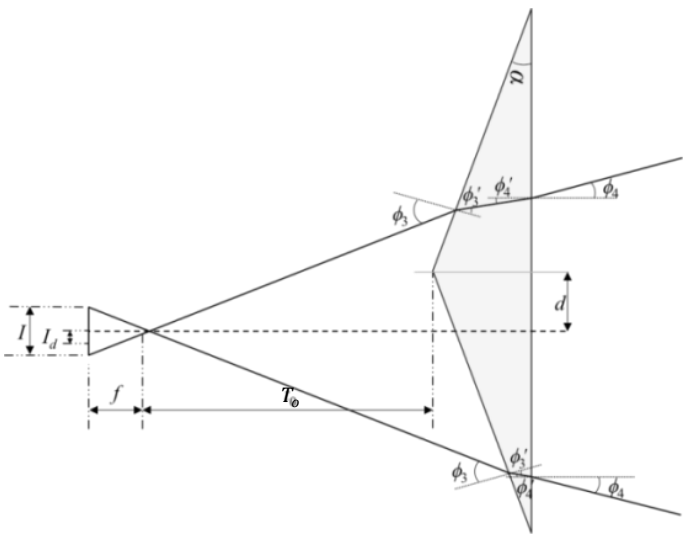


Figure 5.18: Ray tracing through the translated bi-prism half-planes

The angles ϕ_3, ϕ_3', ϕ_4 and ϕ_4' remain unaffected by this translation. Figure 5.19 shows a geometrical diagram of the rays through the bi-prism with and without *x*-axis translation (illustrated in dotted lines) in terms of parameters *u*, *v*, *u′* and *v′*. While *u* and *v* are parameters corresponding to a bi-prism without translation, *u′* and *v′* correspond to a translated bi-prism.

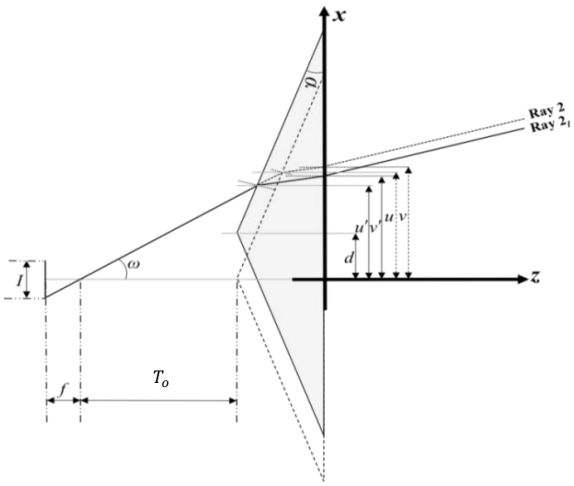


Figure 5.19: Geometrical analysis of *u* and *v* before and after *x*-axis translation

Let $u = u - u'$ and $\Delta v = v - v'$, where u, v, u' and v' are all with respect to the original *x*-*z* coordinate system as shown in Figure 5.19. By geometrical analysis, it can be found that

$$\Delta u = \frac{d \tan\alpha \tan\omega}{1 - \tan\alpha \tan\omega} \tag{5.34}$$

$$\Delta v = \frac{d \tan\alpha \, (\tan\omega - \tan\phi_4')}{1 - \tan\alpha \tan\omega} \tag{5.35}$$

With the above formulae for Δu and Δv, u' and v' can be determined. In essence, v' would be a more relevant parameter in comparing with experimental results as it is the *x*-intercept of Ray 2_l or 2_r of the translated bi-prism. Equations of Rays 1_l and 1_r are given by

$$x = z \tan\mu_3 + h_l \tag{5.36}$$

$$x = z \tan\mu_4 + h_r \tag{5.37}$$

where μ_3 and μ_4 are angles which Rays 1_l and 1_r make with the normal to the bi-prism back-plane and h_l and h_r are the respective *x*-intercepts. Thus, it is possible to theoretically predict how the common FOV is affected by the x-axis translation of the bi-prism.

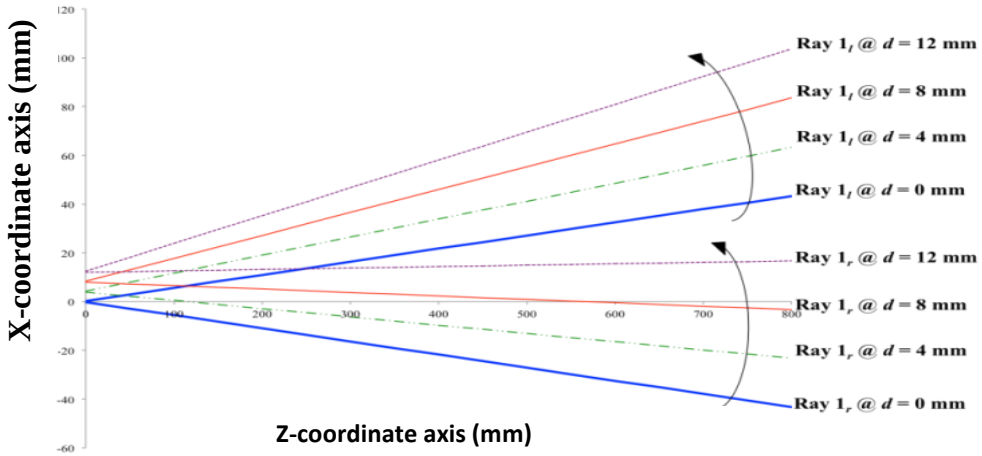


Figure 5.20: Effect of increasing *d* on rays 1_l and 1_r

Figure 5.20 is plotted using theoretical relationships with reference to the original *x*-*z* coordinate frame. Several predictions can be made:

1. As shown in Figure 5.20, the *x*-intercepts of h_l and h_r increase as *d* increases.
2. Translation of the bi-prism in the positive *x*-direction rotates both Rays 1_l and 1_r *clock-wise* about the positive *y*-axis (points out of the page). This also means that the divergent FOV rotates clock-wise about the positive *y*-axis.
3. The common FOV angle γ, which is the sum of μ_3 and μ_4, is constant. Therefore, an *x*-axis translation does not affect the common FOV angle. These predictions will be verified by experiments in the next section.

Conversely, translation of bi-prism in the negative *x*-axis has the effect of rotating the FOV *anti-clockwise* about the positive *y*-axis with common FOV angle remains constant, as well as decreasing the *x*-intercepts of Rays 1_l and 1_r.

5.5 Experimental Results

Comparing Figures 5.20 and 5.21, the actual behavior of Rays 1_l and 1_r with respect to varying d (x-axis translation of bi-prism) conforms closely to what was theoretically

predicted. Both experiment and theory are in agreement to the trend that a translation of bi-prism in the positive x-axis has the effect of rotating both Rays 1_l and 1_r clock-wise about the positive y-axis.

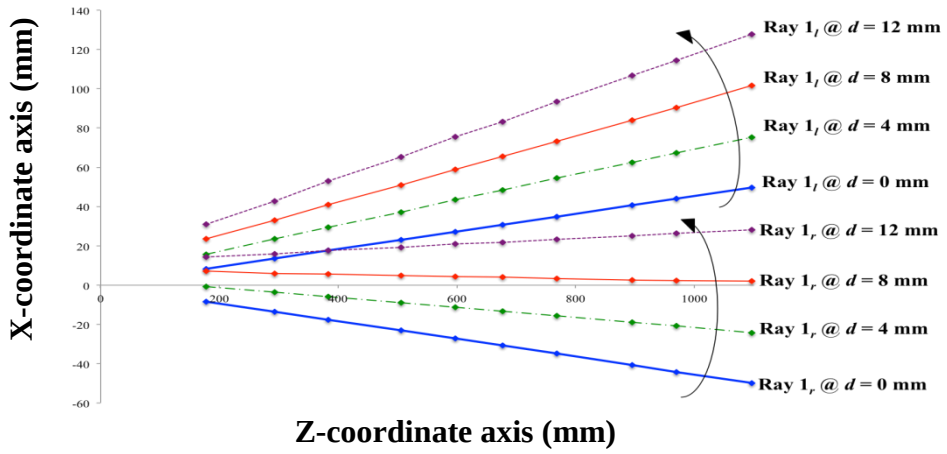


Figure 5.21: Graphs showing rays 1_l and 1_r derived from experimental results at different d for setup 1

Figure 5.22 shows the real scene captured at varying distance of d. The common FOV is shifted to the left or rotated about the y-axis when d increases. The FOV covered by the left sub-image becomes smaller which is due to the limitations of the sensor size (the red line which divides the two sub-images shifts to the left). This is consistent with the theoretical derivations and experimental results in Figures 5.20 and 5.21.

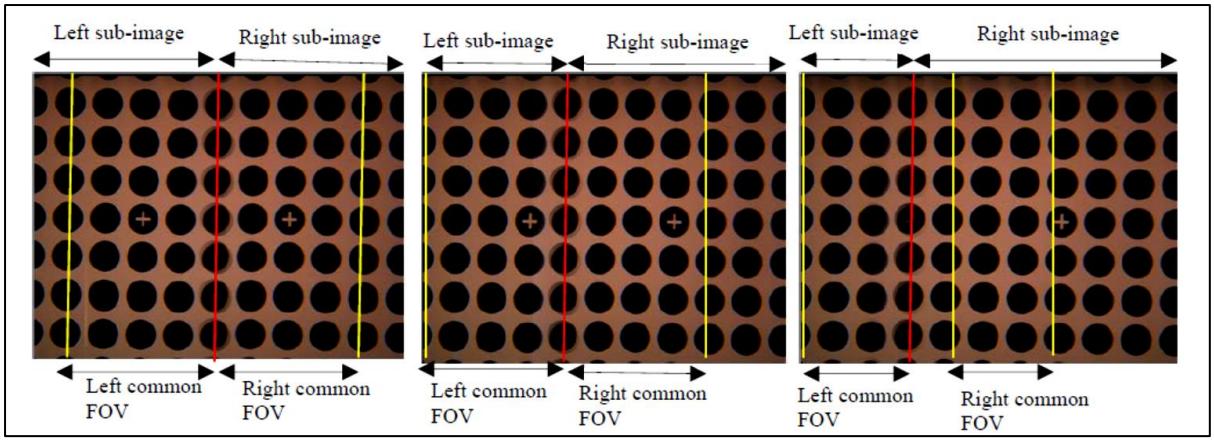


Figure 5.22: Real scene captured using Set-up 1 configuration, the common FOVs are highlighted by the two yellow lines and the two sub-images are divided by the red line. The images are captured at varying d (a) $d = 0mm$ (b) $d = 4mm$ (c) $d = 8mm$

Having examined the effects of bi-prism translation in *z*- and *x*-axes on the FOV of the system, Table 4.4 presents a summary of the important findings.

Table 5.4: A summary of the effect of translation in both *z*- and *x*-axes on the FOV of a system

Effect of Translation of Bi-prism	
***z*-axis translation**	***x*-axis translation**
FOV is bounded by ray 1, which is symmetrical about the optical axis.	FOV is bounded by rays 1_l and 1_r, which are not symmetrical about optical axis.
No effect on ray 1.	Rays 1_l and 1_r rotate clock-wise about positive *y*-axis under positive *x*-axis translation and vice versa.
Ray 2 shifts upwards under positive *z*-axis translation and vice-versa.	No effect on rays 2_l and 2_r.
No effect on FOV.	FOV rotates clock-wise about *y*-axis under positive *x*-axis translation and vice versa.
Common FOV angle is constant.	Common FOV angle is constant.

With the above findings, it is thus possible to theoretically predict how the common FOV and its angle change under *z*- and *x*-axes translation of the bi-prism.

5.6 Summary

In literature, there is thus far no study on single-lens bi-prism stereovision system to investigate how various system parameters affect its FOV. With the development of various theories and validation of these theories in experiments, it is conclusive to say that we have established new understanding specifically on the effect of bi-prism angle as well as the position of the bi-prism on its FOV. The FOV types can be predicted using the proposed coefficient, ε. Theoretical linear equations of Rays 1 and 2 can be utilized to estimate the geometry of a specific FOV type. Besides, this thesis also shows that the common FOV region is largely unaffected by translation of bi-prism in *z*-axis while translation in *x*-axis causes the region to rotate about the *y*-axis. The common FOV angle γ remains constant under bi-prism translation in both *z*- and *x*-axes.

This new understanding on single-lens bi-prism stereovision provides a fundamental background for future studies to build upon. These results can also add value to engineering

processes by allowing engineers to design and evaluate the feasibility of a specific set-up for a particular application, and thus select the best system parameters to produce the most effective outcomes. Our research group has been working with Northeastern University to work on a 3D endoscopy system using the single-lens prism based stereovision system. This project employed part of the findings of this work to obtain an appropriate field of view to capture 3D information of a tumor. Without the knowledge on the types of field of view generated by this system (using different angles of bi-prism), we would face difficulty in choosing the correct parameters to generate a correct field of view in the said medical instrument.

Chapter 6 Parameter Analysis

The design of the conventional stereovision system is dependent on the system parameters such as focal lengths of the viewing cameras, the baseline between the cameras, distance of the viewing objects from the cameras and the interval of image sampling. It has been shown that even with perfect correspondence, the depth error still increase quadratically with the objects' depth. In other words, the accuracy for long range or larger depth is usually low while the accuracy in near range or shorter depth is high. The sources of error such as focal length error, quantization error and camera rotation error have been studied in [109] and more details can be found in Chapter 2.

In this chapter, we will study the parameter analysis and estimate the error of the single-lens bi-prism stereovision system in terms of the system parameters. This unique system has been studied in the past decades [16]-[17] but none of them made any analysis on the quantization error and parameter analysis to estimate the error. Furthermore, our depth recovery results in Chapter 4 show that the error of the system increases when the object is placed further away from the system. In other words, the system is more accurate in short range as compared to long range. Hence, theoretical equations are derived to estimate the error and the trend of error are compared with the results shown in Chapter 4. The relative depth error which is essential to design the system appropriately for practical usage is then formulated. Based on the findings, the possibility of manipulating the system parameters, which is also named as variable parameter, is then studied in detail to reduce or maintain the error of the system for long range application.

This chapter is structured as follows: Section 6.1 will formulate the mathematical equation to estimate depth error of this system and experiment results will be shown in

section 6.2. In section 6.3, the idea of variable parameters is proposed and the concept is verified using experimental results described in section 6.4. Last but not least the conclusion is done in section 6.5.

6.1 Theoretical Analysis

The similar single-lens system shown in the previous chapters is used for both the theoretical and experimental analysis. Thus, similar assumptions are made (Section 3.2). Under these assumptions, it can be assumed that the two virtual cameras have identical properties and are symmetrical about the real camera optical axis. It is also assumed that the following parameters are known, either by measurement or from technical data:

1. α, the corner angle of the bi-prism;
2. *n*, the refractive index of the bi-prism material;
3. T_o, the distance from the real camera optical center to the bi-prism apex;
4. *T*, the distance between the bi-prism apex and the back-plane;
5. *f*, the focal length of the real camera, and;
6. *I*, the real camera image width.

6.1.1 Derivation of the Depth Equation

The geometry of the single-lens bi-prism stereovision with ray tracing from the view zone to the image plane is shown in Figure 6.1. We will not consider the formation of the virtual cameras in this analysis.

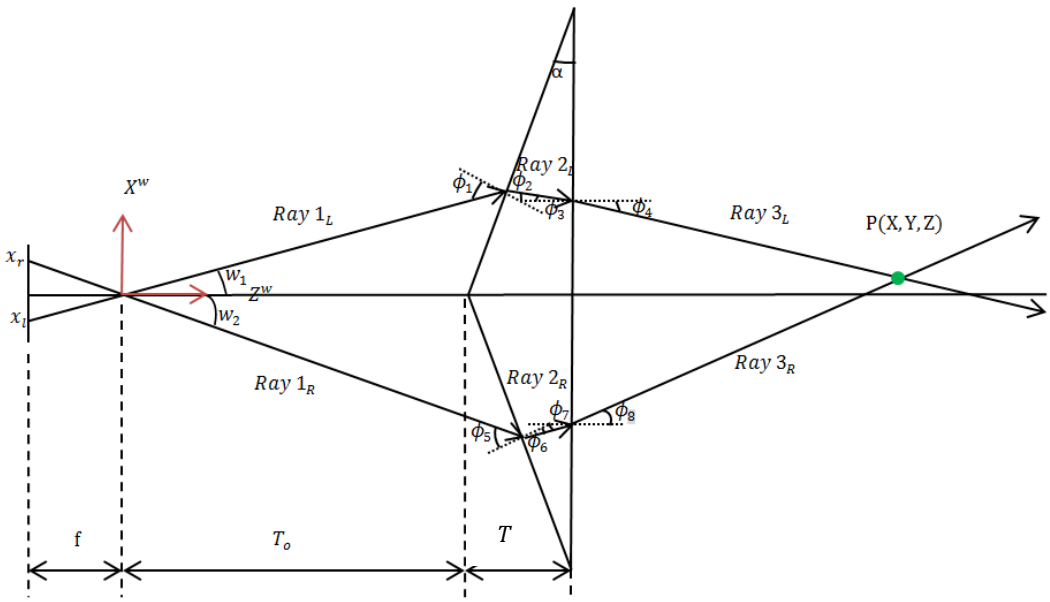


Figure 6.1: Geometrical rays of the system

Any points on the real camera image plane can be projected through the optical centre of the camera into the 3D space. As shown in Figure 6.1, $Ray\ 1_L$ will be refracted twice through the bi-prism forming $Ray\ 2_L$ and $Ray\ 3_L$. A series of angles ϕ_1 , $\phi_2, \phi_3, \phi_4, \phi_5, \phi_6, \phi_7$ and ϕ_8 are formed at the points of refraction on the bi-prism. According to the law of refraction,

$$n = \frac{\sin\phi_1}{\sin\phi_2} = \frac{\sin\phi_4}{\sin\phi_3} = \frac{\sin\phi_5}{\sin\phi_6} = \frac{\sin\phi_8}{\sin\phi_7} \tag{6.1}$$

The following relations can also be established from Figure 6.1 by geometrical analysis and can be expressed in terms of angle α, *n* and *ω* :

$$\phi_1 = \omega_1 + \alpha \tag{6.2}$$

$$\phi_2 = \sin^{-1}\left[\frac{\sin(\omega_1+\alpha)}{n}\right], \text{ where } \omega_1 = \tan^{-1}\frac{x_l}{f} \tag{6.3}$$

$$\phi_3 = \sin^{-1}\left[\frac{\sin(\omega_1+\alpha)}{n}\right] - \alpha \tag{6.4}$$

$$\phi_4 = \sin^{-1}\left\{n\sin\left[\sin^{-1}\left(\frac{\sin(\omega_1+\alpha)}{n}\right) - \alpha\right]\right\} \tag{6.5}$$

Point A, the intersection point of $Ray\ 2_L$ and the back plane of the biprism ($Z = T_o + T$) can be expressed as follows:

$$Point\ A = \begin{bmatrix} \frac{(T_o+T)\tan\omega_1}{1-\tan\omega_1\tan\alpha} - \tan\phi_3\left(d - \frac{(T_o+T)\tan\omega_1\tan\alpha}{1-\tan\omega_1\tan\alpha}\right) \\ 0 \\ T_o + T \end{bmatrix} \tag{6.6}$$

Thus,

$$Ray\ 3_L = \begin{bmatrix} \frac{T_o\tan\omega_1}{1-\tan\omega_1\tan\alpha} - \tan\phi_3\left(T - \frac{(T_o+T)\tan\omega_1\tan\alpha}{1-\tan\omega_1\tan\alpha}\right) \\ 0 \\ T_o + T \end{bmatrix} + q_l \begin{bmatrix} -\sin\phi_4 \\ 0 \\ \cos\phi_4 \end{bmatrix} \tag{6.7}$$

Where q_l is the arbitrary scalar of the direction vector of $Ray\ 3_L$. The same derivation (Equations 6.1-6.7) can be used to derive $Ray\ 3_R$ equation which is shown as bellows:

$$Ray\ 3_R = \begin{bmatrix} \frac{T_o\tan\omega_2}{1-\tan\omega_2\tan\alpha} - \tan\phi_7\left(T - \frac{(T_o+T)\tan\omega_2\tan\alpha}{1-\tan\omega_2\tan\alpha}\right) \\ 0 \\ T_o + T \end{bmatrix} + q_r \begin{bmatrix} -\sin\phi_8 \\ 0 \\ \cos\phi_8 \end{bmatrix} \tag{6.8}$$

Where q_r is the arbitrary scalar of the direction vector of $Ray\ 3_R$.

The intersection point of $Ray\ 3_L$and $Ray\ 3_R$ indicates the position of an object point in the 3D space. Thus, the general depth equation can be expressed as follows (next page):

$$Z = T_o + T + \left[\frac{-(T_o + T)\left(\tan\left(\alpha - \sin^{-1}\left(\frac{\sin\left(\frac{x_l}{f} + \alpha\right)}{n}\right)\right) + \tan\left(\alpha - \sin^{-1}\left(\frac{\sin\left(\frac{x_r}{f} + \alpha\right)}{n}\right)\right)\right) - T_o \frac{\left(\tan\left(\alpha - \sin^{-1}\left(\frac{\sin\left(\frac{x_l}{f} + \alpha\right)}{n}\right)\right) + \tan\frac{x_l}{f}\right)}{\tan\alpha \tan\frac{x_l}{f} - 1} - \frac{T_o\left(\tan\left(\alpha - \sin^{-1}\left(\frac{\sin\left(\frac{x_r}{f} + \alpha\right)}{n}\right)\right) + \tan\frac{x_r}{f}\right)}{\tan\alpha \tan\frac{x_r}{f} - 1}}{\tan\left(\sin^{-1} n \sin\left(\alpha - \sin^{-1}\left(\frac{\sin\left(\frac{x_l}{f} + \alpha\right)}{n}\right)\right)\right) + \tan\left(\sin^{-1} n \sin\left(\alpha - \sin^{-1}\left(\frac{\sin\left(\frac{x_r}{f} + \alpha\right)}{n}\right)\right)\right)} \right] \tag{6.9}$$

Since the angle of the bi-prism α has to be small in order to obtain a divergent system (details can be found in Chapter 5), a small angle approximation can be applied. Thus, the $\sin(\cdot)$ and $\tan(\cdot)$ terms in the depth equation (Equation 6.9) can be simplified as follows

$$Z = \frac{(T_o+T)(n-1)}{n} - \frac{T_o}{n}\left\{\frac{f\alpha(n-1)[\alpha(x_l+x_r)-2f]+(n-1)\boxed{[2\alpha x_l x_r - f(x_l+x_r)]}}{\left[2\alpha(n-1)-\frac{1}{f}(x_l+x_r)\right]\boxed{[\alpha^2 x_l x_r + f^2 - f(x_l+x_r)]}}\right\} \tag{6.10}$$

The two expressions in Equation 6.10 (highlighted in the square box) can be simplified as α is approximated to be a small angle.

$$2\alpha x_l x_r - f(x_l + x_r) \approx -f(x_l + x_r) \tag{6.11}$$

$$\alpha^2 x_l x_r + f^2 - f(x_l + x_r) \approx f^2 - f(x_l + x_r) \tag{6.12}$$

The depth equation becomes

$$Z = \frac{(T_o+T)(n-1)}{n} - \frac{T_o}{n}\left\{\frac{f\alpha(n-1)[\alpha(x_l+x_r)-2f]+(n-1)[-f(x_l+x_r)]}{\left[2\alpha(n-1)-\frac{1}{f}(x_l+x_r)\right][f^2-f(x_l+x_r)]}\right\} \tag{6.13}$$

Based on the convention of the coordinate system (x_l has a negative value), the term $(x_l + x_r)$ is the disparity of this single-lens system.

$$D = (x_l + x_r) \tag{6.14}$$

Thus, the final depth equation is

$$Z = \frac{(T_o+T)(n-1)}{n} - \frac{T_o(n-1)}{n}\left\{\frac{D+2f\alpha}{D+2f\alpha-2f\alpha n}\right\} \tag{6.15}$$

6.1.2 Relative Depth Error

In the previous section, we have derived a depth equation for the single-lens prism based stereovision system. Refer to Equation 6.15, the accuracy of the depth recovered using this system is constrained by the physical limitations of the system (T_o, T, n, f, and α) and

the performance of the correspondence algorithm. The disparity, D is an important parameter which affects the accuracy of the estimated depth. The error of D is due to both the false matching of the correspondence algorithm and the quantization/sampling error of the hardware.

From Equation 6.15, we know that the depth is a function of

$$Z = \boldsymbol{\Psi}(T_o, T, n, , f, \alpha \text{ and } D) \tag{6.16}$$

Assuming linearity, the absolute depth error:

$$dZ = \frac{\partial \boldsymbol{Z}}{\partial D} dD + \frac{\partial \boldsymbol{Z}}{\partial T} dT + \frac{\partial \boldsymbol{Z}}{\partial f} df + \frac{\partial \boldsymbol{Z}}{\partial T_o} dT_o + \frac{\partial \boldsymbol{Z}}{\partial \alpha} d\alpha + \frac{\partial \boldsymbol{Z}}{\partial n} dn \tag{6.17}$$

Thus, the relative depth error:

$$\frac{dZ}{Z} = \frac{\partial Z}{\partial D}\frac{dD}{Z} + \frac{\partial Z}{\partial T}\frac{dT}{Z} + \frac{\partial Z}{\partial f}\frac{df}{Z} + \frac{\partial Z}{\partial T_o}\frac{dT_o}{Z} + \frac{\partial Z}{\partial \alpha}\frac{d\alpha}{Z} + \frac{\partial Z}{\partial n}\frac{dn}{Z} \tag{6.18}$$

Based on the above equation, the depth errors are mainly due to five components: disparity error (dD), geometry of the prism $(dT$ and $d\alpha)$, refractive index of the prism (dn), focal length error (df) and inaccuracy of the distance between prism and the optical centre of the camera (T_o). In order to simplify the error analysis of this single-lens system, the errors caused by each parameter are examined independently. By assuming linearity and using the concept of partial derivative, the disparity, D will be studied first since it is the most important parameter. Thus, Equation 6.17 can be simplified as,

$$dZ \approx \frac{\partial Z}{\partial D} dD \tag{6.19}$$

$$\frac{\partial Z}{\partial D} = \frac{\left(Z - \frac{(T_o+T)(n-1)}{n}\right)^2}{\frac{T_o(n-1)}{n}} \frac{2\alpha f n}{(D+2\alpha f)^2} \tag{6.20}$$

$$\frac{\partial Z}{\partial D} = \frac{(Z-A)^2}{T_o} \frac{n^2}{n-1} \frac{2\alpha f}{(D+2\alpha f)^2} \tag{6.21}$$

$$\frac{\partial Z}{\partial D} = \frac{(Z-A+B)^2}{2} \frac{1}{\alpha f n B} \tag{6.22}$$

where $A = \frac{(T_o+T)(n-1)}{n}$ and $B = \frac{T_o(n-1)}{n}$. It can be concluded that the absolute depth error is quadratically related to Z and proportional to ∂D.

The relative depth error,

$$\frac{\partial Z}{Z} \approx \frac{(Z-A+B)^2}{Z}\frac{1}{2\alpha f n B}\partial D \tag{6.23}$$

From Equation 6.23, the relative depth error increases linearly with the range of the captured object and the disparity error. It becomes apparent that the range estimation of this system is more accurate for close range as compared to far range. The absolute depth error and relative depth error are also formulated in terms of other parameters as follows:

<u>Distance between the prism to the camera, T_o</u>

$$\frac{\partial Z}{\partial T_o} = \frac{n-1}{n}\left[\left(\frac{2\alpha f n}{D+2\alpha f} - 1\right)^{-1} + 1\right] \tag{6.24}$$

Or

$$\frac{\partial Z}{\partial T_o} = \frac{n-1}{n}\left[\frac{Z-A}{B} + 1\right] \tag{6.25}$$

The relative depth error,

$$\frac{\partial Z}{\partial T_o}\frac{1}{Z} = \frac{n-1}{n}\left[\frac{Z-A}{B} + 1\right]\frac{1}{Z} \tag{6.26}$$

<u>Angle of the bi-prism, α</u>

$$\frac{\partial Z}{\partial \alpha} = -\frac{T_o(n-1)}{n}\left(\frac{2\alpha f n}{D+2\alpha f} - 1\right)^{-2}\left[\frac{2fn}{D+2\alpha f} - \frac{2\alpha f n}{(D+2\alpha f^2)}2f\right] \tag{6.27}$$

Or

$$\frac{\partial Z}{\partial \alpha} = -B\left(\frac{Z-A}{B}\right)^2\left[\frac{1}{\alpha}\left(\frac{Z-A+B}{Z-A}\right) - \left(\frac{Z-A+B}{Z-A}\right)^2\frac{1}{n\alpha}\right] \tag{6.28}$$

The relative depth error,

$$\frac{\partial Z}{\partial \alpha}\frac{1}{Z} = -\frac{B}{Z}\left(\frac{Z-A}{B}\right)^2\left[\frac{1}{\alpha}\left(\frac{Z-A+B}{Z-A}\right) - \left(\frac{Z-A+B}{Z-A}\right)^2\frac{1}{n\alpha}\right] \tag{6.29}$$

Focal length, f

$$\frac{\partial Z}{\partial f} = -\frac{T_o(n-1)}{n}\left(\frac{2\alpha f n}{D+2\alpha f} - 1\right)^{-2}\left[\frac{2\alpha n}{D+2\alpha f} - \frac{2\alpha f n}{(D+2\alpha f^2)}2\alpha\right] \tag{6.30}$$

Or

$$\frac{\partial Z}{\partial f} = -B\left(\frac{Z-A}{B}\right)^2\left[\frac{1}{f}\left(\frac{Z-A+B}{Z-A}\right) - \left(\frac{Z-A+B}{Z-A}\right)^2\frac{1}{nf}\right] \tag{6.31}$$

The relative range error,

$$\frac{\partial Z}{\partial f}\frac{1}{Z} = -\frac{B}{Z}\left(\frac{Z-A}{B}\right)^2\left[\frac{1}{f}\left(\frac{Z-A+B}{Z-A}\right) - \left(\frac{Z-A+B}{Z-A}\right)^2\frac{1}{nf}\right] \tag{6.32}$$

Refractive index, n

$$\frac{\partial Z}{\partial n} = \left[\frac{T_o}{n^2}\left(\frac{2\alpha f n}{D+2\alpha f} - 1\right)^{-1} - \frac{T_o(n-1)}{n}\left(\frac{2\alpha f n}{D+2\alpha f} - 1\right)^{-2}\left(\frac{2\alpha f}{D+2\alpha f}\right) - \frac{T_o}{n^2}\right] \tag{6.33}$$

Or

$$\frac{\partial Z}{\partial n} = \left[\frac{T_o}{n^2}\left(\frac{Z-A}{B}\right) - \frac{B}{n}\left(\frac{Z-A}{B}\right)^2\left(\frac{Z-A+B}{Z-A}\right) - \frac{T_o}{n^2}\right] \tag{6.34}$$

The relative depth error,

$$\frac{\partial Z}{\partial n}\frac{1}{Z} = \frac{1}{Z}\left[\frac{T_o}{n^2}\left(\frac{Z-A}{B}\right) - \frac{B}{n}\left(\frac{Z-A}{B}\right)^2\left(\frac{Z-A+B}{Z-A}\right) - \frac{T_o}{n^2}\right] \tag{6.35}$$

Based on the above theoretical equations, the following deductions can be made:

1) The absolute depth error is linearly related to the inaccuracy of T_o. However, it has little or no effect on the relative depth error when the range of the captured object increases. In other words, ∂T_o has the same effect for both close and far range. For this single-lens system, ∂T_o is relatively small as T_o can be adjusted by a precise motorized stand accurately.
2) Equations 6.28 and 6.29 shows that the absolute depth error increases quadratically with the range of the captured object while the relative depth error is proportional to

the $\partial\alpha$ and the range of the captured object. Thus, $\partial\alpha$ will induce more errors of the depth estimation for long range.

3) From Equation 6.32, the relative depth error increases linearly with ∂f and the range of the captured object.

4) For the refractive index, the relative depth error is not affected by the range of the captured object but increase linearly with ∂n.

The theoretical derivations will be verified with experiments in the next section.

6.2. Experiments

The main objective of the experiment is to verify the derived theoretical predictions in previous sections. Figure 6.2 shows the main experimental set-up comprising a bi-prism in prism holder, a CCD camera, a laser sensor and a calibration board on their respective mechanical stands. The specifications of the CCD cameras used in experiments are shown in Table 6.1.

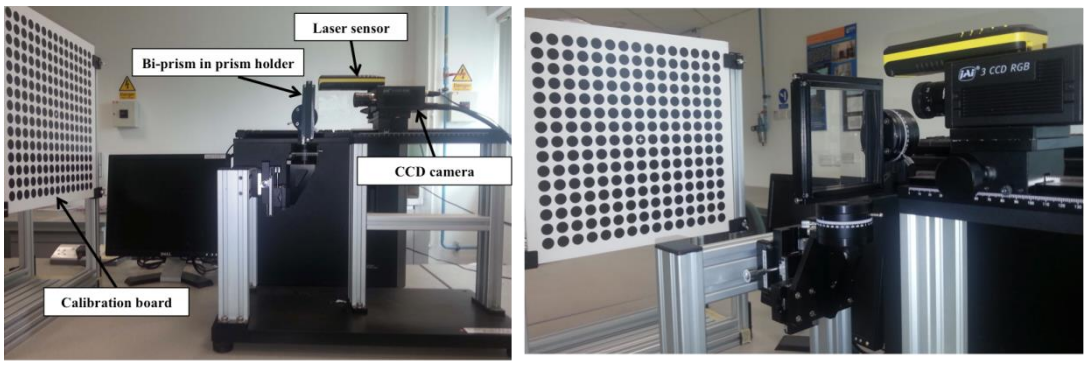


Figure 6.2 Experimental set-up

Table 6.1: The values of parameters of the system used in the experiment

α	n	f	T	T_o
6.4°	1.48	25 mm	5.6 mm	210 mm

6.2.1 Experiment Results

Figures 6.3 and 6.4 show the absolute depth and relative depth error due to the quantization/disparity error while Figure 6.5 shows the relative depth error of other parameters.

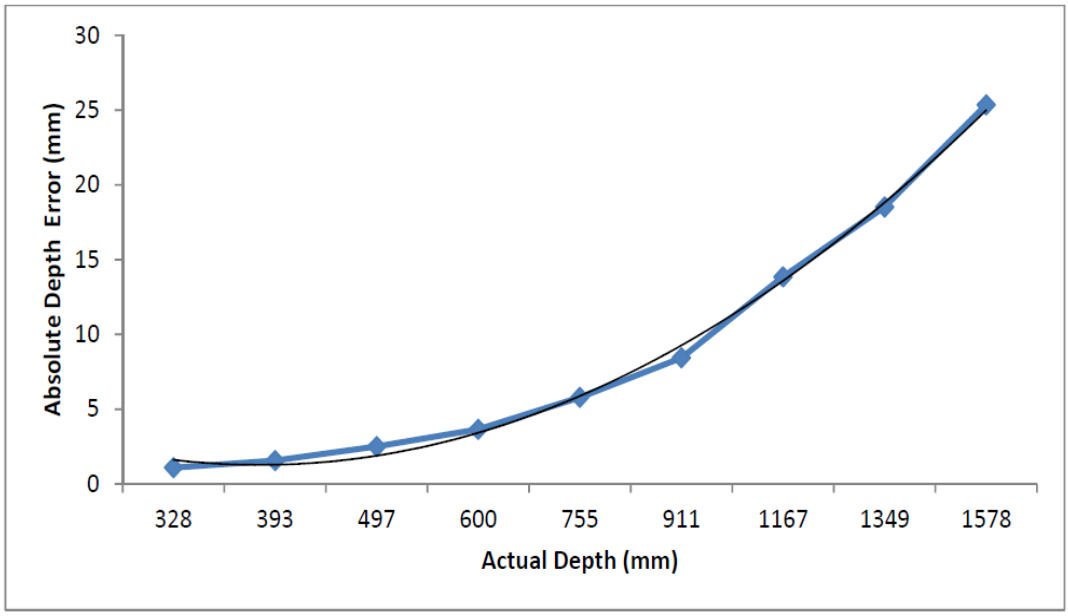


Figure 6.3 Absolute depth errors with vs. actual depth (quantization error, ∂D≈1 pixel)

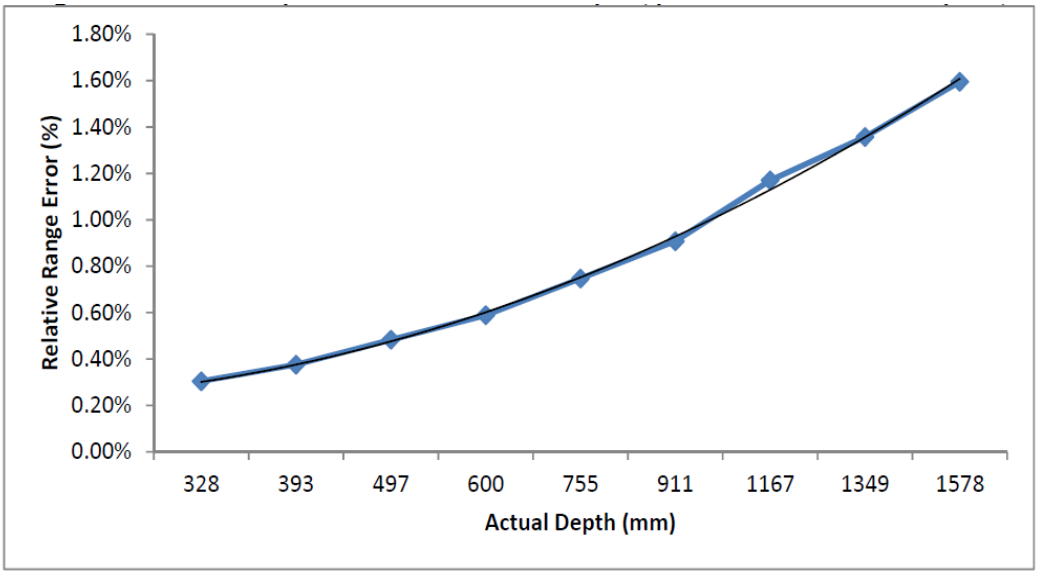


Figure 6.4 Relative depth errors with vs. actual depth (quantization error, ∂D≈1 pixel)

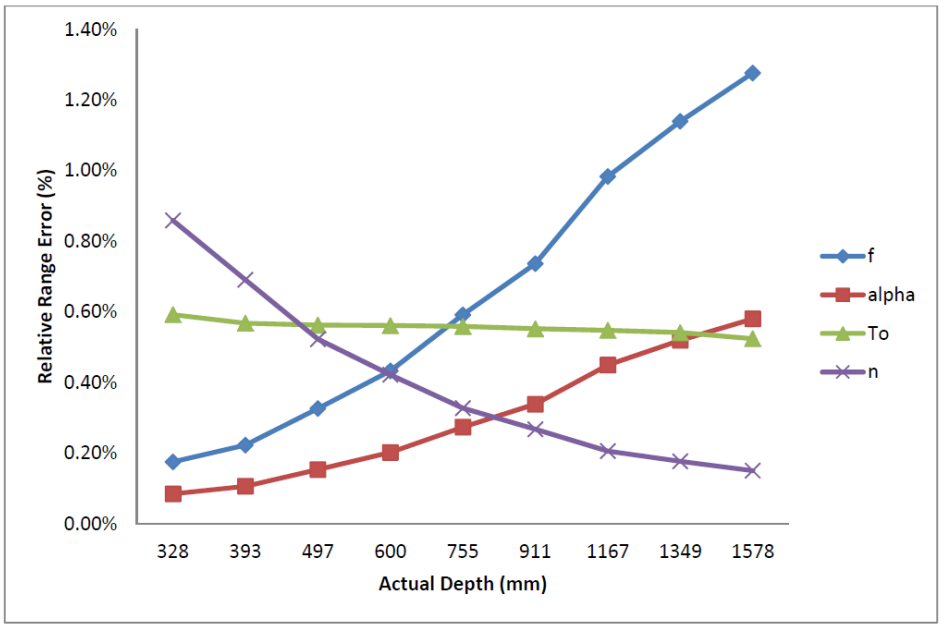


Figure 6.5 Relative depth errors with vs. other parameters ($\partial f \approx 0.1$mm, $\partial \alpha \approx 0.001$rad, $\partial To \approx 1$mm and $\partial n \approx 0.01$)

The overall absolute depth error and relative range error of the single-lens prism based system are estimated in Figure 6.6.

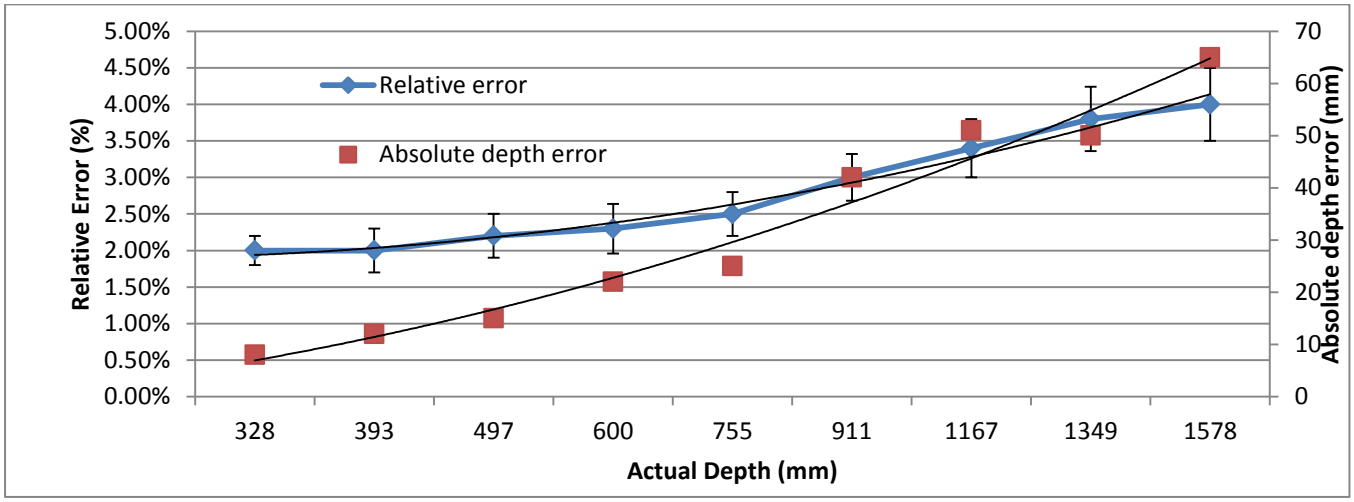


Figure 6.6 The estimated overall absolute depth and relative depth error

6.2.2 Discussion

Based on the results shown in Figures 6.3-6.6, a few conclusions can be made:

1. The relative depth error due to the quantization error (disparity error of 1 pixel) of this single-lens system increase linearly with the increase of the range of the

captured objects. As mentioned previously, quantization error is the hardware sampling error which is hardly avoided. In other words, even with perfect matching and calibration, this single-lens system is more accurate in reconstructing closer objects as compared to longer range objects.

2. The error of T_o, distance between the prism and the camera which is most likely caused by the inaccuracy of the mechanical stand has little or no effect on the trend of relative depth error.
3. Refer to Figure 6.5, for both the error of focal length and angle of the prism, the relative depth error becomes more significant when the objects are further away from the system. These are opposed to the error due to the refractive index, n which depicts an inverse relationship with the relative depth error.
4. The overall relative depth error increases (2-4%) with increasing object distance. This shows that this single-lens system is more accurate or suitable for closer range application. The quantization error is the most significant factor as compared to the other system parameters. Besides, if the errors of the stereo correspondence algorithm are taken into account, the disparity error will become even more significant.
5. The experiment results coincide with the theoretical predictions. In addition, the accuracy of the depth estimation using our proposed algorithm (Chapter 4) also exhibits the similar trend.

This study provides some understandings on the error sources and the range performance of the system. From the derived equations, we can deduce that in order to reduce the relative depth error, we can use larger values of n, α, f and T_o. n, α and f are the properties of the prism and camera lens. Although we can reduce the error by increasing the angle of the prism, it is impractical to use large angle prism as the common field of view of the system

will be reduced. Besides, the prism is made of glasses with refractive index of around 1.5 which is hardly changed. The possibilities of varying these parameters in order to reduce the error will be studied in detailed in the next section.

6.3 Study of Variable Parameters to Reduce Depth Error

In the previous section, it has been shown that the accuracy of the single-lens system depends on the parameters such as $\alpha, T_o, n,$ and f. Since the refractive index, n of the bi-prism is a constant value that cannot be changed, we will focus on the other three parameters in this section. The objective of this study is to examine the possibility in manipulating the parameters of the system in order to obtain an acceptable accuracy for both short and large depth. The concept is also known as *variable parameter*. In other words, by selecting appropriate values for the system parameters, the depth accuracy can be maintained constant for both near and long range. Recall the error equation derived previously,

$$\partial Z = \left(Z - \frac{T(n-1)}{n}\right)^2 \frac{1}{T_o(n-1)2\alpha f} \partial D \tag{6.36}$$

We had shown that when the distance of the object, Z increases, the error or the relative depth errors of the system also increases. At a first glance on Equation 6.36, reducing the values of α, f and T_o will improve the accuracy as all these parameters are at the denominator of the equation.

6.3.1 Variable focal length, f

The focal length of the camera can be increased either by increasing the resolution of the sensor or by narrowing the field of view (zoom). The relationship between the resolution, field of view and focal length can be written as

$$\frac{h}{2} = f \tan \theta_{vc} \tag{6.37}$$

Aspect ratio of the camera, $a = \frac{h}{v}$ (6.38)

where h and v are the horizontal and vertical resolution respectively and θ_{vc} is the field of view of the virtual camera. These parameters are shown geometrically in Figure 6.7.

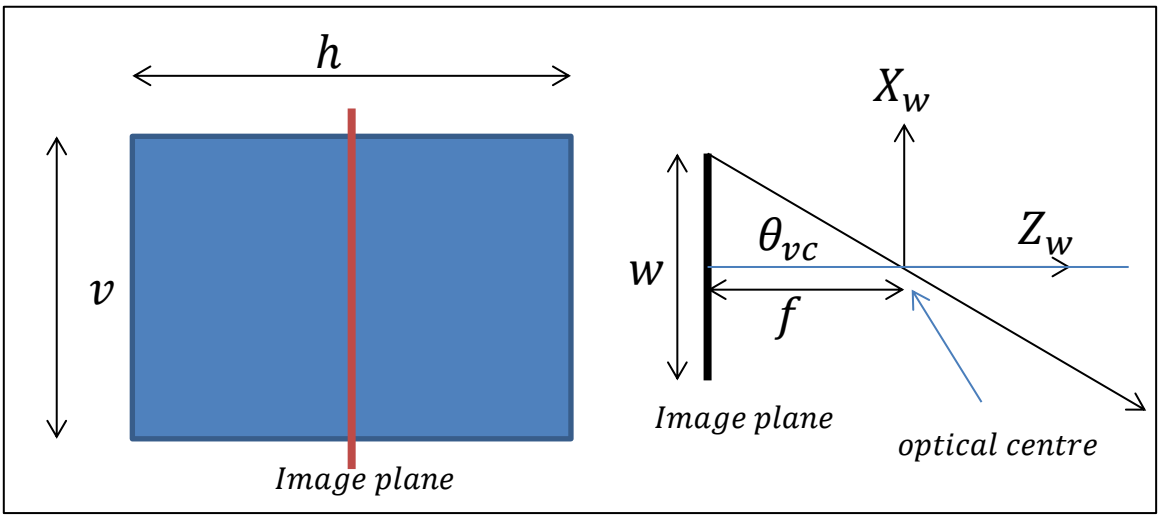


Figure 6.7 Relationship between the resolutions, field of view with focal length

Higher resolution increases the computation time of the stereo correspondence algorithm and there is always a hardware limitation. This will be examined by theoretical derivations. By assuming that the field of view, θ_{vc} has been carefully chosen for a specific application, the target accuracy of the system can be determined by solving for f in the equation below (rearranged from Equation 6.36),

$$f = \left(Z - \frac{T(n-1)}{n}\right)^2 \frac{1}{T_o(n-1)2\alpha} \frac{\partial D}{\partial Z} \tag{6.39}$$

$$\text{No. of pixels} = \frac{hv}{2} = \frac{h^2}{2a} = \frac{2f^2 \tan^2 \theta_{vc}}{a} \tag{6.40}$$

Consider the hardware sampling error, $\partial D \approx 1$ pixel error and combine Equations 6.39 and 6.40,

$$\text{Number of pixels} = \frac{2\tan^2 \theta_{vc}}{a} \left(Z - \frac{T(n-1)}{n}\right)^4 \frac{1}{(T_o(n-1)2\alpha)^2} \frac{1}{\partial Z^2} \tag{6.41}$$

Equation 6.41 shows that increasing the resolution alone to achieve the target accuracy requires the image resolution to grow proportionally to $\left(Z - \frac{T(n-1)}{n}\right)^4$. However, using a higher resolution camera does not necessarily increase the effective resolution.

Besides, a better quality of camera lens and hardware may be required but this will increase the cost of the overall system. Higher resolution also means longer processing time. The cost of stereo correspondence algorithm will become higher due to larger disparity range or longer epipolar line. In other words, the improvement of the accuracy by increasing the focal length or resolution is severely limited!

6.3.2 Variable bi-prism angle, α

Consider the geometry of the bi-prism as shown in Figure 6.8, bi-prism angle, α can be changed either by increasing the bi-prism thickness, T or by reducing the length of the back plane, t_b. However, t_b should be fixed at a given specific field of view (see Figure 6.9). Its length has to be long enough to cover the field of view of the camera but not too long as this will make the system bulky.

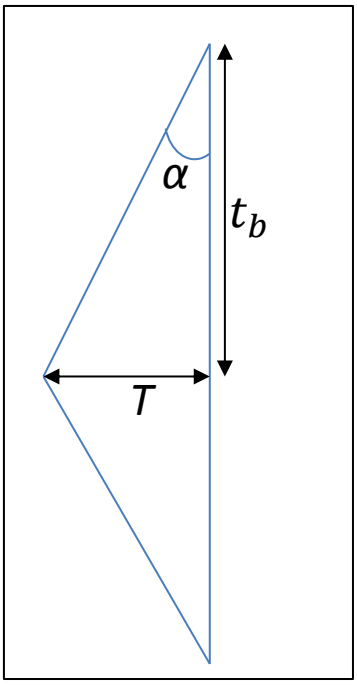


Figure 6.8 2D schematic of the bi-prism geometry

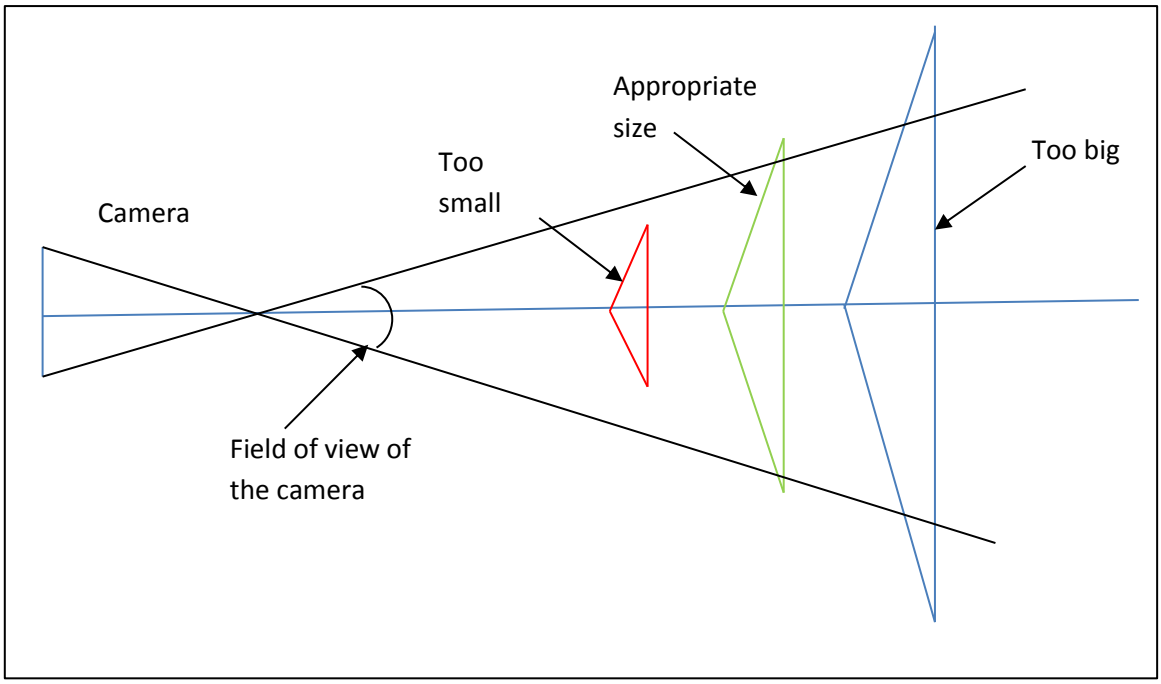


Figure 6.9 Selection of the bi-prism size based on field of view of the camera

The relationship of the bi-prism parameters can be written as (assume small angle α),

$$tan\,\alpha = \frac{T}{t_b} \approx \alpha \tag{6.42}$$

The bi-prism thickness required to meet the target accuracy can be determined by combining Equations 6.36 and 6.42,

$$\partial Z = \left(Z - \frac{T(n-1)}{n}\right)^2 \frac{1}{T_o(n-1)2\frac{T}{t_b}f}\partial D \tag{6.43}$$

Expand and rearrange Equation 6.43 (assume $\partial D \approx 1$ pixel error),

$$T^2 - \left(\frac{2nZ}{n-1} + \frac{2n^2T_of}{(n-1)t_b}\partial Z\right)T + \left(\frac{n}{n-1}Z\right)^2 = 0 \tag{6.44}$$

Solving the quadratic equation for T, we will obtain two solutions,

$$T^+ = \left(\frac{nZ}{n-1} + \frac{n^2T_of}{(n-1)t_b}\partial Z\right) + \sqrt{\left(\frac{nZ}{n-1} + \frac{n^2T_of}{(n-1)t_b}\partial Z\right)^2 - \left(\frac{nZ}{n-1}\right)^2} \tag{6.45}$$

$$T^- = \left(\frac{nZ}{n-1} + \frac{n^2T_of}{(n-1)t_b}\partial Z\right) - \sqrt{\left(\frac{nZ}{n-1} + \frac{n^2T_of}{(n-1)t_b}\partial Z\right)^2 - \left(\frac{nZ}{n-1}\right)^2} \tag{6.46}$$

Using the real single-lens system parameters (see Table 6.2), T^+ and T^- were solved using Equations 6.45 and 6.46. They were plotted against the object's depth, Z and the ∂Z is calculated based on the theoretical predictions in Section 6.1. We found that the T^+ solution does not converge to the its real value of 5.6mm as shown in Figure 6.10 while T^- solution can be approximated to its real value (see Figure 6.11). In other words, the T^+ solution or Equation 6.45 is impractical in describing the physical single-lens system.

Table 6.2 Real system parameters

n	T_o	f	t_b	T
1.48	150mm	25mm	50mm	5.6mm

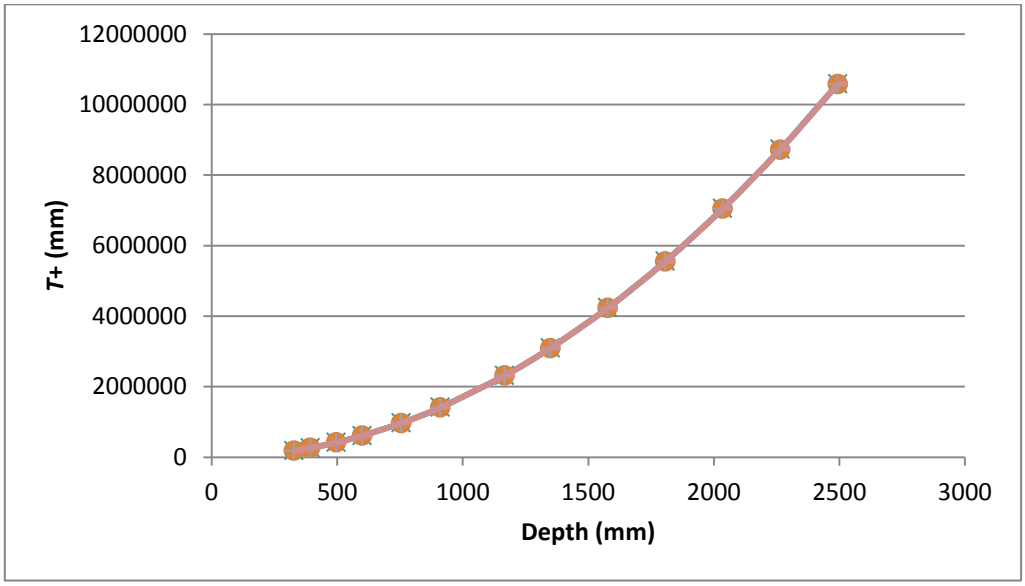


Figure 6.10 Non-convergence of T^+ solution

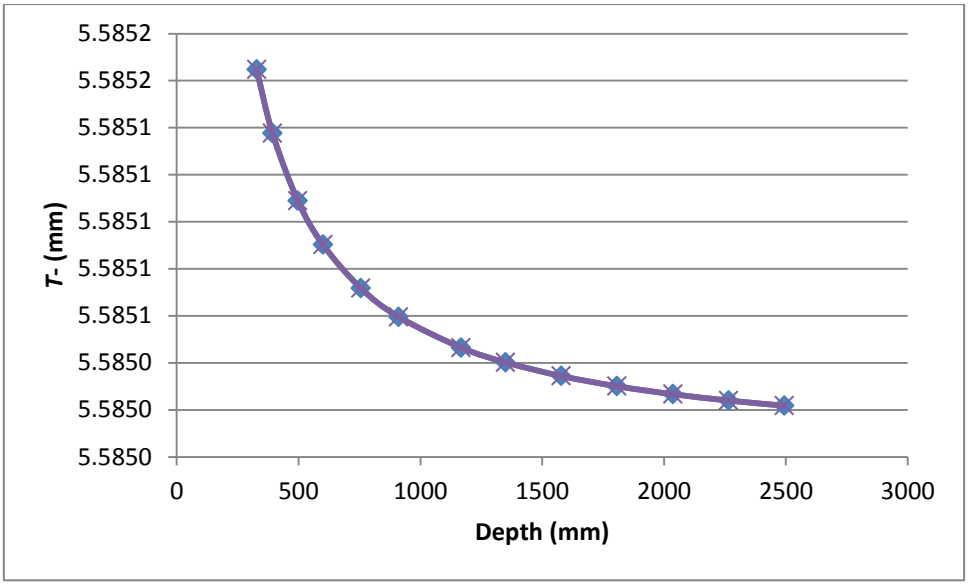


Figure 6.11 T^{-} solution can be approximated to the real value of T

Hence, the thickness of the bi-prism, T and the angle of the bi-prism, α to achieve target accuracy with respect to the object's depth are determined using the mathematical Equation 6.46. The results are shown in Figures 6.12 and 6.13.

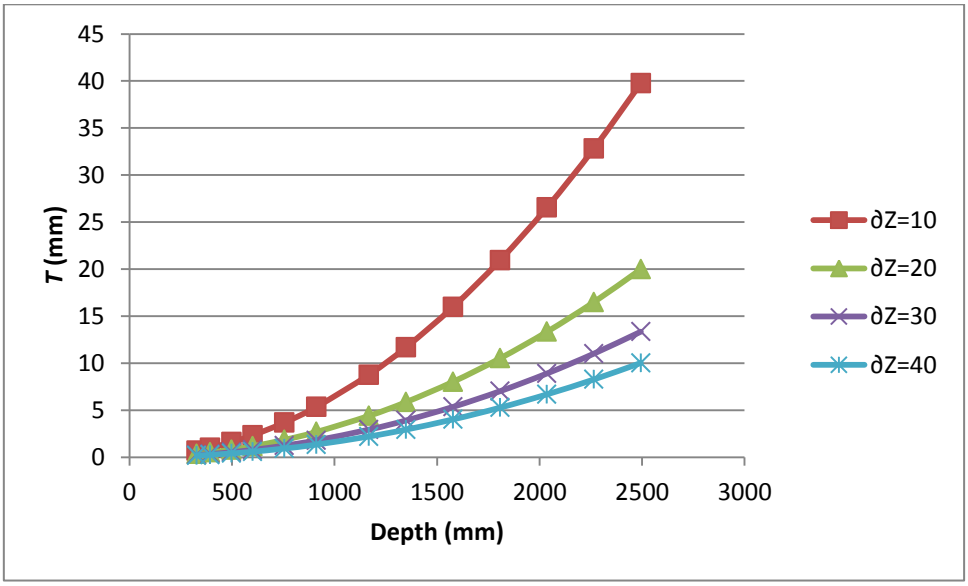


Figure 6.12 Value of T required to obtain the absolute error of 10mm-40mm

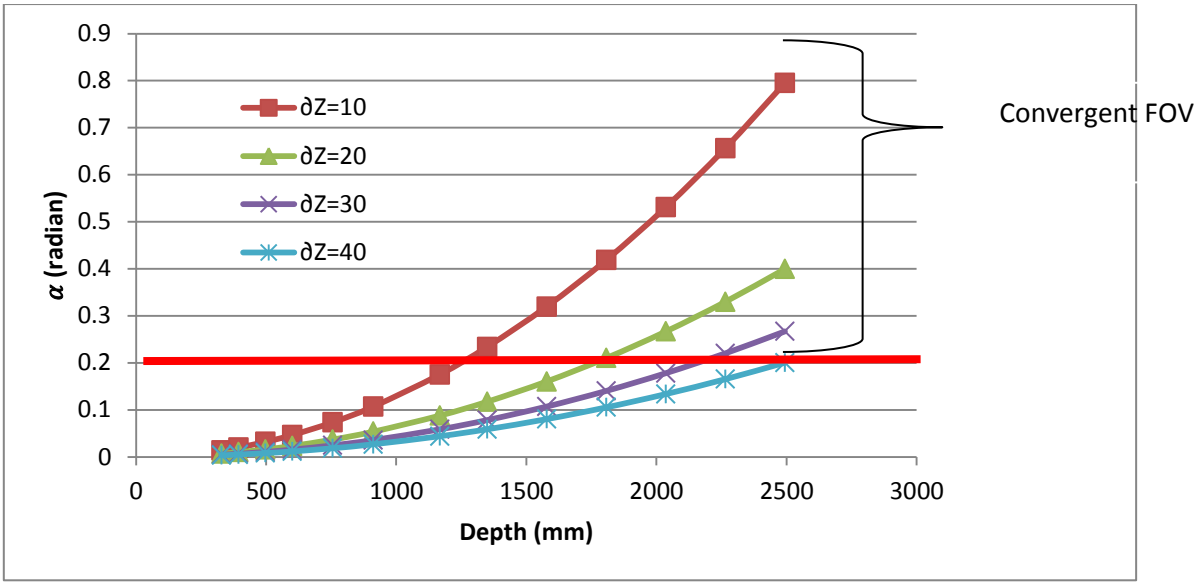


Figure 6.13 Value of α required to obtain the absolute error of 10mm-40mm

As shown in Figure 6.13, to achieve a better accuracy when the object's distances increase, the bi-prism angle has to be increased. Varying the bi-prism angle seems to be an effective method in controlling the error of the system. However, the results in Chapter 5 show that, for $\alpha > 0.2$ radians with focal length of 25mm, the system will produce a convergence field of view (near range system with finite field of view) with limited usages at the near range. More details can be found in Chapter 5. In other words, one cannot rely on increasing the bi-prism angle alone to meet the target accuracy especially for long range application.

6.3.3 Variable T_o

If the focal length and bi-prism angle are held fixed, the depth error of the system can only be improved by varying the distance between the camera with the bi-prism apex, T_o. Thus, equation 6.36 is solved for T_o,

$$T_o = \left(Z - \frac{T(n-1)}{n}\right)^2 \frac{1}{(n-1)2\alpha f} \frac{1}{\partial Z} \tag{6.47}$$

Figure 6.14 shows the theoretical results in varying T_o. A better accuracy can be achieved by using a larger value of T_o. It is a useful parameter in improving the accuracy without

changing any hardware. Varying the focal length and bi-prism angle requires us to use different camera lens and bi-prism which also affect the field of view and system types. On the other hand, varying T_o will shift the field of view of the virtual cameras slightly but leaving the common field of view unchanged as shown in Chapter 5. However, increasing T_o produces a bulky system as the bi-prism has to be placed further away from the camera which is not favoured in some applications with space constraint.

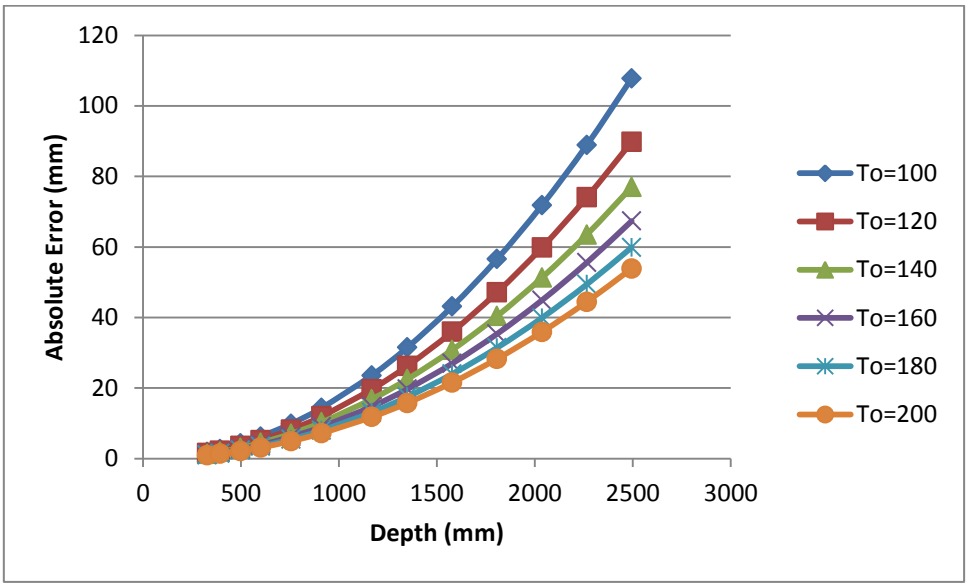


Figure 6.14 Absolute depth errors with different values of T_o

6.4 Experiments

The main objective of the experiments is to verify the correctness of the theoretical predictions on variable parameters. Table 6.3 shows the system parameters used in the experiments.

Table 6.3 System Parameters

α	n	f	T_o
6.4° and 21.6°	1.48	4 and 8 *mm*	100-200 *mm*

The first experiment was to verify the theoretical prediction on the trend of error with varying focal length, f. 4 and 8mm lens were used to capture calibration boards at different

distances as shown in Figure 6.15. 50 pairs of correspondence points were detected using our algorithm discussed in Chapter 4 and their average depths were calculated. The results are shown in Figure 6.16.

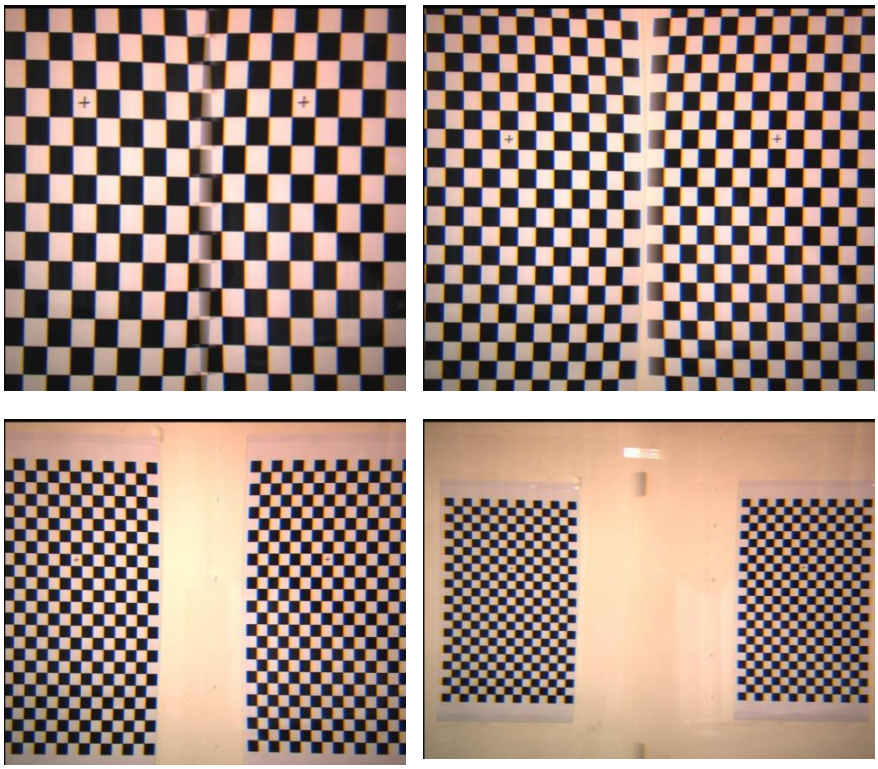


Figure 6.15 Sample images captured at different distances using focal length of 8mm, bi-prism angle of 6.4° and T_o of 100mm

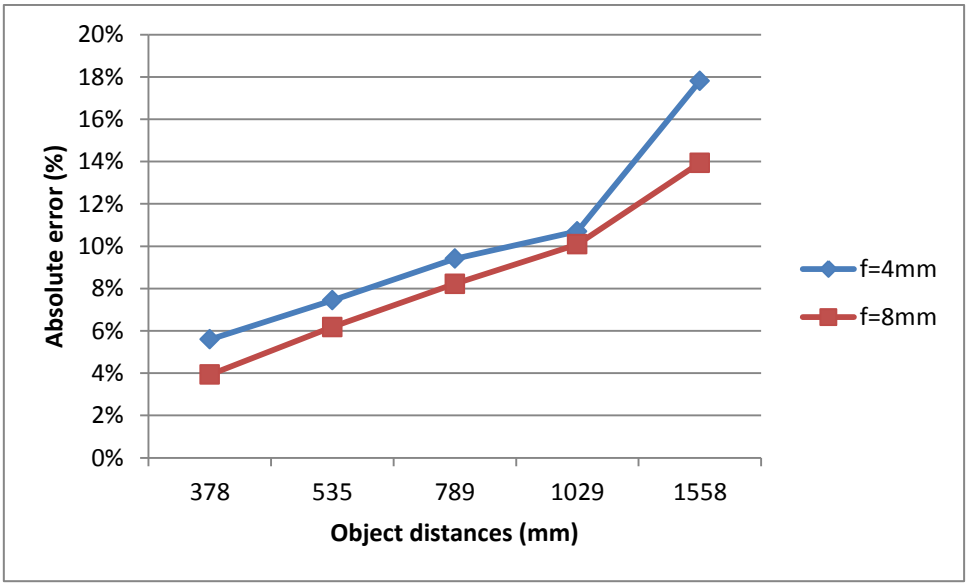


Figure 6.16 Absolute error of the system using bi-prism angle of 6.4°, T_o of 100mm with 4mm and 8mm focal lengths

As shown in Figure 6.16, the system with 8mm focal length has slightly lower error as compared to the system with 4mm focal length. The results show similar trend to our theoretical predictions in the previous section. The error differences of the two setups are small which is also coincide with our predictions that the resolution of the camera, which is proportional to the focal length, is proportional to $\left(Z - \frac{T(n-1)}{n}\right)^4$.

In order to study the effect of variable bi-prism angle (α), a fix focal length of 8 mm is used with a T_o of 100mm. Thus, there will be two setups: one with bi-prism angle of 6.4° and the other one with bi-prism angle of 21.6°. The results are shown in Figure 6.17.

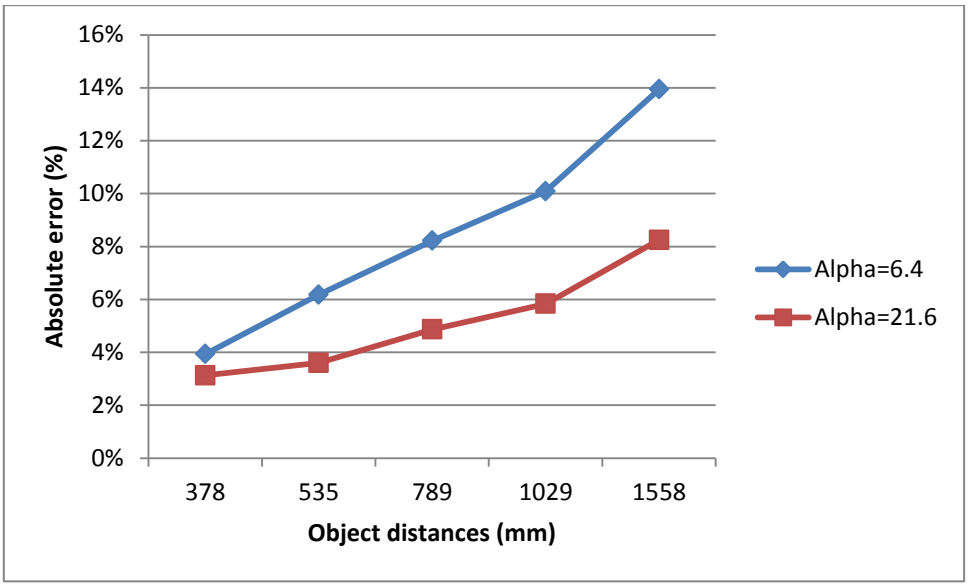


Figure 6.17 Absolute error of the system using T_o of 100mm, focal lengths of 8mm with bi-prism angle of 6.4° and 21.6°

Clearly, the larger the angle of the bi-prism the better the accuracy is. This exhibits the similar behaviour as compared to the theoretical estimations. The final experiment was to fix the focal length and bi-prism angle but vary the value of T_o. The results are shown in Figure 6.18.

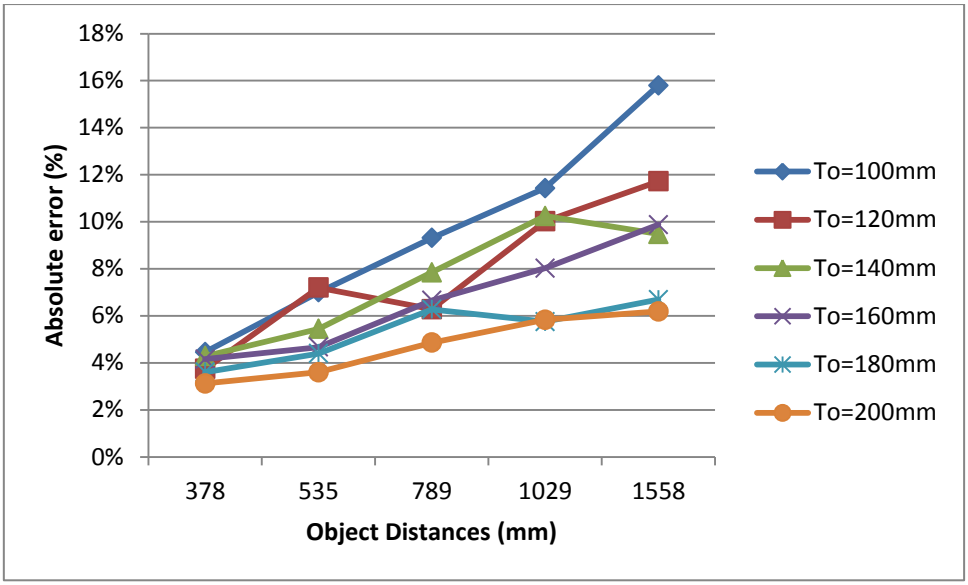


Figure 6.18 Absolute error of the system using focal lengths of 8mm and bi-prism angle of 6.4° with different values of T_o

Refer to Figure 6.18, it is obvious that the increase in T_o improves the accuracy of the system especially for the long range. Based on the experimental results, the following conclusions can be made:

1. Overall, the experiment results show similar behaviour to our theoretical predictions.
2. The errors of the experiment results are larger due to the correspondence error and other sources of error such as distortions. In our theoretical derivations, we assumed the disparity error is only one pixel.
3. Experimentally, T_o and bi-prism angle seems to be good parameters to control the error for both short and long range. However, large value of bi-prism angle will produce a convergent system with limited field of view at near range.
4. The theoretical derivations can be used to predict the error and improve the accuracy of the bi-prism system which will then facilitate the parameter selection process.

6.5 Summary

A general description on the effect of the system parameters of the single-lens prism based stereovision system on the depth accuracy is presented. No study has been done to investigate how various system parameters affect the accuracy of the depth determination. With the development of various relative depth equations as well as validation of these theories with the actual system setup, this study helps to establish a new understanding specifically on the effect of the system parameters on its depth accuracy. The overall error of the system can be predicted once the parameters are finalised. This new understanding on single-lens bi-prism stereovision provides a good foundation for future studies to build upon. These results also add value to engineering processes by allowing engineers to design and evaluate the feasibility of a specific set-up for a particular application, and thus select the best system parameters to produce the most effective outcomes. Furthermore, the concept of variable parameter provides a room for improvement to further reduce the error of the system.

Chapter 7 Conclusions and Future Work

A single-lens prism based stereovision system has been developed, analyzed, implemented and tested. The main objective of the work reported in this thesis is to study the recovery of the depth of a 3D scene using a single-lens prism based stereovision system. The works in this thesis cover: (1) virtual camera calibration, (2) stereo correspondence and depth recovery, (3) determining the field of view of the system, and (4) parameter/error analysis.

In our single-lens stereovision system, a prism is placed in front of the camera to generate multiple views of the same scene. Each image captured by this system can be split into two sub-images and these sub-images can be considered as the images simultaneously captured by the virtual cameras generated by the prism. The advantages of this kind of single-lens stereovision system are its compactness, lower cost, automatic synchronization in view capturing and ease of implementation. Unlike a conventional stereovision system which employs more than one camera, the system used in this thesis employs only one camera. However, having to deal with virtual cameras would add a certain degree of difficulty in the analysis of the system. Besides, this system does have two disadvantages: firstly, all the virtual cameras created by this system share one CCD array, and hence each virtual camera would have fewer CCD elements to represent a captured image; secondly, the common view zone and the baseline between virtual cameras are constrained by the size and shape of the prism, and hence, at the present moment, this system possibly can only find its main application involving short-range stereovision.

7.1 Contributions of the thesis

The contributions of this thesis are summarized in the following sections:

(a) Virtual camera calibration

The aim of this approach is to establish a relationship between the projected digital images of the virtual cameras and the scene - from the 3D World Coordinates of the scene to the 2D Digital Virtual Image Plane. In chapter 3, geometry based approach was presented for virtual camera calibration. This approach also helps to determine the intrinsic and extrinsic parameters of the system without having to carry out the complex calibration process. Experimental results are encouraging, which show that this approach is valid. The results also showed that the proposed approach provides a much simpler implementation requirement and better accuracy as compared to the conventional calibration method. However, this method is applicable only for a thin lens system which is relatively free of distortion.

(b) Stereo correspondence algorithm

Stereo correspondence is a process to search for corresponding points in all the stereo images in order to recover the depth of an object. The search for correspondence points is a fastidious task, more so if the image planes are not co-planar. Our calibration results in Chapter 3 showed that the two virtual cameras of the single-lens system are not co-planar. With a given known point in one of the captured images, the candidates of the corresponding points in the remaining image(s) can be easily determined by making use of the epipolar geometry constraint. This is because the candidate points forms the epipolar line and one can search for the correct corresponding point without much difficulty along the epipolar line. In Chapters 4, an algorithm which solves the binocular stereovision correspondence problem for two views is proposed and verified. The algorithm of constructing virtual epipolar lines using two unique points is developed to facilitate the correspondence search. The major advantage of the two aforesaid proposed approaches is: they can be generalized to handle, from binocular stereovision system to tri-ocular and even multi-ocular stereovision systems. The experimental results validate the reliability and practicality of the proposed virtual epipolar

line approach. Besides, the proposed approach also outperformed the conventional methods with better accuracy.

(c) Field of view of the system

In the literature, few studies on single-lens bi-prism stereovision system have been done to investigate how various system parameters affect its FOV. Hence, one of the main objectives of this thesis is to establish a new understanding specifically on the effect of bi-prism angle as well as the position of the bi-prism on its FOV. With the development of various theories and validation of these theories in experiments, it is conclusive to say that the above objective has been achieved. Our theoretical model showed that the type of FOV corresponding to a fixed bi-prism angle can be predicted by using the coefficient of FOV, ε,. Theoretical linear equations of rays 1 and 2 (Section 5.2.2) can be utilized to estimate the geometry of a specific FOV type. While the common FOV region is largely unaffected by translation of bi-prism in z-axis, translation in x-axis causes the region to rotate about the y-axis, and. The common FOV angle, γ remains constant under bi-prism translation in both z- and x-axes.

These new understandings on single-lens bi-prism stereovision provide a fundamental background for future studies to build upon. These results can also add value to engineering processes by allowing engineers to design and evaluate the feasibility of a specific set-up for a particular application, and thus select the best system parameters to produce the most effective outcomes.

(d) Parameter analysis

In Chapter 6, a general description on the effect of the quantization error and the system parameters of the single-lens prism based stereovision system is presented. In literature, there is no studies on single-lens bi-prism stereovision system have been done to

investigate how various system parameters affect its depth estimation accuracy. With the development of various relative depth equations and validation of these theories with the actual system setup, it is conclusive to say that this thesis has established new understanding specifically on the effect of the system parameters on its depth accuracy. The overall error of the system is predicted. Furthermore, the idea of variable parameter is presented. The main objective is to study the possibility in manipulating the parameters of the single-lens system in order to obtain an acceptable accuracy for both short and long depth applications.

The proposed single-lens prism based stereovision system and its associated 3-D depth recovery approaches derived in this work could have compelling advantages and provide significant improvement in potential applications, such as indoor robot navigation, object detection, small size hand-hold stereovision system for dynamic scene, and economical 3D endoscopy, etc.

7.2 Future work

In this thesis, the developed algorithms for virtual camera calibration and stereo correspondence using a single-lens prism based stereovision have been shown to be better than many existing ones. Besides, a new understanding on the field of view and parameter analysis of the system are presented in detail. However, there are still some limitations and for future endeavors in this research, some recommendations are suggested as follows:

1) Develop calibration algorithm by taking into consideration the lens-distortion, system alignments errors, light refraction errors, and light diffraction errors. The effect of the distortion will be even more severe when the system is used in endoscopic application when the camera lens is smaller in diameter. The proposed geometry based algorithm does not take into account the lens and prism distortion problem. The contributions of

these possible sources or errors should be carefully studied in detail. Both model based and model free approach can be used to study the problem.

2) Stereo correspondence is always the main and most complex problem in stereovision. In our studies, we found that the disparity error is the main source of error of the single-lens system. Thus, in future work, the accuracy of the algorithm must be examined closely as they are the important preludes to the determination of accurate disparity map. An accurate reconstruction depends largely on an accurate disparity map. Besides, the speed of the algorithm should be improved so that it is feasible to be applied in real time applications.
3) As the accuracy of the position and orientation of the system setup affects greatly on the system accuracy, future studies should include the investigation on the effect of mis-alignments of prism, camera etc. This will help to design measures to counter the problem.
4) We note that the range on the recoverable depth depends narrowly on the prism geometry. The current study only makes use of prism of two different prism corner angles. More prisms with different geometrical properties should be used to further investigate the system performance. This will extend the recoverable depth of the system. Besides, the common FOV is always the main concern of this system as it is usually small. Different design of the prism such as curve surface could be considered to overcome this problem.
5) The idea of variable parameter should be integrated into the stereo correspondence algorithm to improve the accuracy of the system for long range. An active system that can vary its system parameters to suit different situations (short or long range) can be designed by making use the variable parameter concept.

6) The detailed studies of the effect of the system parameters on the FOV and accuracy have been established in this thesis. Using this knowledge as the foundation, the study of parameter analysis should be extended to the baseline and orientation of the virtual cameras. These topics are important in improving the accuracy of the system.

List of Publications

1. M. J. Zhao, K. B. Lim, and W. L. Kee, Geometrical-analysis-based algorithm for stereo matching of single-lens binocular and multi-ocular stereovision system. In *Journal of Electronic Science and Technology*, 10(2), 107-112, June 2012.

2. K. B. Lim, D. L. Wang, and W. L. Kee, Virtual cameras rectification with geometrical approach on single-lens stereovision using a biprism. In *Journal of Electronic Imaging*, 21(2), 2012.

3. W. L. Kee, K. B. Lim, and D. L. Wang, Virtual epipolar line construction of single-lens bi-prism stereovision system. In *Journal of Electronic Science and Technology*, 10(2), 97-101 June 2012.

4. D. L. Wang, Kah Bin Lim, and W. L. Kee, Geometrical approach for rectification on single-lens stereovision using a triprism. In *Machine Vision and Applications*, 24(4), 821-833, 2012, doi:10.1007/s00138-012-0467-8.

5. X. Y. Cui, K. B. Lim, and W. L. Kee, and Q. Y. Guo, System optimization for prism based stereovision. In *Advances in Materials Science and Engineering*, 650, 374-378, 2013.

6. K. B. Lim, W. L. Kee, and D. L. Wang, Virtual camera calibration and stereo correspondence of single-lens bi-prism stereovision system using geometrical approach. In *Signal Processing: Image Communication*, 28(9), 1059-1071, October 2013.

7. X. Y. Cui, K. B. Lim, Y. Zhao, and W. L. Kee, Single-lens stereovision system using prism: position estimation of multi-ocular prism. In *Journal of the Optical Society of America A*, 31(5), 1074-1082, 2014.

8. W. L. Kee, K. B. Lim, Z. L. Tun, and Y.D Bai, New Understanding on the effect of angle and position of bi-prism on single-lens bi-prism based stereovision system. In *Journal of Electronic Imaging*, 23(3), May 2014.

9. W. L. Kee, K. B. Lim, and Y. D. Bai, Quantization and Parameter Analysis of Single-lens Prism based Stereovision System. In *Journal of the Optical Society of America A*, (accepted on 14 November 2014).

Bibliography

[1] D. Kim, S. Lee and B. Choi, A Real-time Stereo Depth Extraction Hardware for Intelligent Home Assistant Robot. In *IEEE Transactions on Consumer Electronics,* 56(3), 1782-1788, 2010.

[2] T. P. Pachidis and J. N. Lygouras, Pseudostereo-Vision System: A Monocular Stereo-Vision System as a Sensor for Real-Time Robot Applications. In *IEEE Transactions on Instrumentation and Measurement,* 56(6), 2547-2560, 2007.

[3] A. Naoulou, J. Boizard, J. Fourniols and M. Devy, A 3D Real-time Vision System Based on Passive Stereovision Algorithms: Application to Laparoscopic Surgical Manipulations. In *Information and Communication Technologies,* 1(2), 1068-1073, 2006.

[4] P. Zhou, Y. Liu and Y. Wang, Pipeline Architecture and Parallel Computation-Based Real-Time Stereovision Tracking System for Surgical Navigation. In *IEEE Transactions on Instrumentation and Measurement,* 59(5), 1240-1250, 2010.

[5] Y. Jianxi, Q. Jinwu, Y. Juxia and F. Zengliang, Research on Computer Aided Surgical Navigation based on Binocular Stereovision. In *Mechatronics and Automation, Proceedings of the 2006 IEEE International Conference,* 1532-1536, 2006.

[6] R. Racca and J. Dewey, A method for Automatic Particle Tracking in a Three-dimensional Flow Field. In *Exp. Fluids,* 6(1), 25-32, 1988.

[7] J. Kent, A. Mikulec, L. Rimai, A. Adamczyk, S. Mueller and R. Stein et al, Observation on the Effects of Intake-generated Swirl and Tumble on Combustion Duration. In *SAE Tech. Pap.,* 1989.

[8] D. Barnhart, Adrian RJ and G. Papen, Phase-conjugate Holographic System for High Resolution Particle Image Velocimetry. In *Appl. Opt*, 33(30), 7159-7170, 1994.

[9] J. Wang, P. Huang, C. Chen, W. G and J. Chu, Stereovision Aided Navigation of an Autonomous Surface Vehicle. In *Advanced Computer Control (ICACC), 3rd International Conference*, 130-133, Jan. 2011.

[10] D. Roble, Vision in Film and Special Effects. In *Computer Graphics*, 33(4), 58–60, 1999.

[11] Y. Y. Chuang, et al., Video Matting of Complex Scenes. In *ACM Transactions on Graphics*, 21(3), 243–248, 2002.

[12] N. Snavely , S. M. Seitz, and R. Szeliski, Photo Tourism: Exploring Photo Collections in 3D. In *ACM Transactions on Graphics*, *25(3)*, 835–846, 2006.

[13] M. Goesele, et al., Multi-view Stereo for Community Photo Collections. In *Tenth International Conference on Computer Vision*, Rio de Janeiro, Brasil, 2007.

[14] H. Sidenbladh, and M. J. Black, Learning the Statistics of People in Images and Video. In *International Journal of Computer Vision*, *54(1)*, 189–209, 2003.

[15] J. Sivic, C. L. Zitnick, and R. Szeliski, Finding People in Repeated Shots of the Same Scene. In *British Machine Vision Conference* (BMVC 2006), 909–918, Edinburgh, 2006.

[16] Y. Xiao and K. B. Lim, A prism-based Single-lens Stereovision System: From Trinocular to Multi-ocular. In *Image and Vision Computing*, 25(11), 1725-1736, 2007.

[17] K. B. Lim and Y. Xiao, Virtual Stereovision System: New Understanding on Single-lens Stereovision Using a Biprism. In *Journal of Electronic Imaging*, 14(4), 043020-043020, 2005.

[18] X. Y. Li and R. Wang, Analysis And Optimization Of The Stereo System With A Biprism Adapter. In *International Conference on Optical Instruments and Technology*, 2009.

[19] Lee V. P., Stereovision Using A Single CCD Camera, *Master Thesis*, Department of Mechanical Engineering, NUS, 2001.

[20] Xiao Y, Stereovision Using Single CCD Camera, *PhD Thesis*, Department of Mechanical Engineering, NUS, 2000.

[21] Zhao M. J., Stereo Correspondence and Depth Recovery of Single-lens Bi-prism based Stereovision System, *PhD Thesis*, Department of Mechanical Engineering, NUS, 2013.

[22] Wang D. L., Depth Recovery With Rectification Using Single-lens Prism Based Stereovision System, *PhD Thesis*, Department of Mechanical Engineering, NUS, 2013.

[23] K. B. Lim, and M. J. Zhao, Stereo matching of single-lens bi-prism based stereovision system. In *Journal of Procedia Engineering*, ISSN: 1877-7058, ELSEVIER, 2011.

[24] D. L. Wang and K. B. Lim, Obtaining depth map from segment-based stereo matching using graph cuts. In *Journal of Visual Communication and Image Representation*, 22(4), 325-331, May 2011.

[25] X. Y. Cui, K. B. Lim, Y, Zhao, and Q. Y. Guo, Position estimation method for prism based stereovision system. In *AASRI Conference on Modelling, Identification and Control*, 3, 481-487, 2012.

[26] X. Y. Cui, K. B. Lim, Q. Y. Guo, and D. L. Wang, An accurate geometrical optics model for single-lens stereovision system using prism. In *Journal of the Optical Society of America A (JOSA A)*, 29(9), 1828-1837, 2012.

[27] K.B. Lim, and Y. Xiao, New understanding of single-lens stereovision using a biprism. In *Proc. of IS&T SPIE Symposium on Electronic Imaging*, San Jose, 2004, 5302, 41-52.

[28] Y. Xiao, and K. B. Lim, A single-lens trinocular stereovision system using a 3f filter. In *Proc. of IEEE Conference on Robotics, Automation and Mechantronics*, Singapore, 2004, 306-400.

[29] Y. Xiao, K. B. Lim, and W. M. Yu, A prism based single-lens multi-ocular stereo lmage capture system. In *Multi-Media Modelling Conference Proceedings*, Beijing, 2006 12th International, 469 – 472.

[30] M. J. Zhao, and K. B. Lim, Stereo correspondence problem of single-lens bi-prism stereovision system. In *Int. Conf. of ISDM*, Wuhan, China, 2009, 244-250.

[31] D. L. Wang, and K. B. Lim, A new segment-based stereo matching using graph cuts. In *Computer Science and Information Technology (ICCSIT)*, Chengdu, China, 2010, 410-416.

[32] M. J. Zhao, and K. B. Lim, Stereo matching of single-lens bi-prism based stereovision system. In *Int. Conf. of ICFIT*, Changsha, China, 2010, 179-183.

[33] K. B. Lim, D. L. Wang, and W. L. Kee, Virtual cameras rectification with geometrical approach on single-lens stereovision using a biprism. In *Journal of Electronic Imaging*, 21(2), 2012.

[34] D. L. Wang, K. B. Lim, and W. L. Kee, Geometrical approach for rectification on single-lens stereovision using a triprism. In *Machine Vision and Applications*, 24(4), 821-833, 2012, doi:10.1007/s00138-012-0467-8.

[35] Emanuele Trucco and Alessandro Verri, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 2006.

[36] L. Grewe and A. Kak, Handbook of Pattern Recognition and Image Processing: Computer Vision. In *Academic Press Inc.*, 2, 239-317, 1994.

[37] Y. Nishimoto and Y. Shirai, A Feature-based Stereo Model Using Small Disparities. In *Proceedings of Computer Vision and Pattern Recognition*, 1987, 192-196.

[38] C. Gao and N. Ahuja, A Refractive Camera for Acquiring Stereo and Super-Resolution Images. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference* , 2, 2316-2323, 2006.

[39] H. Kim, C. Lin, J. Song and H. Chae, Distance Measurement Using a Single Camera with a Rotating Mirror. In *International Journal Of Control Automation And Systems,* 3(4), 542, 2005.

[40] W. Teoh and X. Zhang, An Inexpensive Stereoscopic Vision System For Robots. In *IEEE International Conference on Robotics and Automation,* 1, 186-189, 1984.

[41] R.J. Francois, G. Medioni and W. Roman, Reconstructing Mirror Symmetric Scenes From A Single View Using 2-View Stereo Geometry. In *Pattern Recognition, 2002,* 4, 12-16, 2002.

[42] A. Goshtasby and W. A. Gruver, Design Of A Single-Lens Stereo Camera System. In *Pattern Recognition,* 26(6), 923-937, 1993.

[43] M. Inaba, T. Hara and H. Inoue, A Stereo Viewer Based On A Single Camera With View-Control Mechanisms. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems,* 3, 1857-1865, 1993.

[44] S. Nene and S. Nayar, Stereo With Mirrors, in *Computer Vision, Sixth International Conference,* 1087-1094, 1998.

[45] J. Segen and S. Kumar, Shadow Gestures: 3D Hand Pose Estimation Using A Single Camera. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* 1, 1999.

[46] D. Moore and M. Hayes, Tracking 3d Position And Orientation From 2D Sequences Using Simple Geometry. In *Conf. Record on Asilomar on Signals, Systems & Computers,* 1, 125-129, 1997.

[47] R. LeGrand and R. Luo, Position Estimation Of Selected Targets. In *IEEE International Conference on Robotics and Automation,* 2, 1714-1719, 1996.

[48] E. Adelson and J. Wang, Single Lens Stereo With A Plenoptic Camera. In *IEEE Trans Pattern Anal Mach Intell,* 14(2), 99-106, 1992.

[49] J. Cardillo, M. Sid Ahmed and J. Soltis, 3d Position Sensing Using A Single Camera Approach. In *Proceedings of the 32nd Midwest Symposium Circuits and Systems,* 1, 325-328, 1989.

[50] Y. Suzuki and M. Takano, Estimation Of The Position Of A Rigid Body With A Single Camera. In *Proc. Int. Conf. on Industrial Electronics, Control and Instrumentation* , 2, 1309-1331, 1991.

[51] G. Lester, J. Watts, and D. Wilmington, Ferroelectric Liquid Crystal Device for a Single Camera Stereoscopic Endoscope System. In *Electronics Letters,* 33(10), 857-858, 1997.

[52] G. Lester, J. Watts, and D. Wilmington, Single Camera Three-dimensional Endoscope System Using a Ferroelectric Liquid Crystal Device. In *IEE Proceedings of Science, Measurement and Technology,* 145(2), 49-51, 1998.

[53] D. Lee and I. Kweon, A Novel Stereo Camera System By A Biprism. In *IEEE Transactions on Robotics and Automation,* 16(5), 528-541, 2000.

[54] K. Genovese, L. Casaletto, J. Rayas, V. Flores, and A. Martinez, Stereo-Digital Image Correlation (DIC) measurements with a single camera using a biprism. In *Optics and Lasers in Engineering,* 51(3), 278-285, 2013.

[55] R.Y. Tsai, A Versatile Camera Calibration Technique For High-Accuracy 2D Machine Vision Metrology Using Off-The-Shelf TV Cameras And Lenses. In *IEEE Journal of Robotics and Automation,* 3(4), 323-344, 1987.

[56] S. Ganapathy, Decomposition of Transformation Matrices for Robot Vision. In *Proc. IEEE International Conference on Robotics and Automation*, 74-79, 1984.

[57] O. D. Faugeras and G. Toscani, Calibration Problem for Stereo. In *Proceedings of International Conference on Computer Vision & Pattern Recognition*, 15-20, 1986.

[58] K. W. Wong, Mathematical Formulation and Digital Analysis in Close Range Photogrammetry. In *Photogrammetric Engineering and Remote Sensing*, 44(11), 1355-1373, 1975.

[59] I. W. Faig, Calibration of Close Range Photogrammetric Systems: Mathematical Formulation. In *Photogrammetric Eng. Remote Sensing*, 41(12), 1479-1486, 1975.

[60] H. Bacakoglu and M. S. Kamel, A Three-Step Camera Calibration Method. In *IEEE Transactions on Instrumentation and Measurement*, 46(5), 1165-1172, 1997.

[61] H. Gao, C. Wu, L. Gao and B. Li, An Improved Two-Stage Camera Calibration Method. In *Proceedings of the 6th World Congress on Intelligent Control and Automation*, 9514-9518, 2006.

[62] J. Weng, P. Cohen, and M. Herniou, Calibration of Stereo Cameras Using a Non-Linear Distortion Model. In *Proceedings of the 10th International Conference on Pattern Recognition*, Atlantic City, NJ, 246-253, 1990.

[63] G. Q. Wei and S. Ma, A Complete Two-plane Camera Calibration Method and Experimental Comparisons. In *Proceedings of the Fourth International Conference on Computer Vision*, Berlin, 439-446, 1993.

[64] Q. T. Luong and O. D. Faugeras, Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices. In *International Journal of Computer Vision*, 22(3), 261-289, 1997.

[65] Olivier Faugeras, *Three-Dimensional Computer Vision:A Geometric Viewpoint*, MIT Press, 1993.

[66] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, 2000.

[67] Milan Sonka, Vaclav Hlavac and Roger Boyle, *Image Processing, Analysis and Machine Vision*, Thomson Asia, 2002.

[68] Z. Zhang, A Flexible New Technique for Camera Calibration. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330-1334, 2000.

[69] H. Kwon, J. Park and A.C. Kak, A New Approach for Active Stereo Camera Calibration. In *IEEE International Conference on Robotics and Automation* , 10-14, 2007.

[70] M. Brown and D. G. Lowe, Invariant Features from Interest Point Groups. In *Proceedings of British Machine Vision Conference*, Cardiff, Wales, 656-665, 2002.

[71] R. Hartley, R. Gupta, and T. Chang, Stereo from Uncalibrated Cameras. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Champaign, IL, 761-764, 1992.

[72] Z. Zhang, Determining the Epipolar Geometry and Its Uncertainty: A Review. In *International Journal of Computer Vision*, 27(2), 161-195, 1998.

[73] B. K. P. Horn and B. G. Schunck, Determining Optical Flow. In *Artificial Intelligence in Perspective*, 81-87, 1994.

[74] V. Venkateswar and R. Chellappa, Hierarchical Stereo and Motion Correspondence Using Feature Grouping. In *International Journal of Computer Vision*, 5(3), 245-269, 1995.

[75] S. Todorovic and N. Ahuja, Region-based Hierarchical Image Matching. In *International Journal of Computer Vision*, 78(1), 47-66, 2007.

[76] T. Kanade, H. Kano, and S. Kimura, Development of a Video-rate Stereo Machine. In *Image Understanding Workshop*, Monterey, CA, 549–557, 1994.

[77] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, Introduction to Algorithms, *New York: McGraw-Hill*, 1990.

[78] Y. Ohta and T. Kanade, Stereo by Intra- and Intra-Scanline Search Using Dynamic Programming. In *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7(2), 139-154,1985.

[79] I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs, A Maximum Likelihood Stereo Algorithm. In *Computer Vision and Image Understanding*, 63(3), 542-567, 1996.

[80] S.S. Intille and A.F. Bobick, Incorporating Intensity Edges in the Recovery of Occlusion Regions. In *Proc. International Conference on Pattern Recognition*, 1, 674-677, 1994.

[81] H.H. Baker, Depth from Edge and Intensity Based Stereo, Technical Report *AIM-347*, Artificial Intelligence Laboratory, Stanford Univ., 1982.

[82] T. H. Cormen, et al., Introduction to algorithms. In *MIT press*, 2001.

[83] S. S. Intille and A. F. Bobick, Incorporating Intensity Edges in the Recovery of Occlusion Regions. In *Proceedings of the 12th IAPR International Conference on Pattern Recognition*, Jerusalem, 674-677, 1994.

[84] S. Birchfield and C. Tomasi, Depth Discontinuities by Pixel-to-Pixel Stereo. In *International Journal of Computer Vision*, 35(3), 269-293, 1999.

[85] P. N. Belhumeur, A Bayesian Approach to Binocular Stereopsis. In International Journal of Computer Vision, 19(3), 237-260, 1996.

[86] T. H. Cormen, C. E. Leiserson, and R. L. Rivest, Introduction to Algorithms, *New York: McGraw-Hill*, 1990.

[87] H. Zhao, Global Optimal Surface from Stereo. In *Proc. International Conference on Pattern Recognition*, 1, 101-104, 2000.

[88] I. Thomos, S. Malasiotis, and M.G. Strintzis, Optimized Block Based Disparity Estimation in Stereo Systems Using a Maximum-Flow Approach. In *Proc. SIBGRAPI'98 Conf.*, 1995.

[89] S. Roy and I. J. Cox, A Maximum-Flow Formulation of the N-Camera Stereo Correspondence Problem. In *Proceesings of the Sixth International Conference on Computer Vision*, Bombay, 492-499, 1998.

[90] Y. Boykov, V. Kolmogorov, An Experimental Comparison of Min-Cut/Max-Flow Algorithms for Energy Minimization in Vision. In *Proc. Third International Workshop Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2001.

[91] V. Kolmogorov and R. Zabih, Computing Visual Correspondence with Occlusions Using Graph Cuts. In *Proc. International Conference on Computer Vision*, 2001.

[92] X. Huang, A Cooperative Search Approach to Global Optimization. In *Proceedings of the First International Conference on Optimization Methods and Software*, China, December 2002.

[93] X. Huang, Cooperative Optimization for Solving Large Scale Combinatorial Problems. In *Theory and Algorithms for Cooperative Systems*, ser. Series on Computers and Operations Research. World Scientific, 117–156, 2004.

[94] X. Huang, A General Framework for Constructing Cooperative Global Optimization Algorithms. In *Frontiers in Global Optimization*, ser. Nonconvex Optimization and Its Applications. Kluwer Academic Publishers, 179–221, 2003.

[95] X. Huang, A General Global Optimization Algorithm for Energy Minimization from Stereo Matching. In *ACCV*, Korea, 480–485, 2004.

[96] D. Marr and T. Poggio, Cooperative Computation of Stereo Disparity. In *Science (New York, NY)*, 194(4262), 283, 1976.

[97] Y. Zhang, C. Kambhamettu , Stereo Matching with Segmentation-based Cooperation. In *European Conference on Computer Vision*, 556-571, 2003.

[98] H. Yong,L. Kyoung and L. Sang, Robust Stereo Matching Using Adaptive Normalized Cross Correlation. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (99), 2011.

[99] R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov and A. Agarwala, A Comparative Study of Energy Minimization. In *Microsoft Research,* 16-29, 2006.

[100] Y. Ruichek, Multilevel and Neural-network-based Stereo Matching Method for Real-time Obstacle Detection Using Linear Cameras. In *IEEE Trans on Intelligent Transportation Systems*, 6(1), 54-62, 2005.

[101] X. J. Hua, M. Yokomichi, M. Kono, Stereo Correspondence Using Color Based on Competitive-cooperative Neural Network. In *Proc of the 6th International Conference on Parallel and Distributed Computing*, Applications and Technologies, Denver, 856-860, 2005.

[102] J. Shah, A Nonlinear Diffusion Model for Discontinuous Disparity and Half-Occlusion in Stereo. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 34-40, 1993.

[103] J. Sun, N. N. Zheng, and H. Y. Shum, Stereo Matching using Belief Propagation. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(7), 787-800, 2003.

[104] R. Balasubramanian, D. Sukhendu , S. Udayabaskaran, and K. Swaminathan, Quantization error in stereo imaging systems. In *International Journal of Computer Mathematics*, 79 (6), 671-691. ISSN 0020-7160, 2002.

[105] C. Chang, C.; Chatterjee, S., Quantization error analysis in stereo vision. In *Signals, Systems and Computers, 1992 Conference Record of The Twenty-Sixth Asilomar Conference*, 2, 1037-1041, Oct 1992, doi: 10.1109/ACSSC.1992.269140 .

[106] S. D. Blostein and T. S. Huang, Error analysis in stereo determination of 3-D point positions. In *IEEE Trans. On Patt. Anal. And Mach. Intel, PAMI-9:752-765*, Dec. 1987.

[107] E. Krotkov, R. Kories, and K. Henriksen, Stereo ranging with verging cameras: a practical calibration procedure and error analysis. In Technical Report MS-CIS-86-36, Dept. of Computer and Information Science, University of Pennsylvania, Dec. 1986.

[108] E.S. McVey and J.W. Lee, Some accuracy and resolution aspects of computer vision distance measurements. In *IEEE Trans. on Patt. Anal. & Mach. Intel., PAMI-*(6):646-649, Nov. 1982.

[109] J.J. Rodriguez and J.K. Aggarwal, Stochastic analysis of steno quantization error. In *IEEE Trans. on Patt. Anal. and Mach. Iniel*, PAMI-12, (5):467-470, May 1990.

[110] A. Verri and V. Torre, Absolute depth estimates in stereopsis. In *Journal of Optical Society of America*, 3(3), 297–299, 1986.

[111] L. Matthies and S.A. Shafer, Error modeling in stereo navigation. In *IEEE Journal of Robatics and Automation*, 3, 239–248, 1987.

[112] Gallup, D.; Frahm, J.-M.; Mordohai, P.; Pollefeys, M., Variable baseline/resolution stereo. In *Computer Vision and Pattern Recognition, 2008. CVPR 200*, 23-28 June 2008, doi: 10.1109/CVPR.2008.4587671.

[113] K. Mikko, N. Mikko, and O. Pirkko, Method for measuring stereo camera depth accuracy based on stereoscopic vision. In *Proc. SPIE 7864, Three-Dimensional Imaging,Interaction,and Measurement*, 78640I, January 2011,doi:10.1117/12.872015.

[114] J. Pujol, The solution of non-linear inverse problems and the Levenberg–Marquardt method. In *Geophysics(SEG)*, 72(4), doi:10.1190/ 1.2732552.

[115] D. Scharstein and R. Szeliski, A Taxonomy and Evaluation of Dense Two-frame Stereo Correspondence Algorithms. In *Int. Jour. Computer Vision*, 47(1/2/3): 7-42, 2002.

[116] Y. Lamdan and H. Wolfson, Geometric Hashing: A General And Efficient Model-based Recognition Scheme. In *Proceedings of the Second International Conference on Computer Vision*, 1988, 238-249.

[117] L. F. Cheong. 3D Computer Vision [Lecture Note]. Available: http://couses.nus.edu.sg/course/eleclf/ee6901

[118] E. Boyer and M. O. Berger, 3D Surface Reconstruction Using Occluding Contours. In *International Journal of Computer Vision*, 22(3), 219-233, 1997.

[119] P. Aschwanden and W. Guggenbuhl, Experimental Results from a Comparative Study on Correlation-Type Registration Algorithms. In *Robust Computer Vision*, 1992, 268-289.

[120] J. Banks and P. Corke, Quantitative Evaluation of Matching Methods and Validity Measures for Stereo Vision. In *The International Journal of Robotics Research*, 20(7), 512-532, 2001.

[121] W. L. Kee, K. B. Lim, and D. L. Wang, Virtual epipolar line construction of single-lens bi-prism stereovision system. In *Journal of Electronic Science and Technology*, 10(2), June 2012.

[122] K. B. Lim, W. L. Kee, and D. L. Wang, Virtual camera calibration and stereo correspondence of single-lens bi-prism stereovision system using geometrical approach. In *Signal Processing: Image Communication*, 28(9), 1059-1071, October 2013.

[123] W. L. Kee, K. B. Lim, Z. L. Tun, and Y. D. Bai, New Understanding on the effect of angle and position of bi-prism on single-lens bi-prism based stereovision system. In *Journal of Electronic Imaging*, 23(3), May 2014.

Appendices

A. Law of Refraction (Snell's Law)

Snell's law (also known as the Snell–Descartes law and the law of refraction) is a formula used to describe the relationship between the angles of incidence and refraction, when referring to light or other waves passing through a boundary between two different isotropic media, such as water and glass. Snell's law states that:

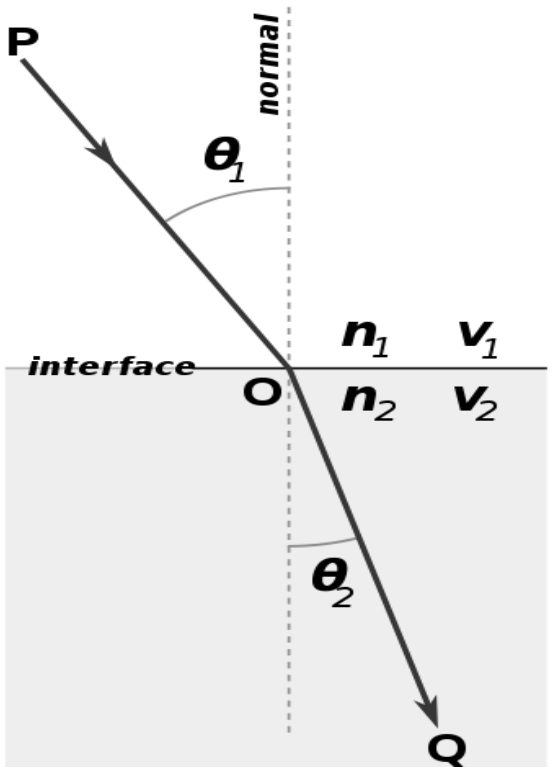


Figure A1 Demonstration of the Snell's Law

1) The incident ray *P*, refracted ray *Q* and normal of the boundary is coplanar;
2) The ratio of the sines of the angles of incidence and refraction is equivalent to the ratio of phase velocities in the two media, or equivalent to the reciprocal of the ratio of the indices of refraction:

$$\frac{\sin\theta_1}{\sin\theta_2} = \frac{v_1}{v_2} = \frac{n_2}{n_1}$$

with each θ as the angle measured from the normal of the boundary, v as the velocity of light in the respective medium (SI units are meters per second, or m/s) and n as the refractive index (which is unit less) of the respective medium.

B. **Zhang's calibration algorithm**

Without the loss of generality, we assume the object plane captured is on Z=0 of the world coordinates frame. The projection equation is reduced to:

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = A[r_1 \quad r_2 \quad t\ ]\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}$$

$$s\tilde{m} = H\tilde{M}$$

Where $[r_1 \quad r_2 \quad t]$ represent the column vector of the rotation and translation matrix, homographies, $H = A[r_1 \quad r_2 \quad t\ ]$, is 3x3 matrix up to a scale factor.

$$H = [h_1 \quad h_2 \quad h_3] = A[r_1 \quad r_2 \quad t\ ]$$

Since r_1 and r_2 are orthonormal, we can impose two constraints:

$$h_1^T A^{-T} A^{-1} h_2 = 0$$

$$h_1^T A^{-T} A^{-1} h_1 = h_2^T A^{-T} A^{-1} h_2$$

Because a homography has 8 degrees of freedom and there are 6 extrinsic parameters (3 for rotation and 3 for translation), we can only obtain 2 constraints on the intrinsic parameters. For the five unknown parameters in *A*, three images of the model plane are required to solve for all the intrinsic parameters. Once the intrinsic parameters are obtained, the extrinsic parameters are readily computed:

$$r_1 = \frac{1}{\lambda} A^{-1} h_1 ; r_2 = \frac{1}{\lambda} A^{-1} h_2 ; r_3 = r_1 \times r_2 ; t = \frac{1}{\lambda} A^{-1} h_3$$

C. Mid-point theorem

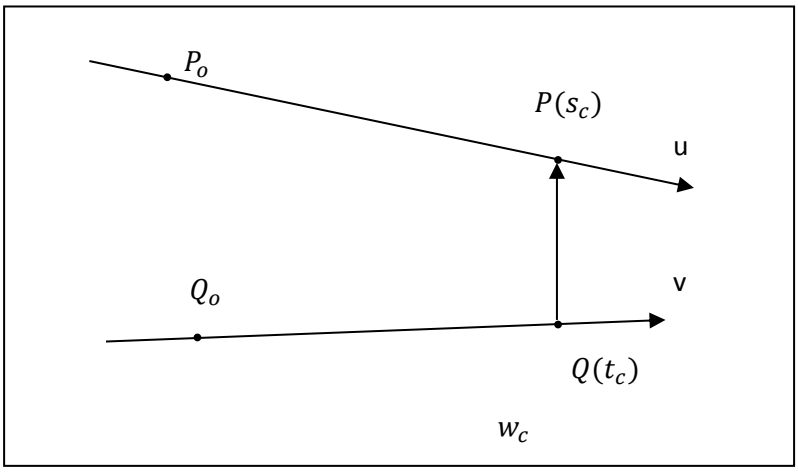


Figure C1 Mid-point of two skew lines

If $\vec{P}$ and $\vec{Q}$ are two skew lines in 3-D space which are not parallel and w_c is uniquely perpendicular to $\vec{u}$ and $\vec{v}$, then

$$u.w_c = 0 \; and \; v.w_c = 0$$

$$w_c = P(s_c) - Q(t_c)$$

$$= w_o + s_c u - t_c v$$

Where $w_o = P_o - Q_o$.

$$u.(w_o + s_c u - t_c v) = 0$$

$$v.(w_o + s_c u - t_c v) = 0$$

Let $a = u.u, b = u.v, c = v.v, d = u.w_o, e = v.w_o$, we can solve for s_c and t_c as shown below:

$$s_c = \frac{be - cd}{ac - b^2}, t_c = \frac{ae - bd}{ac - b^2}$$

After solving for s_c and t_c, $P(s_c)$ *and* $Q(s_c)$ can be computed, the mid-point of the shortest distance is

$$midpoint = \frac{P(s_c) + Q(s_c)}{2}$$

D. Convergent System

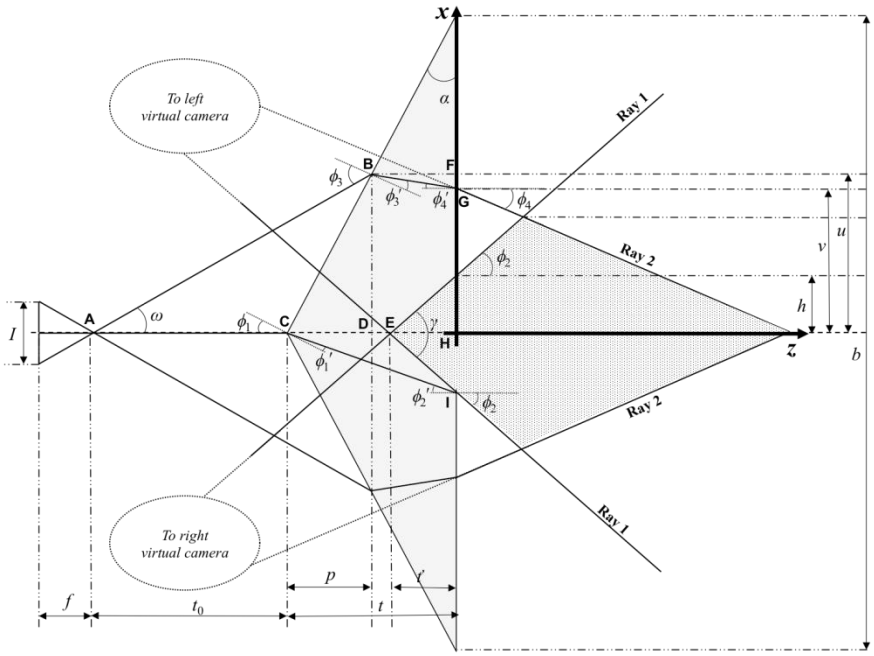


Figure D1 Detailed geometry of a convergent system

The linear equations of rays 1 and 2 for the convergent case are the same as in the divergent case.

That is, equation of ray 1 is given by

$$x = z \tan \phi_2 + h,$$

where $\phi_2 = \sin^{-1}\left\{n_r \sin\left[\alpha - \sin^{-1}\left(\frac{\sin\alpha}{n_r}\right)\right]\right\}$ and $h = \frac{b}{2}\tan\alpha \tan\left[\alpha - \sin^{-1}\left(\frac{\sin\alpha}{n_r}\right)\right]$.

On the other hand, ray 2 is defined by

$$x = z \tan \phi_4 + v,$$

where $$\phi_4 = \sin^{-1}\left\{n_r \sin\left[\sin^{-1}\left(\frac{\sin\left(\tan^{-1}\frac{I}{2f}+\alpha\right)}{n_r}\right) - \alpha\right]\right\}$$

and $$v = \frac{t_0 I}{2f - I\tan\alpha} + \left[\frac{2ft-(t+t_0)I\tan\alpha}{2f-I\tan\alpha}\right]\tan\left[\sin^{-1}\left[\frac{\sin(\omega+\alpha)}{n_r}\right] - \alpha\right].$$

The only difference between convergent and divergent FOV is that the slope of ray 2 in the former case has a negative gradient, that is, ϕ_4 is negative.

E. Results of Set-up 3 and 4

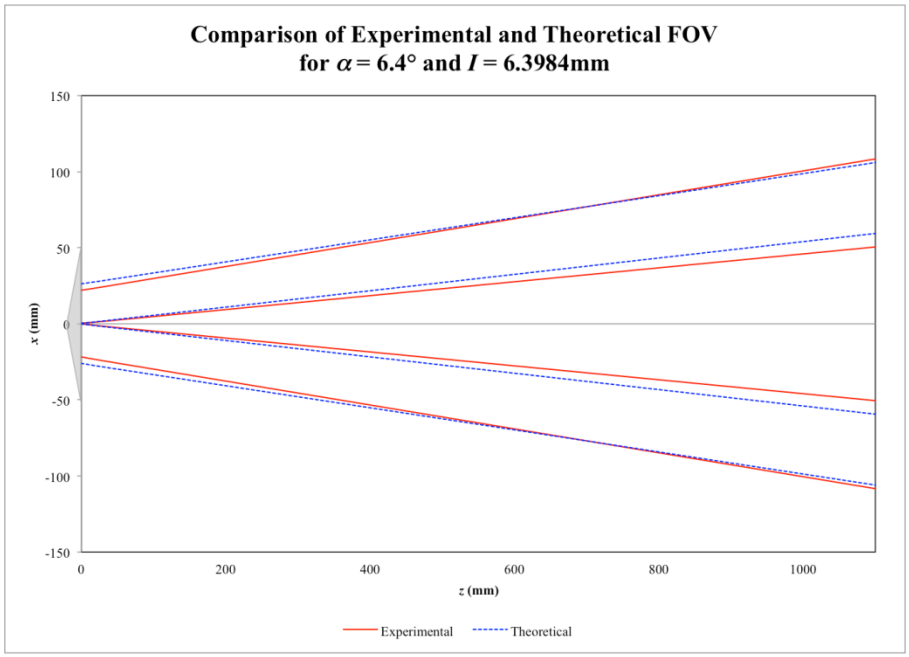


Figure E1 Comparison of experimental and theoretical FOV for set-up 3

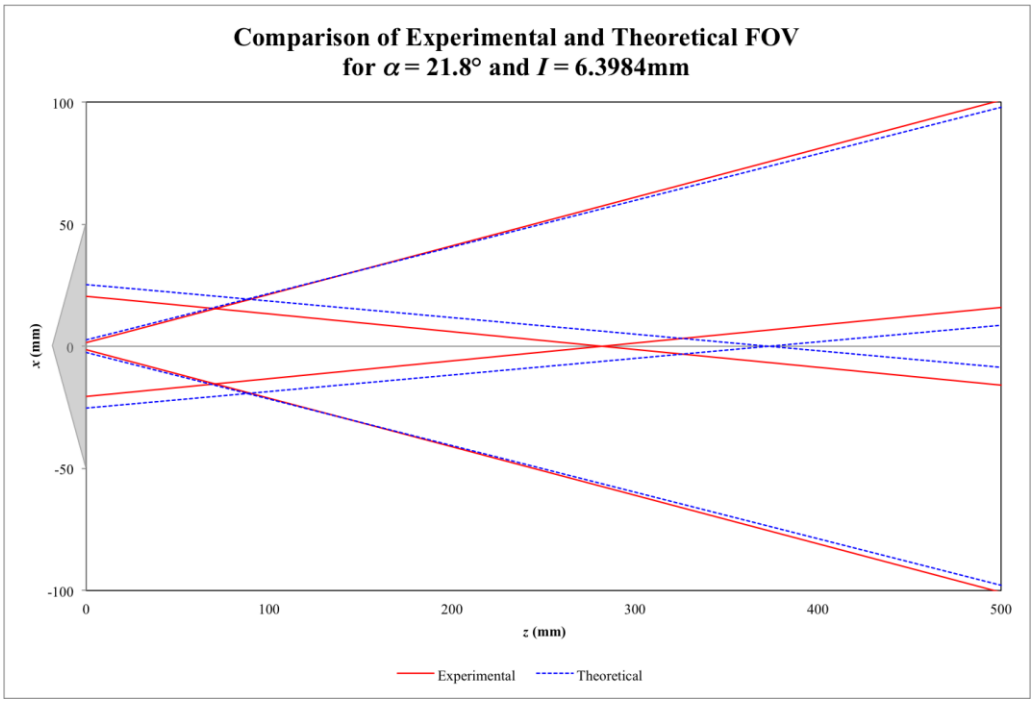


Figure E2 Comparison of experimental and theoretical FOV for set-up 4