

**STEREO CORRESPONDENCE AND DEPTH RECOVERY
OF SINGLE-LENS BI-PRISM BASED STEREOVISION SYSTEM**

ZHAO MEIJUN

(B.Eng., Harbin Institute of Technology, Harbin, China;
M.Sc., Shanghai Academy of Spaceflight Technology, Shanghai, China)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

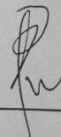
2013

Declaration

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.



Zhao Meijun

10 May, 2013

Acknowledgements

Acknowledgements

I would like to express my sincere appreciation to my supervisor, Associate Professor Lim Kah Bin, for his invaluable guidance, insightful comments, strong encouragements and personal concerns both academically and otherwise throughout the course of the research. I benefit a lot from his comments and critiques. I would also like to thank Dr. Xiao Yong, who has given me invaluable suggestions for this research.

I gratefully acknowledge the financial support provided by the National University of Singapore through Research Scholarship that makes it possible for me to study for academic purposes.

Thanks are also given to my friends and technicians in Control and Mechatronics Laboratory for their support and encouragements. They have provided me with helpful comments, great friendship and a warm community during the past few years in NUS.

My deepest thanks go to my families for their moral support and love.

Last but not least, I would like to thank the examiners of this report for their reviewing, attending her oral examination and giving many helpful advices for the future research.

Table of Contents

DECLARATION I

ACKNOWLEDGEMENTS II

TABLE OF CONTENTS III

SUMMARY VI

LIST OF TABLES VIII

LIST OF FIGURES IX

LIST OF SYMBOLS XI

CHAPTER I INTRODUCTION 1

1.1 Stereovision and stereo correspondence 1

1.2 Objective of this thesis 3

1.3 Organisation of the thesis 5

CHAPTER II LITERATURE REVIEW 6

2.1 Overview of the single-lens stereovision systems 6

2.1.1 Conventional two or more camera stereovision system 6

2.1.2 Single camera stereovision system 8

2.2 Review of the stereo correspondence algorithms 18

2.2.1 *Local stereo correspondence methods*..... 19
2.2.2 *Global stereo correspondence methods*..... 22

**CHAPTER III CAMERA CALIBRATION BASED APPROACH FOR STEREO
CORRESPONDENCE AND DEPTH RECOVERY OF SINGLE-LENS BI-
PRISM BASED STEREOVISION SYSTEM 27**

3.1 Real and virtual camera calibration technique 27
3.1.1 *Introduction of the virtual camera model*..... 27
3.1.2 *Calibration of the real camera and the two virtual cameras*..... 29
**3.2 Stereo correspondence of the single-lens bi-prism based stereovision system through camera
calibration 41**
3.3 Depth recovery of single-lens bi-prism stereovision system 43
3.4 Summary 46

**CHAPTER IV RAY SKETCHING BASED APPROACH FOR STEREO
CORRESPONDENCE AND DEPTH RECOVERY OF SINGLE-LENS BI-
PRISM BASED STEREOVISION SYSTEM 48**

4.1 Introduction of epipolar geometry..... 48
4.2 Stereo correspondence by ray sketching based method 51
4.2.1 *Theoretical basis of the novel ray sketching based method*..... 51
4.2.2 *Stereo correspondence by ray sketching based approach*..... 54
4.3 Depth recovery of the single-lens bi-prism based stereovision system 76
4.3.1 *Triangulation of general stereo image pairs* 76
4.3.2 *Triangulation of single-lens bi-prism based stereovision system* 78

4.4 Summary	79
CHAPTER V EXPERIMENT AND EXPERIMENTAL RESULTS	80
5.1 Setup of the single-lens prism based stereovision system	80
5.2 Experimental results by camera calibration based approach	82
5.2.1 <i>Experimental procedures of the camera calibration based approach</i>	82
5.2.2 <i>Results of stereo correspondence by camera calibration based approach</i>	84
5.2.3 <i>Results of the depth recovery by camera calibration based approach</i>	88
5.3 Experimental results by ray sketching based approach	92
5.3.1 <i>Results of stereo correspondence by ray sketching based approach</i>	93
5.3.2 <i>Results of depth recovery by ray sketching based approach</i>	96
5.4 Evaluation and discussion of the experimental results	99
5.4.1 <i>Evaluation and discussion on the camera calibration based approach</i>	100
5.4.2 <i>Evaluation and discussion on ray sketching based method</i>	104
5.4.3 <i>Summary</i>	107
CHAPTER VI CONCLUSIONS	108
BIBLIOGRAPHY	113
APPENDICES	123
Appendix A- the Snell's Law and 3D geometrical analysis	123
Appendix B- experimental results	126
PUBLICATIONS	138

Summary

Stereovision refers to the problem of determining the three-dimensional structure of a scene from two or more digital images taken from distinct viewpoints. The basis of stereovision is that a single three-dimensional physical scene is projected to a unique pair of images in two observing cameras. However, the reconstruction of the same 3D scene is only possible when one is able to locate the two points from the image pairs which correspond to the same point in the scene. This is known as the stereo correspondence, which poses the greatest challenge in stereovision. The solution of this problem is necessary in the depth recovery of the 3D scene in question.

In this thesis, the 2D images pairs are captured simultaneously by the single-lens binocular stereovision system using a bi-prism (2F filter). This system offers several advantages over that which uses two cameras, such as compactness, lower costs and ease in operation. The image of the 3D scene is split by the prism into two different sub-images, which are regarded as an image pair acquired by two virtual cameras. The concept and formation of the virtual cameras are also introduced. Two approaches are developed for the stereo correspondence and 3D scene recovery: camera calibration and ray sketching approaches. In addition, we assume that the camera lens is distortion-free.

The first approach yields the relationship between a point in the 2D digital image and its corresponding 3D world point, given by a linear 3 by 4 projection matrix. However, the results are highly dependent on the calibration accuracy. The ray sketching approach requires no complex calibration process. It is based on the geometrical characteristics and the optical principles of the system to solve the stereo correspondence. This novel approach has not been attempted before by the researchers.

A specially designed experimental setup, with high precision was fabricated to conduct the experiments. The results show that both approaches are effective and robust. The depth recovery accuracy is of the order of 3% to 4% depending on the value of the target depth. The experiments are carried out with a maximum target depth of 1800mm. Future works can explore the effectiveness of our works to recover depth with a longer range. To improve the accuracy, future development in both the approaches should also consider the effect of lens distortion. Inaccuracy due to the experimental setup, such as the mis-positioning and mis-alignment of the prism and the camera, should also be investigated.

List of Tables

TABLE 5.1: SPECIFICATION OF THE JAI CV-M9CL CAMERA.....	81
TABLE 5.2: RESULTS OF STEREO CORRESPONDENCE AT POSITION A	87
TABLE 5.3: RESULTS OF DEPTH RECOVERY BY CALIBRATION BASED APPROACH AT POSITION A.....	90
TABLE 5.4: RESULT OF STEREO CORRESPONDENCE BY RAY SKETCHING BASED APPROACH (POSITION A).....	94
TABLE 5.5: THE RECOVERED DEPTH OF POINTS BY RAY SKETCHING BASED APPROACH (POSITION A).....	97
TABLE B.1: RESULTS OF STEREO CORRESPONDENCE OF RANDOM 20 POINTS AT DISTANCE 1400MM.....	128
TABLE B.2: RESULTS OF STEREO CORRESPONDENCE OF RANDOM 20 POINTS AT DISTANCE 1800MM.....	130
TABLE B.3: RESULTS OF DEPTH RECOVERY AT DISTANCE OF 1400MM BY CALIBRATION BASED APPROACH.....	131
TABLE B.4: RESULTS OF DEPTH RECOVERY AT DISTANCE OF 1800MM BY CALIBRATION BASED APPROACH.....	132
TABLE B.5: RESULT OF STEREO CORRESPONDENCE BY RAY SKETCHING BASED APPROACH AT 1400MM.....	133
TABLE B.6: RESULT OF STEREO CORRESPONDENCE BY RAY SKETCHING BASED APPROACH AT 1800MM.....	134
TABLE B.7: THE RECOVERED DEPTH OF POINTS AT DISTANCE OF 1400MM BY RAY SKETCHING BASED APPROACH	135
TABLE B.8: THE RECOVERED DEPTH OF POINTS AT DISTANCE OF 1800MM BY RAY SKETCHING BASED APPROACH	136

List of Figures

FIG. 2.1: MODELING OF A TWO CAMERA CANONICAL STEREOVISION SYSTEM	7
FIG. 2.2: A CONVENTIONAL STEREOVISION SYSTEM USING TWO CAMERAS.....	8
FIG. 2.3: SINGLE CAMERA STEREOVISION SYSTEM WITH MIRRORS/PLATES	12
FIG. 2.4 SINGLE CAMERA STEREOVISION USING TWO PLANAR MIRRORS	13
FIG. 2.5: FOUR STEREOVISION SYSTEM SETUP USING MIRRORS: (1) TWO PLANAR MIRRORS; (2) TWO ELLIPSOIDAL MIRRORS; (3) TWO HYPERBOLOIDAL MIRRORS; (4) TWO PARABOLOIDAL MIRRORS	13
FIG. 2.6: ILLUSTRATION OF LEE AND KWEON’S BI-PRISM STEREOVISION SYSTEM	15
FIG. 2.7: DIAGRAM OF STEREO CORRESPONDENCE SOLVED BY LEE AND KWEON	16
FIG. 2.10: DSI DEFINED BY (A) LEFT-RIGHT SCAN-LINE; (B) LEFT SCAN-LINE AND LEFT-DISPARITY	23
FIG. 3.1: SINGLE-LENS BINOCULAR STEREOVISION SYSTEM USING BI-PRISM	27
FIG. 3.2: GENERATION OF THE LEFT VIRTUAL CAMERA USING BI-PRISM (TOP VIEW).....	29
FIG. 3.3: GEOMETRICAL REPRESENTATION OF THE COORDINATES SYSTEM.....	31
FIG. 4.1: ILLUSTRATION OF THE EPIPOLAR GEOMETRY	48
FIG. 4.2: THE ILLUSTRATION OF NON-VERGED GEOMETRY OF STEREOVISION SYSTEM	50
FIG. 4.3: THE GEOMETRY OF VERGED STEREO WITH THE EPIPLAR LINE (SOLID) AND THE COLLINEAR SCAN-LINES (DASHED) AFTER RECTIFICATION	51
FIG. 4.4: RAY SKETCHING BASED STEREO CORRESPONDENCE CONFIGURATION (TOP VIEW)	52
FIG. 4.5: DEMONSTRATION OF RAY SKETCHING BASED APPROACH (ISOMETRIC VIEW)	53
FIG. 4.6: PARAMETERS OF THE BI-PRISM SINGLE-LENS STEREOVISION SYSTEM (ISOMETRIC VIEW).....	54
FIG. 4.7: DERIVATION OF THE INTERSECTION POINT E (TOP VIEW)	57
FIG. 4.8: ILLUSTRATION OF RAY 3 DERIVATION (TOP VIEW)	60
FIG. 4.9: LOCAL COORDINATES H ATTACHED AT POINT B RELATED WITH RAY 4 AND RAY 5 (TOP VIEW)	65
FIG. 4.10: LOCAL COORDINATE G ATTACHED AT POINT A RELATED WITH RAY 5 AND RAY 6 (TOP VIEW).....	69
FIG. 4.11: ILLUSTRATION OF THE STEREO CORRESPONDING SEARCH	75
FIG. 4.12: TRIANGULATION WITH NONINTERSECTING.....	77
FIG. 4.13: OBJECT POINT DETERMINATION WHEN RAY 3 AND RAY 4 ARE NOT INTERSECTED IN SPACE	78
FIG. 5.1: EXPERIMENTAL SETUP OF THE SINGLE-LENS PRISM BASED STEREOVISION SYSTEM.....	80
FIG. 5.2: VERNIER CLIPERS AND ROTATIONAL STAGE	80

List of Figures

FIG. 5.3: OPTICAL BI-PRISM USED FOR OUR EXPERIMENT	81
FIG. 5.4: CUSTOMIZED CALIBRATION BOARD.....	82
FIG. 5.5: THE 7 BY 15 CALIBRATION PATTERN FOR CAMERA CALIBRATION	83
FIG. 5.6: THE SETUP OF THE POSITIONS OF CALIBRATION BOARD IN THE EXPERIMENT.....	83
FIG. 5.7: STEREO IMAGE PAIR TAKEN BY OUR SINGLE-LENS STEREOVISION SYSTEM AT POSITION <i>A</i>	86
FIG. 5.8: DISPLAY OF THE STEREO CORRESPONDING POINTS (POSITION <i>A</i>).....	86
FIG. 5.9: ACTUAL DEPTH AND RECOVERED DEPTH BY CALIBRATION BASED APPROACH (POSITION <i>A</i>)	91
FIG. 5.10: ACTUAL DEPTH AND RECOVERED DEPTH BY CALIBRATION BASED APPROACH (POSITION <i>B</i>)	91
FIG. 5.11: ACTUAL DEPTH AND RECOVERED DEPTH BY CALIBRATION BASED APPROACH (POSITION <i>C</i>).....	92
FIG. 5.12: EPIPOLAR LINE AND STEREO CORRESPONDENCE BY RAY SKETCHING BASED APPROACH (POSITION <i>A</i>)	93
FIG. 5.13: EPIPOLAR LINE AND STEREO CORRESPONDENCE BY RAY SKETCHING BASED APPROACH (POSITION <i>B</i>)....	95
FIG. 5.14: EPIPOLAR LINE AND STEREO CORRESPONDENCE BY RAY SKETCHING BASED APPROACH (POSITION <i>C</i>)	95
FIG. 5.15: ACTUAL AND RECOVERED DEPTH OF POINTS BY RAY SKETCHING BASED APPROACH (POSITION <i>A</i>)	96
FIG. 5.16: ACTUAL AND RECOVERED DEPTH OF POINTS BY RAY SKETCHING BASED APPROACH (POSITION <i>B</i>)	98
FIG. 5.17: ACTUAL AND RECOVERED DEPTH OF POINTS BY RAY SKETCHING BASED APPROACH (POSITION <i>C</i>)	98
FIG. 5.18: DEPTH RECOVERY OF THE 20 POINTS BY TWO APPROACHES AT POSITION <i>A</i>	99
FIG. 5.19: DEPTH RECOVERY OF THE 20 POINTS BY TWO APPROACHES AT POSITION <i>B</i>	99
FIG. 5.20: DEPTH RECOVERY OF THE 20 POINTS BY TWO APPROACHES AT POSITION <i>C</i>	100
FIG. 5.21: ILLUSTRATION OF THE ASSUMPTION OF $Y_{LF} = Y_{RF}$	102
FIG. 5.22: PRISM MISALIGNMENT IN THE STEREOVISION SYSTEM SETUP	106
FIG. A.1: DEMONSTRATION OF THE SNELL'S LAW	123
FIG. B.1: EXPERIMENTAL STEREO IMAGE PAIR TAKEN AT DISTANCE 1400MM.....	126
FIG. B.2: DISPLAY OF THE STEREO CORRESPONDENCES (DISTANCE 1400MM)	127
FIG. B.3: EXPERIMENTAL STEREO IMAGE PAIR TAKEN AT DISTANCE 1800MM.....	129
FIG. B.3: DISPLAY OF THE STEREO CORRESPONDENCES (DISTANCE 1800MM)	131

LIST OF SYMBOLS

λ	: baseline, i.e. the distance between the two camera optical centres
d	:disparity of the corresponding points between the left and right image
O_c	:optical center of the camera
Z_w	:depth of object in world coordinates system
f	:effective real camera focal length
R	:rotational matrix
T	:translational vector
P_w	:object point in world coordinate frame
p_l	:image point on the left image plane
p_r	:image point on the right image plane
(X_w, Y_w, Z_w)	:World Coordinates
$M_{intrinsic}$:camera intrinsic parameters
$M_{extrinsic}$:camera extrinsic parameters
e_l	:epipole of left image
e_r	:epipole of right image
δ	:corner angle of the bi-prism
n_r	:refractive index of the prism glass material

Chapter I Introduction

1.1 Stereovision and stereo correspondence

Human beings have the ability to perceive depth easily through the stereoscopic fusion of the pair of images registered from the eyes, although this visual system is still not well understood. Nevertheless, by modelling the way human beings perceive range information mathematically, the depth of a scene point can be retrieved if the same scene point is viewed in two or more different orientations. Stereovision is a 3D computer vision technique based on this model and comprehensive researches have been devoted to this area in recent decades in search of more unambiguous and quantitative measurements of interested scenario. Its broad application covers surgery navigation [1-3], real-time robotic application [4-5] and object detection and tracking [6-7], etc. According to Barnard and Fischler [8], any stereo analysis can be carried out in six steps: image acquisition, camera modelling, feature extraction, stereo correspondence, depth recovery and interpretation. Among these steps, stereo correspondence is considered to be the most challenging and time-consuming task and depth recovery is the objective to be addressed. Marr [9] depicts 3D vision as follows: *‘Form an image (or a series of images) of a scene, derive an accurate three-dimensional geometric description of the scene and quantitatively determine the properties of the object in the scene’*. This also implies that 3D computer vision consists of stages of data capturing, reconstruction and interpretation.

Stereovision refers to the problem of determining three-dimensional structure of the scene from two or more stereo images taken from distinct viewpoints. Conventional binocular stereovision setups utilize two cameras to capture any pair of images for depth analysis, or three cameras for the case of tri-ocular stereovision. When a point in the scene is projected into different locations in each of the image planes, the difference in

positions of its projections, called disparity, is evaluated. Its depth information is then determined through the knowledge of disparity, geometric relationships between the cameras and the properties of individual cameras. 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 systems. The problems include difficulties in the synchronizing of image capturing, variations in the intrinsic parameters of the hardware used, etc. The solutions of the problems form the motivation of our earlier works in a single-lens prism based stereovision system as well as the concept of virtual camera in the year 2004 by Lim and Xiao [10]. 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 analyzing the system error.

The main objective of stereovision is to recover the depth and to reconstruct the 3D scene through captured image pairs. However, given a two-dimensional view of a 3D scene, there is no unique way to reconstruct it. This is indeed an ill-posed problem, and there is no unique or definitive solution even if the 2D image pairs are perfectly captured. This problem could only be solved if we are able to determine the information of all corresponding objects in both the captured images. The process is referred to as stereo correspondence. It is indeed the most essential and probably the most challenging step in stereovision. There are many algorithms which have been developed in addressing this

issue, especially for the determination of disparity map. However, the performances of these algorithms are adversely affected by the presence of random occlusion, repeated patterns, image noise, poor illumination and high computational load, etc. There are also methods designed to alleviate some of these difficulties and to reduce the spurious matches. Notably, Grewe and Kak [11] reported on the existence of the epipolar geometry, which is inherent in stereoscopic geometry. Their work enables us to clarify what information is needed in order to perform the search for corresponding elements only along epipolar lines. It has greatly simplified the stereo correspondence search.

1.2 Objective of this thesis

The aim of this research reported in this thesis is to develop the faithful and efficient methods to solve the stereo correspondence and hence to obtain the depth map of the scene for a bi-prism based single-lens stereovision system developed in our laboratory.

Two methods are proposed and presented in the thesis.

(1) Camera calibration based method

This method will achieve stereo correspondence and depth recovery by calibrating the real camera and the two associated virtual cameras. The basic of this approach is to mathematically express the relationship between the 3D world coordinates where a scene locates and the corresponding 2D image coordinates where the digital images are observed. In this thesis, a three-step linear calibration technique is developed based on the works of Tsai [12] and Zhang [13-14]. The linear 3 by 4 perspective matrices, representing the relationship between the 3D scene and 2D digital images are generated for both real and virtual cameras in Homogenous Coordinates. The perspective matrices contain all the intrinsic and extrinsic parameters of the cameras and enable the stereo

correspondence between image points to be established. The depth information is then determined through the matching results and the intrinsic and extrinsic camera properties.

(2) Ray sketching based method

This method employs 3D geometrical analyses and simple optical principle: Snell's Law to attain the objectives of stereo correspondence and depth recovery. Unlike the calibration based approach, this method constructs the epipolar geometry for the stereo images. With a known image point in one of the captured images, the candidates of the corresponding points in the remaining image(s) can be determined. Subsequently, the corresponding epipolar line can be constructed through the ray sketching approach, which inherently expresses all the pertinent points, lines and planes in the 3D camera coordinates. In this manner, the search for the correspondence points is limited to be along a straight line – epipolar line, instead of the whole image. Whenever a pair of corresponding points is found, the disparity can be computed, and thence the depth recovery can be obtained straightforward through triangulation.

Among the two methods, the first one is less efficient as it involves cumbersome calibration setup and operations, which are not required in the second approach. The ray sketching based method provides an interesting way of understanding the system, and it is simpler to implement. In addition, the accuracy in depth recovery is acceptable.

An experimental setup has been established and several experiments have been carried out to test the effectiveness of the single-lens binocular stereovision systems and to verify the efficiency of the two methods. Results from the experiments demonstrate that the two developed methods are valid in the study of the single-lens bi-prism based stereovision system. Due to the simple under-lying principles used and the characteristics of the system, the methods and the setup can be generalized easily from a binocular to a

multi-ocular system. We believe that most of the works presented in this thesis, especially the ray sketching based method is novel and also useful.

1.3 Organisation of the thesis

Organization of the thesis is as follows. Chapter II provides the background study of the single-lens stereovision system and the stereo correspondence. In Chapter III and IV, two theoretical frameworks, namely, camera calibration based and ray sketching based methods, are proposed to characterize and analyze the stereo correspondence and depth recovery issues of the single-lens bi-prism based stereovision system. The experimental and simulation results are thereafter provided in chapter V, followed by an error analysis. The sources of error are also identified. Finally, the conclusion and suggested future work of this project are addressed.

Chapter II Literature Review

2.1 Overview of the single-lens stereovision systems

By using merely a single static camera with known intrinsic parameters, it is not possible to obtain the three-dimensional location of a point in a 3D scene. 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 carry out the stereovision analyses of a scene comparable to the human vision system, we must ensure that the same scene point can be viewed from two or more different viewpoints. As long as this criterion is satisfied, we can implement stereovision analyses even by using a single camera. Stereovision system in literature can be classified into two broad categories according to the way of stereovision effect generation: 1) conventional two or more camera stereovision system and 2) single camera stereovision system.

2.1.1 Conventional two or more camera stereovision system

The conventional stereovision system employs two or more cameras to capture the images from different viewpoints. Grewe and Kak [11] gave an elaborate overview of the camera modelling and geometry for a binocular stereovision system. They considered the classical stereo camera configuration in Fig. 2.1, which consists of two cameras translated by a baseline distance λ in the x direction. The optical axes of these cameras were parallel and perpendicular to the baseline connecting the image plane centres. Their depth recovery equation is given below:

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

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.

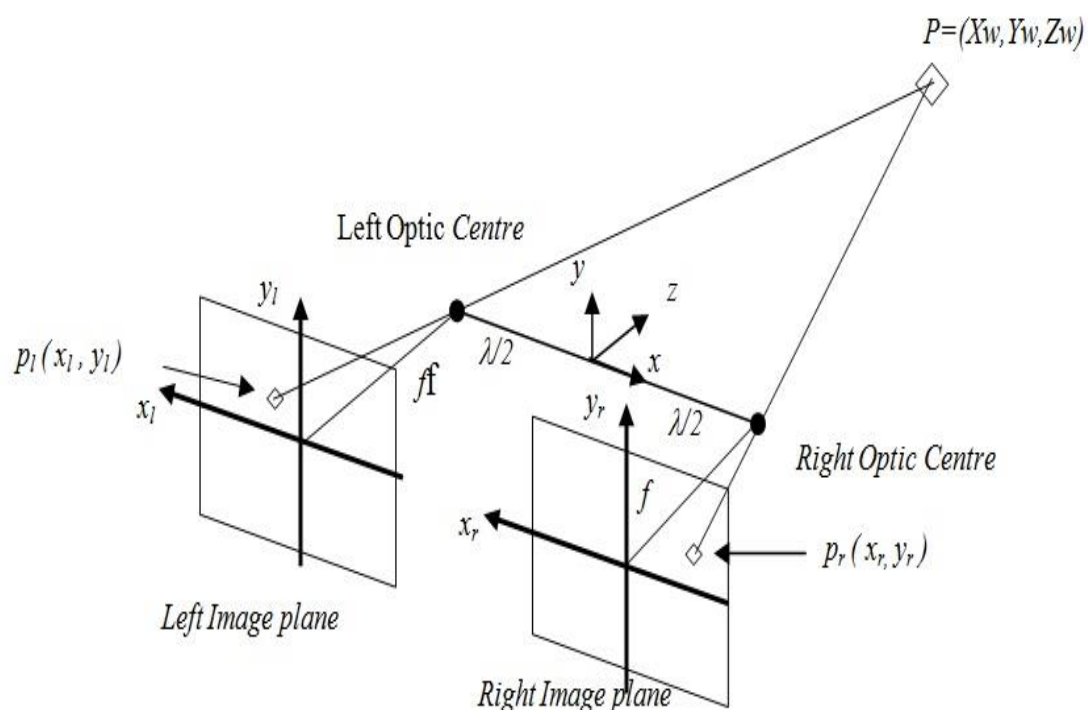


Fig. 2.1: Modeling of a two camera canonical stereovision system

In practice, the cameras are placed at a certain angle to capture the stereo images from different viewpoints as shown in Fig. 2.2. Under this circumstance, the optical axes are no longer parallel and the vanishing point does not exist at infinity.

From the information of the objects/scene in the images (position, disparity, epipolar line, etc.) and the intrinsic and extrinsic parameters of stereovision system, the stereo correspondence and 3D object reconstruction could be obtained.

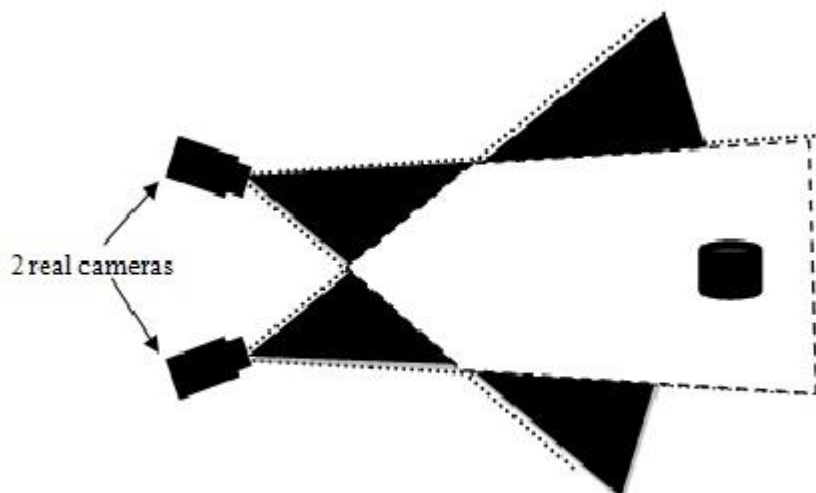


Fig. 2.2: A conventional stereovision system using two cameras

2.1.2 Single camera stereovision system

Although a conventional stereovision system is easy to realize, a great deal of efforts have been made to the single camera stereovision techniques due to the following advantages over the traditional two or multiple camera stereovision system [16]. The immediate benefit of using a single camera is the ease to construct a more compact stereo system at a relatively low cost. Moreover, three main advantages that can be derived from using a single camera over multiple cameras are:

- 1) Identical system parameters: there will be no variation in the intrinsic properties of the camera and lens when the images are captured. This is especially important if the stereo correspondence algorithm is based on the colour information, as it can easily lead to mismatch errors when the images are captured using more than one camera. This advantage will definitely lessen errors when determining the depth of a scene point;
- 2) Synchronized data acquisition: camera synchronization is not an issue any more because only a single camera is used. Stereo data can easily be acquired and

conveniently stored with a standard video recorder without the need to synchronize multiple cameras;

- 3) Ease of calibration: there is only one set of intrinsic calibration parameters of the system, thereby reducing the total number of calibration parameters and hence the computational complexity.

There are many ways to achieve single camera stereovision, depending on what type of depth cue the system employs to capture the depth information of a scene. Based on the different mechanisms, single camera stereovision techniques can be classified into two categories: 1) stereovision by exploiting depth cues from the system and its surroundings, such as shadows and camera motion; and 2) stereovision by exploiting additional optical devices, such as plates, mirrors and optical prisms.

First category: single camera stereovision system using known cues

In this kind of system, one camera is used to capture images of the scene, and in conjunction, with information that can be gathered from other devices or visual cues to recover the depth of the objects in the scene in question.

Shadow can be a possible visual cue. Segan et al. [17] designed a system which used a camera and a point light source to track a user's hand in 3D space. The projections of the hand and its shadow were used as the visual cues to obtain the depth information of the user's hand in space. However, the light source in this system had to be calibrated through a standard procedure, which restricted the applications domain according to the light source setting limitations.

Using the camera motion as the depth cue is another alternative to achieve the depth recovery of 3D information of the single camera stereovision system. LeGrand and Luo

[18] presented an estimation technique that retained the nonlinear camera dynamics and helped to provide an accurate 3D position estimate of the selected targets within the environment. They made use of the known cues, which were the motion of the camera mounted on the robot arm, to acquire 3D information of the work space.

Object geometry was also employed as the depth cue in Moore and Hayes' [19] work to obtain and track the 3D position and orientation of the objects. Three coplanar points on the object need to be identified and their distances from the camera lens are measured. Using photometric techniques and simple geometry, location and orientation of the points in 3D space could be estimated accurately from their projections on the image plane. One drawback was that the method would not work if these points failed to be projected onto the image plane due to occlusion or if these points were highly susceptible to noise. Similar attempt was also reported by Suzuki et al. [20]. Obviously, this type of cue-based methods is not applicable to an unknown object or uncontrolled environment.

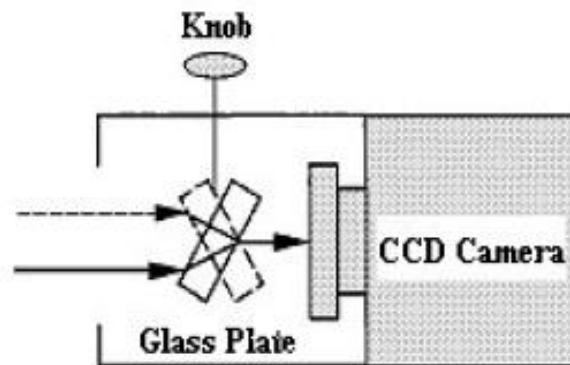
There are also other approaches to achieve single camera stereovision. For example: Adelson and Wang [21] proposed to infer the depth of a scene through the difference in the optical structure when light was striking on adjacent sub regions of the camera aperture. Cardillo et al. [22] introduced a similar method to capture depth information through the investigation of the blurring effects of the camera's lens. This technique worked the best when the image scene constituted sharp contrasting edges. Lester et al. [23-24] developed an unconventional way to achieve single camera stereovision through the application of a ferroelectric liquid crystal (FELC) shutter. The use of the crystal shutter allows the optical path to be switched at video frame rates and offers the advantages of being lightweight, simple to be driven without moving parts. Moreover, the system uses the field sequential display of the two images combined with FELC shutter glasses to present the left and right images to the user's eyes, which allows distance

information being obtained and changes in the relative positions of objects to be evaluated.

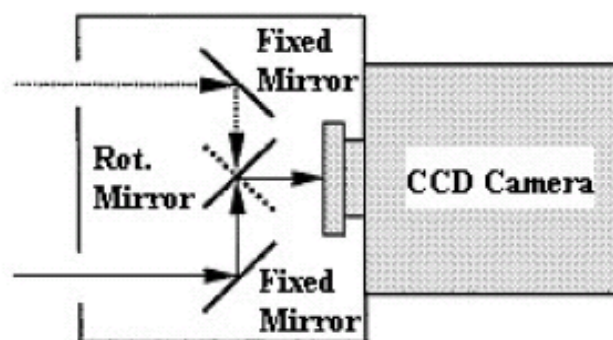
Second category: single camera stereovision system using optical devices

The basis of the technique is to generate the different viewpoints of the object by using the optical devices to change the direction of light path from the objects to the imaging sensor. The use of plane mirrors to create a series of virtual cameras for depth recovery is relatively uncommon in stereovision. The mirrors or plates change the light direction by following the Law of Reflection.

D. Murray [25] proposed an imaging device that consisted of a static camera and rotating plane mirror (inclined to the horizontal) for passive range recovery. Based on Fermat's principle, a reflected scene point that is viewed by a real camera is equivalent to the point being viewed by a virtual camera created by the mirror reflection. As the mirror rotates, the virtual camera moves, thereby generating a panoramic view for stereovision analyses. The device is capable of recovering range in a plane by using only 1D image measurements to track features along the central horizontal raster. It can also recover range over a wide field of view except at two blind spots when the mirror is edge on and when the camera looks at itself. This is remedied in Nishimoto and Shirai's [26] and Teoh and Zhang's [27] work by using different configuration - rotated glass plates and rotated mirrors to capture images. However, the disadvantage that arises is that 1D image scanning is no longer possible.



(a) Rotated glass plate



(b) Rotated mirror

Fig. 2.3: Single camera stereovision system with mirrors/plates

The systems described above require the camera to take two separate shots to obtain one pair of stereo images, their applications are probably limited to static scene or slow changing environment only (even though fast rotation speed of the glass or mirrors reduces the negative effect of this limitation). Gosthasby and Gruver [28] described another mirror-based single camera stereovision system as shown in Fig. 2.4, which can overcome the problem. The acquired images are reflected by the mirrors and transformation processes of these images are needed before carrying out the correspondence and depth measurement as in a normal two camera stereovision system.

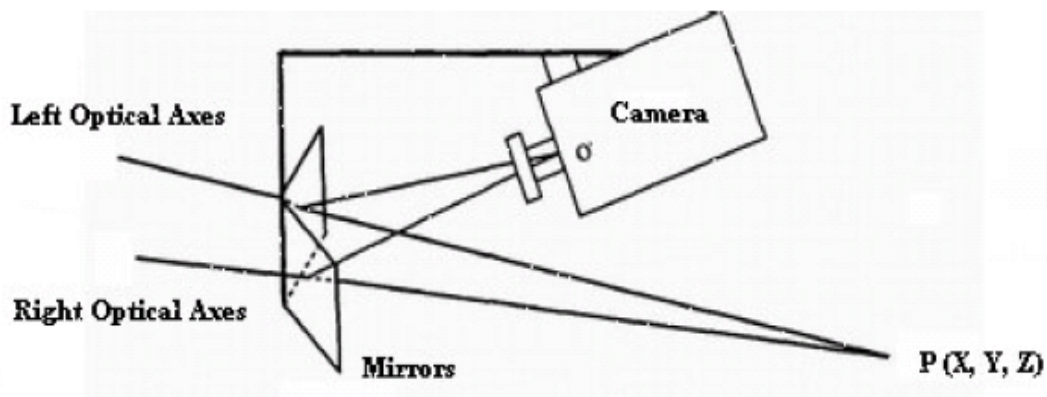


Fig. 2.4 Single camera stereovision using two planar mirrors

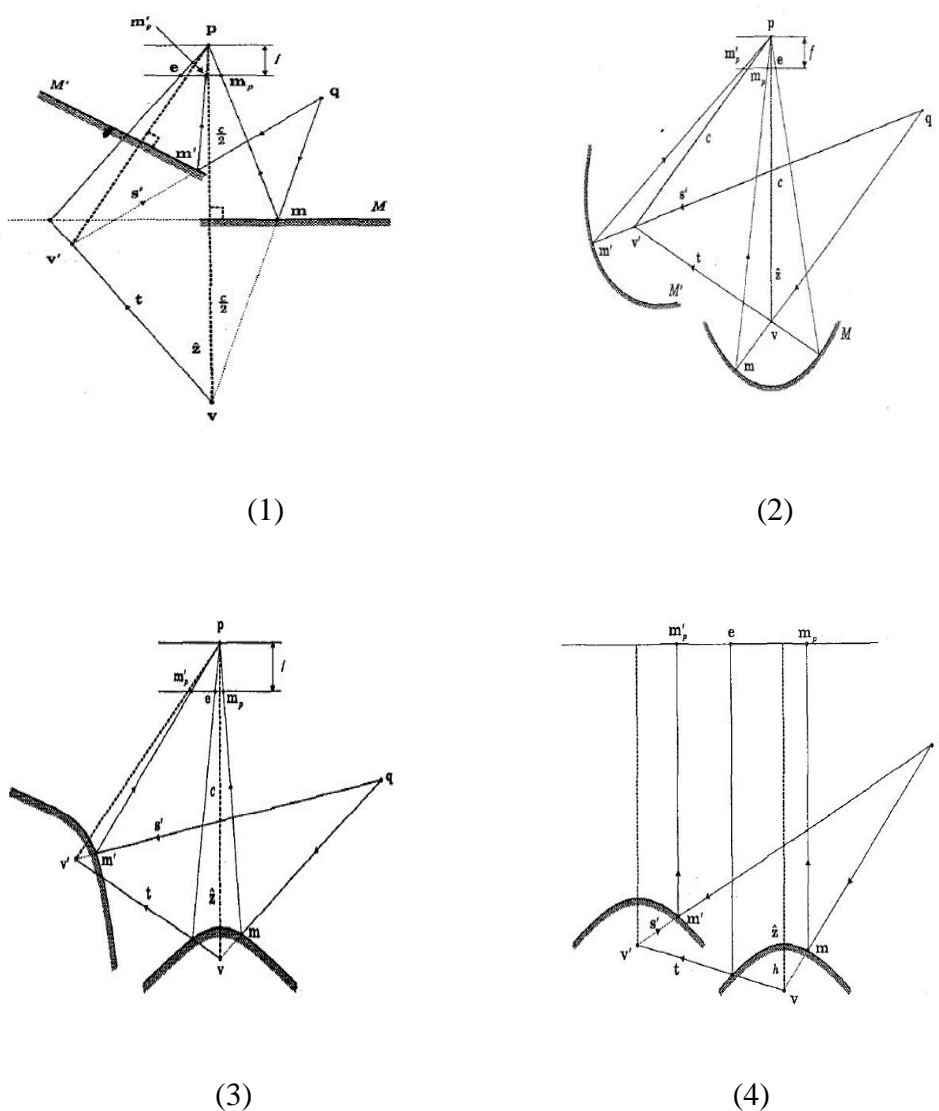


Fig. 2.5: Four stereovision system setup using mirrors: (1) two planar mirrors; (2) two ellipsoidal mirrors; (3) two hyperboloidal mirrors; (4) two paraboloidal mirrors

Taking place of the plane mirrors, S. A. Nene [29] gave an elaborate analysis on stereovision using different types of mirrors as shown in Fig. 2.5. Four stereo systems were proposed using a single camera pointing towards planar, ellipsoidal, hyperbolic and parabolic mirrors, respectively. By using non-planar reflecting surfaces, a wide field of view (FOV) can be achieved. For each scenario, the epipolar constraints were derived and the results of the experiments demonstrated the viability of using these mirrors for stereovision analyses. However, in such systems, the projection of the scene produced by the curved mirrors is not from a single viewpoint. Violation of the “single viewpoint assumption” implies that the pinhole camera model cannot be used, thus making calibration and correspondence a more difficult task.

In the Control and Mechatronics Laboratory of the Department of Mechanical Engineering, National University of Singapore (NUS), continuous effort is being made into the study of the single camera stereovision. A mirror based binocular stereovision system was designed successfully and a preliminary discussion on a bi-prism based binocular single camera stereovision was done by Lim, Lee and Ng [30-31].

Lee and Kweon [32] proposed a single camera stereovision system using one bi-prism which has a similar setup of the binocular system that was presented by Lim and Lee [30]. However, in the approaches used to understand such a system, there are fundamental differences between the methods reported. The following diagram and figure show the details of Lee and Kweon’s system.

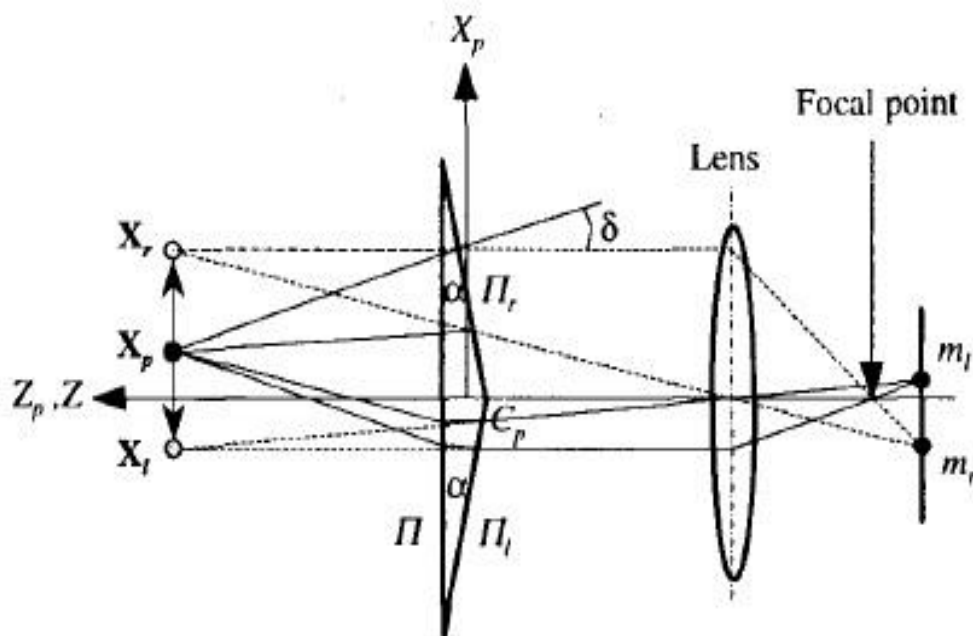


Fig. 2.6: Illustration of Lee and Kweon's bi-prism stereovision system

Lee and Kweon proposed the concept of virtual points in their work. Any arbitrary point in the view zone of the vision system was transformed into two virtual points in 3D space which are determined by the refractive index and the angle of the bi-prism. A simple mathematical model was derived to obtain the stereo correspondence of the system but it works only when the angle between the two image planes is zero. This implied that, an assumption that the two virtual cameras are coplanar was made for their system analyses. This will eventually render the model invalid when the angle of the prism becomes larger.

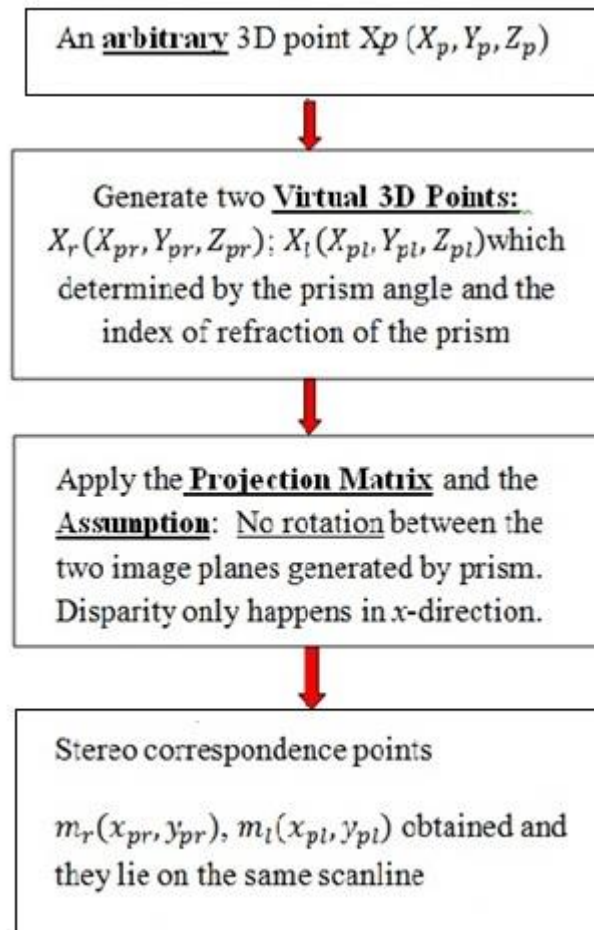


Fig. 2.7: Diagram of stereo correspondence solved by Lee and Kweon

Recently, Lim and Xiao further developed, analyzed, implemented and tested a single camera stereovision system using pyramid-like multi-face optical prism [33-34]. A systematic investigation of their stereovision system, including binocular, tri-ocular and multi-ocular systems, has been carried out. They are the first to analyze the tri-ocular single-lens stereovision system with the ray sketching-based approach in two-dimensional space (Fig. 2.8). The design issues, including virtual camera generation, blurring zone, weak reflections, depth error analyses, and zone of overlap and search range, were also discussed in details [35]. The main constraints encountered during their study are the limited hardware that was available and financial constraints, in particular, the available optical prism did not have the required quality and property. In this thesis, we have

specially fabricated a tailor-made hardware system. We use it to further investigate the stereo correspondence and depth recovery of the single-lens prism based stereovision system. Simple illustration of the system configuration is presented in Fig. 2.9 together with the stereo images captured.

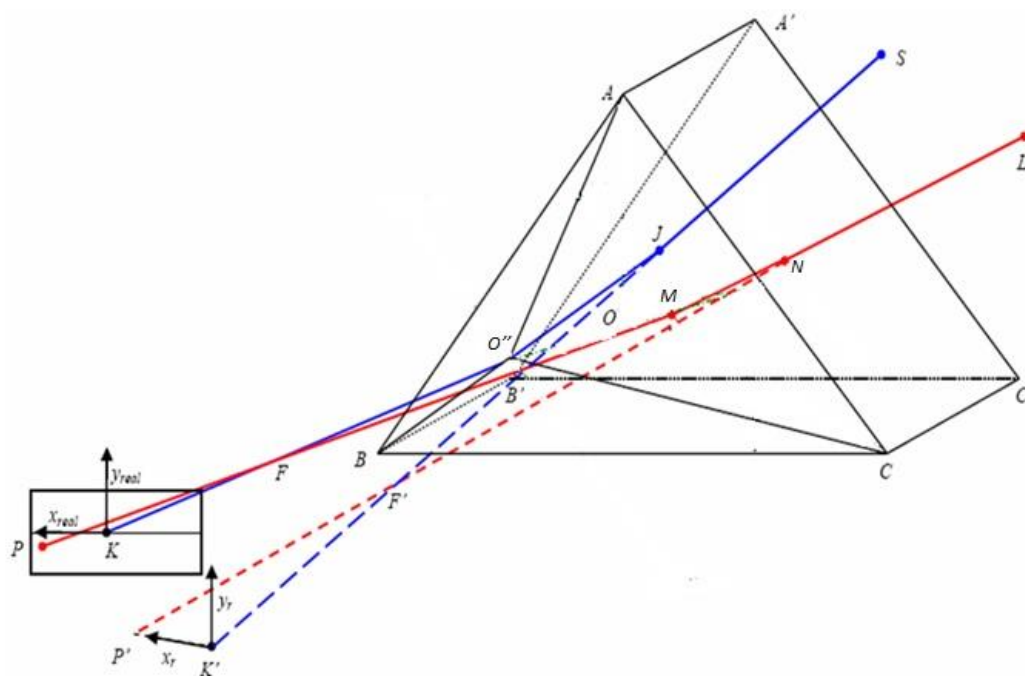


Fig. 2.8: Illustration of virtual camera modelling by using a three-face prism [33-34]

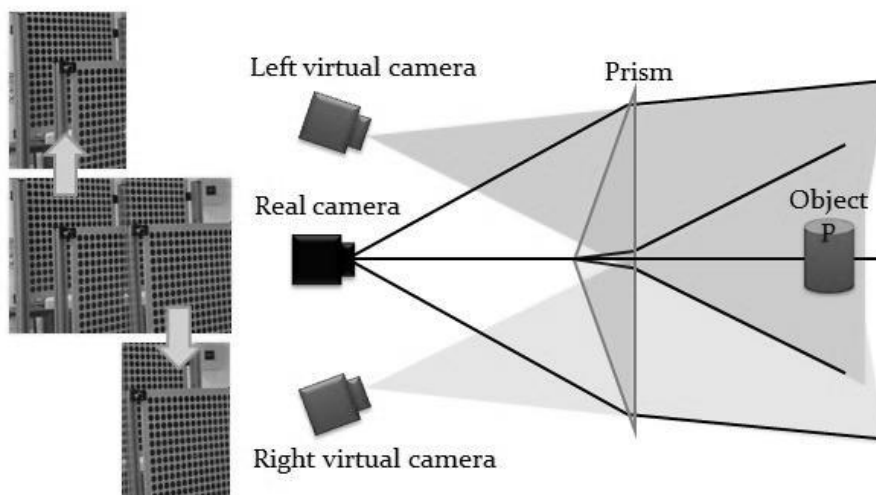


Fig. 2.9: Single-lens bi-prism-based stereovision system

2.2 Review of the stereo correspondence algorithms

Stereo correspondence has traditionally been, and it continues to be, one of the most heavily investigated topics in stereovision. No general solution exists because the performance of the algorithms is often degraded by occlusions, lacks in texture, variation in illumination, etc. A large number of algorithms for stereo correspondence have been developed. In 1989, Dhond and Aggarwal [36] gave a preliminary review of the developed algorithms and they have classified them into six broad categories:

- 1) Computational theory of stereopsis, e.g., Mayhew-Frisby theory of disparity gradient [37];
- 2) Area-based stereo, e.g., Moravec [38];
- 3) Relaxation process in stereo, such as Kim-Aggarwal algorithm [39] and Marr-Poggio cooperative algorithm [40];
- 4) Stereo matching using edge segments. Minimum differential disparity algorithm [41] is one of algorithms that falls in this category;
- 5) Hierarchical approaches to stereo matching. Hierarchical stochastic optimization [42] and concurrent multilevel relaxation [43-44] are two typical algorithms;
- 6) Stereo matching by dynamic programming [45-46].

Another recent survey was done by M. Z Brown et al. [47] in 2003, which classified the stereo corresponding algorithms into two different approaches. One of them is called local approach where the corresponding matching process is locally applied to the pixel of interest and a small number of its surroundings. The other one is called global approach, in which searching process works on the entire image loosely. To reduce the complication and computational complexity, constraints from image geometry, and those making use of object properties and assumptions [48] are commonly exploited to make the problem tractable. Among them, epipolar constraint [49-50], continuity constraint and uniqueness

constraint [51] are the three common and powerful constraints in solving the stereo correspondence.

It is worthwhile to mention that, D. Scharstein and R. Szeliski [52] characterized the performance of the stereo corresponding algorithms by presenting the taxonomy of dense, two-frame stereo methods. Their taxonomy was designed to assess the different components and design decisions made in individual stereo algorithms. Evaluation and comparison of the algorithms were also included to provide a good reference for beginners. Our detailed literature review on stereo correspondence will follow Brown's classification [47].

2.2.1 Local stereo correspondence methods

Local stereo correspondence algorithms are sensitive to the presence of local ambiguous regions in the image, but they are efficient in performance. Gradient approach (optical flow) is a well-known local technique leading to the solutions of local stereo correspondence problem and it is applicable for most of the real-time project. It formulates the differential Eq. (2.1) relating the motion and image brightness to determine the small local disparities between two images. $E(x,y,t)$ is the image intensity at points (x,y) , which is a continuous and differentiable function of space and time. ∇E and E_t are the spatial image intensity derivative and the temporal image intensity derivative, respectively.

$$E(x, y, t) = E(x + dx, y + dy, t + dt),$$

$$\text{Let} \quad \frac{dE}{dt} = 0,$$

$$\frac{dE}{dt} = E_x V_x + E_y v_y + E_t = \nabla E^T (V_x, V_y) + E_t = 0 \quad . \quad (2.1)$$

Among them, ∇E and E_t are known parameters which can be measured directly from the images, while (V_x, V_y) are the unknown optical flow components $(dx/dt, dy/dt)$ in the X and Y directions. Based on the observation, the optical flow has two components while the basic gradient equation for the rate of change of image brightness provides only one constraint. Horn and Schunck [53] introduced the smoothness of the flow as the second constraint to compute the optical flow for a sequence of images. The algorithm is robust and insensitive to the quantization of brightness levels and additive noise. However, it is applicable only when the change in intensity is entirely due to motion and the motion must be “small”.

In contrast, local feature matching algorithm can deal with the sequence of images when the motion is “large”. Significant attention has been put on the feature matching algorithms due to their insensitivity to depth discontinuities and the presence of regions with uniform texture. Hierarchical feature matching and segmentation matching are two main classes of feature matching in recent research. Venkateswar and Chellappa [54] discussed the hierarchical feature matching where the matching starts at the highest level of the hierarchy (surface) and proceeds to the lowest level (lines). The matching process was more distinct in form and was easier due to the lower number of higher level features. Segmentation matching presented by Todorovic and Ahuja [55] aimed to identify the largest part in one image and its match in another image by measuring the maximum similarity that was defined in term of geometric and photometric properties of regions as well as the regions topology. However, no in depth discussion has been made on the hierarchical feature matching and segmentation matching. All the matched features used by the various researchers are local features represented by mathematical models, such as line, circle and corner [56]. The ground-breaking work of Low [57-58] is the development of the invariant features, which are invariant to image scale and rotation, and robust

matching results were shown across a substantial range of affine distortion, change in 3D viewpoint, noise, and change in illumination. Obviously, the invariant features [59-60] are highly distinctive so that a single feature can be correctly matched with high probability. It is a prosperous area attracting attentions and research efforts.

The last category of local correspondence algorithm is the correlation-based algorithms, which seek to find the corresponding points on the basis of similarity (correlation) between the corresponding areas in the left and right images. Sum of Absolute Differences (SAD) is one of the similarity measures which are calculated by subtracting pixels within a square neighbourhood between the left image I_l and the right image I_r followed by the aggregation of absolute difference within the square window. The corresponding points are then given by the window that has the maximum similarity. Although correlation-based algorithms are easier to implement and provide dense disparity map, it is sensitive to the changes in illumination directions and viewpoints. Some research works to achieve better and robust performance of the correlation-based algorithms can be found in the work of Aschwanden and Guggenbuhl [61] and Banks and Corke [62]. The former provided an extensive description and comparison on correlation-type registration algorithms, while, the latter presented the performance comparison of rank and census matching with those of correlation and difference metrics.

The most challenging problems for the local correspondence methods are the matching ambiguity and sensitivity because they are dealing with single pixel or group of neighbourhood pixels in the entire image. For a given feature in one image, there are multiple candidates in the other one which can be paired with this feature. Although in different forms, all algorithms use one common constraint, the disparity continuity or smoothness, to choose one matching point from among the multiple candidates. To reduce the mismatching chances and increase the computational speed of the local

matching approaches, a well-known constraint, epipolar geometry constraint was developed. It reduces the corresponding searches from the entire image to a straight line. We will be establishing the epipolar line of our single-lens bi-prism based stereovision system in this thesis and use it to realize the stereo correspondence. Details of the epipolar geometry are presented in Chapter IV.

2.2.2 Global stereo correspondence methods

Global matching approach is a powerful technique in stereo correspondence. The approach begins with the setting up of Global Energy function. It exploits non-local constraints in order to reduce sensitivity to the local regions of the image that fail to match due to occlusion and uniform texture, etc. The smoothness assumption is first made in the global matching for the disparity map calculation. Generally speaking, the global energy function contains two parts, smoothness energy and data energy, as shown below:

$$E(d) = E_{data}(d) + \lambda E_{smooth}(d) \quad , \quad (2.2)$$

Where the data term $E_{data}(d)$ measures how well the disparity agrees with the input image pair, and the smoothness term $E_{smooth}(d)$ encodes the smoothness assumption by the algorithm. There are many research papers published in the area of global stereo matching. Our review touches only on dynamic programming, graph cut and cooperative matching algorithms.

Dynamic Programming

Dynamic Programming (DP) is one of the global matching algorithms that reduce the computational complexity of optimization problems by decomposing them into smaller and simpler sub-problems [63]. A global cost function is computed in stages, with the transition between stages defined by a set of constraints. For stereo matching, the epipolar monotonic ordering constraint allows the global cost function to be determined by the

minimum cost path of a disparity space image (DSI). Generally, there are two common ways to construct DSI in DP, as shown in Fig. 2.10.

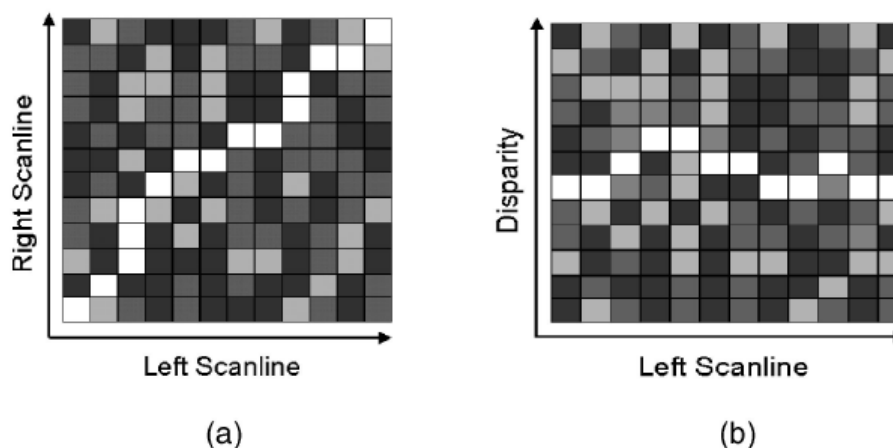


Fig. 2.10: DSI defined by (a) left-right scan-line; (b) left scan-line and left-disparity

The first step is to define DSI with the left and right scan lines [46][64] and in this case, DP is used to determine the minimum cost path from the lower left corner to the upper right corner of the DSI. With N pixels in a scan-line, the computational complexity using dynamic programming for this type of DSI is $O(N^4)$, in addition to the time required to compute the local cost functions. The second step is to construct DSI by the left scan line and left-disparity, as is done by Intille and Bobick [65]. With such construction, the minimum cost path computed by DP is from the first column to the last column of the DSI. With N pixels in a scan-line and a disparity range of D pixels, full global optimization requires $O(N^N)$ operations per scan-line, in addition to the time required to compute local cost functions. One greedy algorithm discussed by Birchfield and Tomasi [66] has reduced the complexity to $O(ND \log D)$ by pruning nodes when locally lower cost alternatives are available.

Other than the computational efficiency, DP also seeks to resolve the occlusion problem although it is more difficult because the cost function applied near an occlusion

boundary is typically high. Belhumeur [67] has proposed the methods to deal with the difficulties by replacing matching cost at occlusion boundaries with a small fixed occlusion cost. Furthermore, DP also provides global support for the local regions that lack texture and would otherwise cause mismatching. These local regions pose challenge for the global search since any cost function in these regions is low.

However, the principal disadvantage of DP is the possibility that local errors may be propagated along a scan-line, corrupting other potentially good matches. Horizontal streaks caused by this problem may be observed in many disparity map results. Another significant limitation of DP for stereo matching is its inability to strongly cooperate both horizontal and vertical continuity constraints. Many approaches have been proposed to improve this situation while maintaining the framework of DP. Among them, graph cut is one of the approaches.

Graph cut

Graph cut is a global matching approach which is efficient on the integration of the horizontal and vertical continuity constraints while maintaining the DP framework. It realizes the stereo correspondence by finding the maximum flow in graph through exploiting constraints. Naturally, graph cut methods require more computational efforts than DP. A great deal of effort has been spent in search of efficient solutions, such as, the well-known preflow-push lift-to-front algorithm proposed by Roy and Cox [68] and Zhao [69]. The complexity of this algorithm is $O(N^2 D^2 \log(ND))$, which is significantly greater than that of DP algorithms. However, the average observed time reported by Roy and Cox is $O(N^{1.2} D^{1.3})$, which is much closer to that of DP. One limitation of the left-to-front algorithm is that classical implementations require huge memory resources, making this approach cumbersome for large images. Thomos et al. [70] investigated an efficient

data structure that reduces one quarter of the memory resources, making the algorithm manageable for large data sets.

Recent works on graph cuts have produced both new graph architectures and energy minimization algorithms. Boykov and Kolmogorov [71] developed an approximate Ford-Fulkerson style augmenting paths algorithm, which is much faster than the standard push-re-label approach. Kolmogorov and Zabini [72] proposed a graph architecture in which the vertices represented the pixel correspondence (rather than pixels themselves) and imposed certain unique constraint to handle occlusion. The performance of these graphs cut methods has been proven to be among the best [52].

Cooperative optimization algorithm

Inspired by the computational models of human vision, cooperative algorithm is a newly developed global matching approach among the earliest ones for disparity computation [73]. Such algorithms iteratively perform local computations for matching scores by using nonlinear operations, such as uniqueness and continuity constraints and result in an overall behaviour similar to global optimization algorithms. Cooperative optimization algorithms have been proposed to apply for DNA image analyses [74], shape from shading [75] and stereo matching [76], etc. When applied in the case of stereo matching, the cooperative algorithms achieve comparable results with of graph cut in terms of solution quality, whereas it is twice as fast as graph cuts in software simulation using the common evaluation framework [52].

Recently, a promising variant of Marr and Poggio's original cooperative algorithm has been developed by Zintnick and Kanade [77]. They computed locally the matching scores using match windows and demonstrated the global behaviours by iteratively refining the correlation scores using the uniqueness and continuity constraints. In Zhang and Kambhamettu' work, the matching score was proposed to be calculated by image

segmentation [78]. The results of the segmentation were analyzed to prevent the support area from overlapping a depth discontinuity.

Global matching techniques for stereo correspondence are popular due to their powerful and robust solutions. Other than the aforesaid approaches, researchers are continuously investigating other approaches, such as nonlinear diffusion [79] and belief propagation [80] for better and robust stereo correspondence solutions.

In this chapter, we have first reviewed the available techniques of the single camera stereovision system, and then the existing algorithms for both the local and global stereo correspondence. Knowledge gained from the review provides the motivation in developing the single-lens bi-prism based stereovision system in this thesis. In addition, we shall employ both the camera calibration and ray sketching based approaches to solve the stereo correspondence problem.

Chapter III Camera calibration based approach for stereo correspondence and depth recovery of single-lens bi-prism based stereovision system

3.1 Real and virtual camera calibration technique

3.1.1 Introduction of the virtual camera model

In our work, we achieve the single-lens binocular stereovision effect with the aid of a bi-prism (2F filter) positioned in front of the CCD camera. Two stereo images of the same scene are formed at the real camera image plane as the results of refraction due to the prism. We can assume that the images are captured by two identical and symmetrically positioned *virtual cameras* which do not exist physically. The concept of virtual camera is illustrated in Fig. 3.1 and it forms the most important basis of our stereovision work.

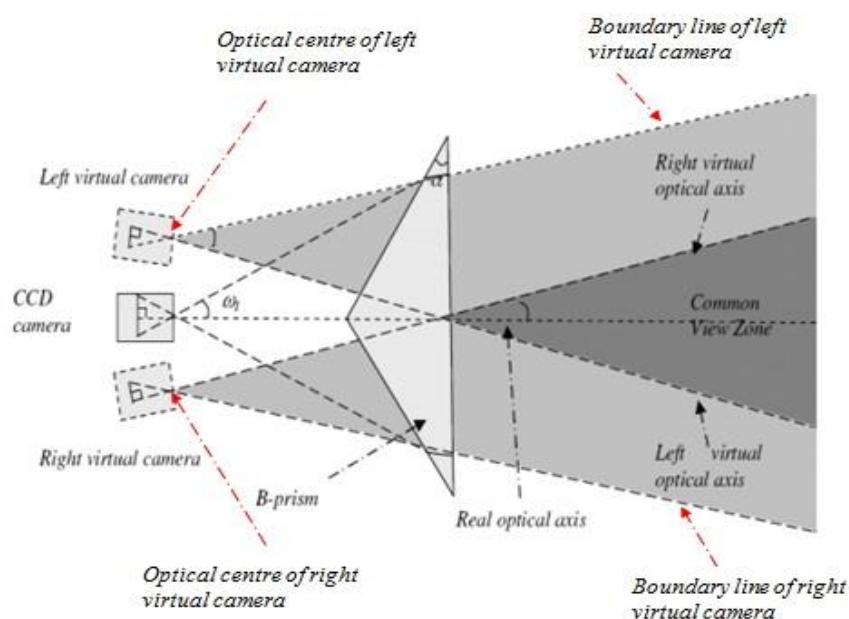


Fig. 3.1: Single-lens binocular stereovision system using bi-prism

Fig. 3.1 shows the generation of the two virtual cameras by the single-lens prism based stereovision setup. Each virtual camera is defined by two boundary lines: one is the optical axis of the virtual camera, which can be determined by back-extending the refracted ray along the real camera optical axis; and the another boundary line is determined by back-extending the refracted ray along the real camera field of view (FOV) boundary line(s). The intersection between the two above-mentioned lines defines the virtual camera optical centre. Their positions, orientations, and the focal lengths can be determined by virtual camera calibration, which will be presented in Section 3.1.2. 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. However, for the virtual camera generated in Fig. 3.1 to be valid, 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;
- 5) The two virtual cameras have identical properties and they are positioned symmetrically with respect to the camera axis.

The last condition can be achieved through proper selection of prism and accurate experimental setup. Fig. 3.2 gives a detail illustration of the left virtual camera generation as the right virtual camera formation can be inferred in a similar way. The image plane

of each virtual camera corresponds to only one half of the real image plane, i.e., the left (right) virtual camera image plane corresponds to the right (left) half of the real image plane, and hence its Principle Point (defined as the intersection between the image plane and the optical axis of a camera) is not at the centre of the virtual image plane centre, but at its boundary.

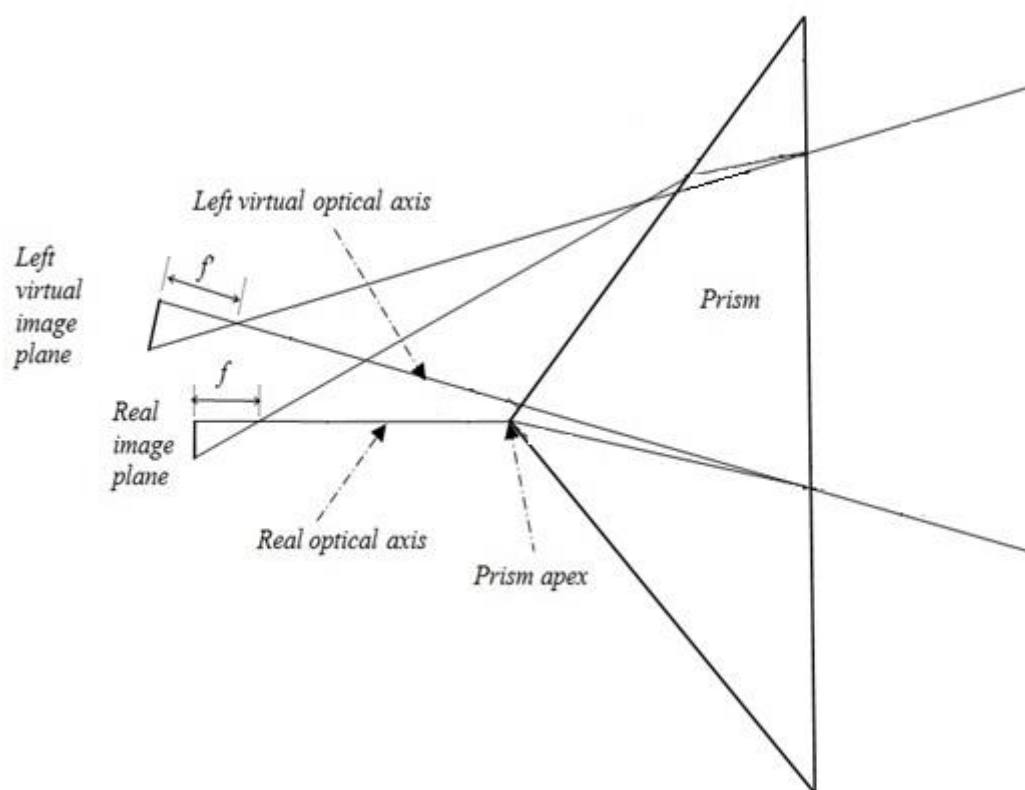


Fig. 3.2: Generation of the left virtual camera using bi-prism (top view)

3.1.2 Calibration of the real camera and the two virtual cameras

Camera calibration in stereovision is a process to determinate the system parameters related to camera operation, such as *the intrinsic parameters* which give the information about the optical and geometrical characteristics of the camera, for example, focal length, unknown scale factor and distortion parameters, and *the extrinsic parameters* which are the location and orientation of the camera in the 3D world coordinates. These parameters

are presented as rotation matrix and translation vector. The accuracy of camera calibration will affect directly on the performance of a stereovision system, especially in stereo correspondence and depth recovery. Therefore, many research efforts are made to deal with this challenge. The camera calibration techniques can be divided into three categories: direct non-linear minimization, linear transformation and hybrid methods.

Direct non-linear minimization methods are classical calibration techniques [81-82], which usually use the interactive algorithms to minimize the residual errors of some governing equations which can be obtained directly from the relationship between the intrinsic and extrinsic parameters. The accuracy of these methods could be quite good by considering the camera lens distortions, such as radial distortion and tangential distortion in the camera model. However, they are computational intensive, and the results could be unstable or erroneous if the boundary and/or initial conditions are not chosen correctly. In contrast, *linear transformation methods* present the relationship between the intrinsic and extrinsic parameters in the form of linear equations and therefore, the calibration parameters are to be determined in close-form solutions [83-85]. The algorithms are fast but the camera lens distortions are not taken into account. Hence, it is only applicable when the system error is quite small or camera lens distortions are negligible.

Hybrid methods make use of the advantages of the two previous categories to obtain the camera intrinsic and extrinsic parameters. Generally, these methods involve two steps: the first step is to solve most of the camera parameters linearly; while the second step employs simple nonlinear optimization to obtain the remaining parameters, such as lens distortion parameters. These methods are relatively fast and accurate. Therefore, they are widely studied and used in recent works [86-90].

For our single-lens bi-prism based stereovision system, calibration is a process to determinate the intrinsic and extrinsic parameters of a real camera as well as its two associated virtual cameras. Our proposed camera calibration algorithm falls on the category of the "*Linear Transformation Methods*", consisting of 4-coordinates system in Euclidean Coordinates and 4-steps of linear transformation to obtain all the required parameters. The four coordinates system is shown in Fig. 3.3 together with the relevant parameters.

We show graphically the geometry of the system setup. There are four coordinate system defined in Fig 3.3.

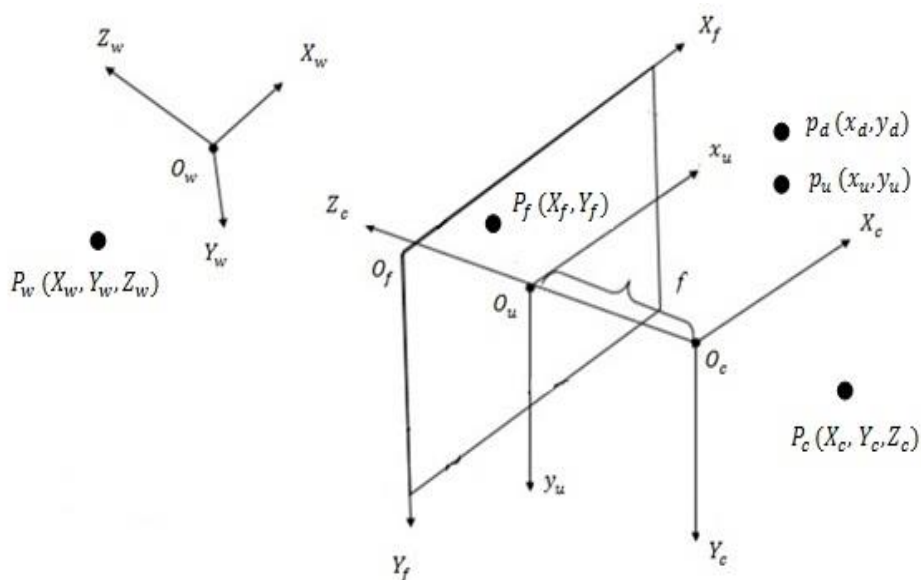


Fig. 3.3: Geometrical representation of the coordinates system

Screen Coordinates (view on computer screen) : $O_f[X_f, Y_f]$;

Camera Image Plane Coordinates (coordinates of the image CCD array) : $O_u[x_u, y_u]$;

Camera Coordinates : $O_c[X_c, Y_c, Z_c]$;

World Coordinates : $O_w[X_w, Y_w, Z_w]$;

The World Coordinates $[X_w, Y_w, Z_w]$ and its origin O_w shown in Fig. 3.3 is determined by users, which could be arbitrary chosen. The Camera Coordinates $[X_c, Y_c, Z_c]$ is defined with the origin O_c at the camera optical center. Its Z_c axis is aligned with the camera optical axis. X_c and Y_c are set to be parallel with x_u and y_u , respectively. The Camera Image Plane Coordinates has its origin at O_u , which is the intersection of the camera optical axis and the camera image plane. Lastly, the Screen Coordinates $[X_f, Y_f]$ is defined with the origin O_f located at the top-left corner of the image frame.

As illustrated in Fig. 3.3, an arbitrary point $P_w(X_w, Y_w, Z_w)$ in World Coordinates has its corresponding point $P_c(X_c, Y_c, Z_c)$ in Camera Coordinates. The image point $p_u(x_u, y_u)$ in the Camera Image Plane Coordinates is the perspective projection of $P_c(X_c, Y_c, Z_c)$ based on the pin-hole camera model. Point $p_d(x_d, y_d)$ in Camera Image Plane Coordinates is the real image point corresponding to $P_c(X_c, Y_c, Z_c)$, which is different from $p_u(x_u, y_u)$ due to the system distortions. Through the camera frame grabber, we will have the digital image point $P_f(X_f, Y_f)$ in the Screen Coordinates. The relationship between $P_w(X_w, Y_w, Z_w)$ and $P_f(X_f, Y_f)$ can be described by our developed 4-steps linear transformation. The camera system calibration is then achieved by determining all the parameters involving in this relationship. The mathematical presentations of the 4-steps transformation are shown as follows.

(1) First transformation: from World Coordinates to Camera Coordinates

Rigid body transformation is involved to transform the object from World Coordinates $O_w[X_w, Y_w, Z_w]$ to Camera Coordinates $O_c[X_c, Y_c, Z_c]$, which can be

represented by a rotational matrix $R = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix}$ and translational vector $T = \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$

as:

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}, \quad (3.1)$$

Here the subscripts c and w denote Camera Coordinates and World Coordinates, respectively.

(2) Second transformation: from Camera Coordinates to Camera Image Plane Coordinates

Perspective projection and pin-hole camera model are the bridge to transform the Camera Coordinates to Camera Image Plane Coordinates. Eq. (3.2) describes the projection model with the focal length f to be calibrated.

$$x_u = f \frac{X_c}{Z_c}, \quad y_u = f \frac{Y_c}{Z_c}. \quad (3.2)$$

(3) Third transformation: from $p_u(x_u, y_u)$ to distorted image point $p_d(x_d, y_d)$

Due to the imperfection of the camera lens, actual image $p_d(x_d, y_d)$ projected on the image plane differs from the ones obtained theoretically through the perspective projection. In general, radial lens distortion is the major distortion of camera lens, which can be represented by the distortion coefficients g_1 and g_2 is in the form of:

$$x_d + D_x = x_u, \quad y_d + D_y = y_u, \quad (3.3)$$

$$D_x = x_d(g_1 r^2 + g_2 r^4), \quad D_y = y_d(g_1 r^2 + g_2 r^4),$$

$$r = \sqrt{x_d^2 + y_d^2}, \quad (3.4)$$

Where D_x and D_y represent the camera distortion model. This model is determined based on the physical camera properties. If D_x and D_y are equal to zero, the camera model is taken as a distortion-free model.

(4) Fourth transformation: from Camera Image Plane Coordinates to Screen Coordinates

The final transformation is carried out to convert the Camera Image Plane Coordinates $O_u[x_u, y_u]$ to Screen Coordinates $O_f[X_f, Y_f]$.

$$X_f = \frac{s_x \cdot N_{fx}}{d_x \cdot N_{cx}} x_d + C_x, Y_f = \frac{N_{fy}}{d_y \cdot N_{cy}} y_d + C_y, \quad (3.5)$$

Where:

- (X_f, Y_f) : an image pixel coordinates in the Screen Coordinates;
- (C_x, C_y) : x and y coordinates values of the centre point of the computer frame;
- d_x : centre to centre distance between adjacent sensor elements along X_f direction;
- d_y : centre to centre distance between adjacent sensor elements along Y_f direction;
- N_{cx} : number of sensor elements along x_u direction;
- N_{cy} : number of sensor elements along y_u direction;
- N_{fx} : number of pixels in X_f direction of the Screen;
- N_{fy} : number of pixels in Y_f direction of the Screen;
- s_x : uncertainty image scale factor.

Usually, the manufacturer of the CCD cameras supplies the information of d_x and d_y . If quantization error is negligible, we will have $N_{cx} = N_{fx}$ and $N_{cy} = N_{fy}$. Due to a variety of factors, such as slight timing mismatch between the image acquisition

hardware and the computer monitor, an uncertainty parameter s_x , mentioned by Ngiam [12], was introduced into the above equations. However, this parameter s_x is found to be close to 1 for this project through many repeated experiments.

By substituting Eq. (3.1) to Eq. (3.2), which combines effectively the first two transformations, we have:

$$\begin{aligned} x_u &= f \frac{r_{1,1}X_w + r_{1,2}Y_w + r_{1,3}Z_w + T_x}{r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z}, \\ y_u &= f \frac{r_{2,1}X_w + r_{2,2}Y_w + r_{2,3}Z_w + T_y}{r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z}. \end{aligned} \quad (3.6)$$

In this thesis, we assume that the camera system is distortion-free, it means $D_x = D_y = 0$. Eq. (3.3) will then become:

$$x_d = x_u, \quad y_d = y_u. \quad (3.7)$$

Combining the 4-steps transformation by substituting Eq. (3.6) and Eq. (3.7) into Eq. (3.5), the digital image point $P_f(X_f, Y_f)$ is linked to its corresponding 3D world point $P_w(X_w, Y_w, Z_w)$ by:

$$\begin{cases} X_f = f \frac{s_x}{d'_x} \frac{r_{1,1}X_w + r_{1,2}Y_w + r_{1,3}Z_w + T_x}{r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z} + C_x, \\ Y_f = \frac{f}{d'_y} \frac{r_{2,1}X_w + r_{2,2}Y_w + r_{2,3}Z_w + T_y}{r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z} + C_y. \end{cases} \quad (3.8)$$

Rewriting Eq. (3.8) into the following form:

$$\begin{cases} X_f k = (fs_x d'_x{}^{-1} r_{1,1} + C_x r_{3,1})X_w + (fs_x d'_x{}^{-1} r_{1,2} + C_x r_{3,2})Y_w + (fs_x d'_x{}^{-1} r_{1,3} + C_x r_{3,3})Z_w + (fs_x d'_x{}^{-1} T_x + C_x T_z) \\ Y_f k = (fd'_y{}^{-1} r_{2,1} + C_y r_{3,1})X_w + (fd'_y{}^{-1} r_{2,2} + C_y r_{3,2})Y_w + (fd'_y{}^{-1} r_{2,3} + C_y r_{3,3})Z_w + (fd'_y{}^{-1} T_y + C_y T_z) \end{cases} \quad (3.9)$$

Where $k = r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z$

Rearranging the two equations of Eq. (3.9) into the form of matrix, we are able to generate a 3 by 4 projection matrix M consisting of all the need-to-be-calibrated parameters in Homogenous Coordinates.

$$\begin{bmatrix} X_f k \\ Y_f k \\ k \end{bmatrix} = \begin{bmatrix} fs_x d'_x{}^{-1} r_{1,1} + C_x r_{3,1} & fs_x d'_x{}^{-1} r_{1,2} + C_x r_{3,2} & fs_x d'_x{}^{-1} r_{1,3} + C_x r_{3,3} & fs_x d'_x{}^{-1} T_x + C_x T_z \\ fd'_y{}^{-1} r_{2,1} + C_y r_{3,1} & fd'_y{}^{-1} r_{2,2} + C_y r_{3,2} & fd'_y{}^{-1} r_{2,3} + C_y r_{3,3} & fd'_y{}^{-1} T_y + C_y T_z \\ r_{3,1} & r_{3,2} & r_{3,3} & T_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3.10)$$

$$\Rightarrow \begin{bmatrix} X_f k \\ Y_f k \\ k \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

Further decomposition of M in Eq. (3.10) leads to Eq. (3.11), which defines intrinsic and extrinsic parameters of the camera system as:

$$M_{3 \times 4} = \begin{bmatrix} fs_x d'_x{}^{-1} & 0 & C_x \\ 0 & fd'_y{}^{-1} & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & T_x \\ r_{2,1} & r_{2,2} & r_{2,3} & T_y \\ r_{3,1} & r_{3,2} & r_{3,3} & T_z \end{bmatrix}, \quad (3.11)$$

Where:

$$M_{intrinsic} = \begin{bmatrix} fs_x d'_x{}^{-1} & 0 & C_x \\ 0 & fd'_y{}^{-1} & C_y \\ 0 & 0 & 1 \end{bmatrix}, \quad M_{extrinsic} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & T_x \\ r_{2,1} & r_{2,2} & r_{2,3} & T_y \\ r_{3,1} & r_{3,2} & r_{3,3} & T_z \end{bmatrix}.$$

In short, the camera calibration model can be expressed in terms of the projection matrix M in Homogenous Coordinates as:

$$\begin{bmatrix} X_f k \\ Y_f k \\ k \end{bmatrix} = M_{intrinsic} \cdot M_{extrinsic} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (3.12)$$

Our proposed model combines the aforesaid four transformations, and will involve all the 13 unknown parameters, which is different from that of Tsai [6]. With the 13 unknowns, a minimum of 7 data points is required to yield a non-trivial solution. For $n > 6$

data points selected for the calibration purpose, we can form at least $2n$ equations to solve the 13 unknowns. The objective of the camera system calibration is then fulfilled. The following shows the derivations in details.

From Eq. (3.8), we obtain Eq. (3.13) and Eq. (3.14):

$$\begin{cases} d_x' X_f' = fs_x \frac{r_{1,1}X_w + r_{1,2}Y_w + r_{1,3}Z_w + T_x}{r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z} \\ d_y' Y_f' = f \frac{r_{2,1}X_w + r_{2,2}Y_w + r_{2,3}Z_w + T_y}{r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z} \end{cases}, \quad (3.13)$$

$$\begin{cases} d_x' X_f' (r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z) - fs_x (r_{1,1}X_w + r_{1,2}Y_w + r_{1,3}Z_w + T_x) = 0 \\ d_y' Y_f' (r_{3,1}X_w + r_{3,2}Y_w + r_{3,3}Z_w + T_z) - f (r_{2,1}X_w + r_{2,2}Y_w + r_{2,3}Z_w + T_y) = 0 \end{cases}, \quad (3.14)$$

with $X_f' = X_f - C_x$ and $Y_f' = Y_f - C_y$

Dividing throughout Eq. (3.14) by T_z , we have:

$$\begin{bmatrix} -X_w \\ -Y_w \\ -Z_w \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ d_x' X_f' X_w \\ d_x' X_f' Y_w \\ d_x' X_f' Z_w \end{bmatrix}^T \cdot \begin{bmatrix} fs_x r_{1,1} / T_z \\ fs_x r_{1,2} / T_z \\ fs_x r_{1,3} / T_z \\ fs_x T_x / T_z \\ fs_x r_{2,1} / T_z \\ fs_x r_{2,2} / T_z \\ fs_x r_{2,3} / T_z \\ fs_x T_y / T_z \\ r_{3,1} / T_z \\ r_{3,2} / T_z \\ r_{3,3} / T_z \end{bmatrix} = -d_x' X_f'$$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ -X_w \\ -Y_w \\ -Z_w \\ -1 \\ d_x'Y_f'X_w \\ d_x'Y_f'Y_w \\ d_x'Y_f'Z_w \end{bmatrix}^T \begin{bmatrix} fs_x r_{1,1} / T_z \\ fs_x r_{1,2} / T_z \\ fs_x r_{1,3} / T_z \\ fs_x T_x / T_z \\ fs_x r_{2,1} / T_z \\ fs_x r_{2,2} / T_z \\ fs_x r_{2,3} / T_z \\ fs_x T_y / T_z \\ r_{3,1} / T_z \\ r_{3,2} / T_z \\ r_{3,3} / T_z \end{bmatrix} = -d_x'Y_f' \quad (3.15)$$

The two equations in Eq. (3.15) are derived from one image point and its corresponding 3D point. To recover all the unknown parameters, at least 7 data points ($n > 6$) are required. Hence, there are $2n$ equations. By combining all the $2n$ equations, and representing them into matrix form $AX = B$, we have Eq. (3.16).

$$A = \begin{bmatrix} -X_{w,1} & -Y_{w,1} & -Z_{w,1} & -1 & 0 & 0 & 0 & 0 & d_x'X_{f,1}'X_{w,1} & d_x'X_{f,1}'Y_{w,1} & d_x'X_{f,1}'Z_{w,1} \\ 0 & 0 & 0 & 0 & -X_{w,1} & -Y_{w,1} & -Z_{w,1} & -1 & d_x'Y_{f,1}'X_{w,1} & d_x'Y_{f,1}'Y_{w,1} & d_x'Y_{f,1}'Z_{w,1} \\ " & " & " & " & " & " & " & " & " & " & " \\ " & " & " & " & " & " & " & " & " & " & " \\ -X_{w,n} & -Y_{w,n} & -Z_{w,n} & -1 & 0 & 0 & 0 & 0 & d_x'X_{f,n}'X_{w,n} & d_x'X_{f,n}'Y_{w,n} & d_x'X_{f,n}'Z_{w,n} \\ 0 & 0 & 0 & 0 & -X_{w,n} & -Y_{w,n} & -Z_{w,n} & -1 & d_x'Y_{f,n}'X_{w,n} & d_x'Y_{f,n}'Y_{w,n} & d_x'Y_{f,n}'Z_{w,n} \end{bmatrix};$$

$$X = \begin{bmatrix} fs_x r_{1,1} / T_z \\ fs_x r_{1,2} / T_z \\ fs_x r_{1,3} / T_z \\ fs_x T_x / T_z \\ fs_x r_{2,1} / T_z \\ fs_x r_{2,2} / T_z \\ fs_x r_{2,3} / T_z \\ fs_x T_y / T_z \\ r_{3,1} / T_z \\ r_{3,2} / T_z \\ r_{3,3} / T_z \end{bmatrix}; \quad B = \begin{bmatrix} -d_{x,1}'X_{f,1}' \\ -d_{x,1}'Y_{f,1}' \\ " \\ " \\ -d_{x,n}'X_{f,n}' \\ -d_{x,n}'Y_{f,n}' \end{bmatrix} \quad (3.16)$$

With the n data points $P_{w,1} (X_{w,1}, Y_{w,1}, Z_{w,1}) \dots \dots P_{w,n} (X_{w,n}, Y_{w,n}, Z_{w,n})$ ($n > 6$) in the World Coordinates obtained directly from the calibration board and their corresponding image points $P_{f,1} (X_{f,1}, Y_{f,1}) \dots \dots P_{f,n} (X_{f,n}, Y_{f,n})$ on the Screen Coordinates, X is solvable using the Least Square method as:

$$X = (A^T * A)^{-1} A^T * B \quad (3.17)$$

Subsequently, rearranging all the solved parameters in X into a matrix M' , we obtain

$$M' = \begin{bmatrix} fs_x r_{1,1} / T_z & fs_x r_{1,2} / T_z & fs_x r_{1,3} / T_z & fs_x T_x / T_z \\ fr_{2,1} / T_z & fr_{2,2} / T_z & fr_{2,3} / T_z & fT_y / T_z \\ r_{3,1} / T_z & r_{3,2} / T_z & r_{3,3} / T_z & 1 \end{bmatrix} \quad (3.18)$$

$$= \begin{bmatrix} a_1 & a_2 & a_3 & a_4 \\ a_5 & a_6 & a_7 & a_8 \\ a_9 & a_{10} & a_{11} & 1 \end{bmatrix}$$

With all the known parameters in M' , the 13 camera calibration parameters, including 9 rotational parameters, 2 translational parameters and 2 camera intrinsic parameters are determined as shown in Eq. (3.19) by considering the orthogonal property and the right-handed property of R.

$$r_{1,1} = \frac{a_1 * T_z}{f * s_x}; r_{1,2} = \frac{a_2 * T_z}{f * s_x}; r_{1,3} = \frac{a_3 * T_z}{f * s_x};$$

$$r_{2,1} = \frac{a_5 * T_z}{f}; r_{2,2} = \frac{a_6 * T_z}{f}; r_{2,3} = \frac{a_7 * T_z}{f};$$

$$r_{3,1} = a_9 * T_z; r_{3,2} = a_{10} * T_z; r_{3,3} = a_{11} * T_z; \quad (3.19)$$

$$T_x = \frac{a_4 * T_z}{f * s_x}; T_y = \frac{a_8 * T_z}{f};$$

$$f = \sqrt{T_z^2 (a_5^2 + a_6^2 + a_7^2)}; s_x = \sqrt{T_z^2 (a_5^2 + a_6^2 + a_7^2) / f^2};$$

$$T_z = \sqrt{a_9^2 + a_{10}^2 + a_{11}^2} .$$

Intuitively, by substituting all the solved parameters in Eq. (3.19) into Eq. (3.11), the system is then considered to have been calibrated. The above calibration procedures work not only on the real camera but also on the two virtual cameras of our stereovision system. We just need to ensure that there are $n>6$ calibration input data points and construct the equations and matrix accordingly. In summary, our camera model in Homogenous Coordinates is:

For the real camera (Eq. (3.12)):

$$\begin{bmatrix} X_f k \\ Y_f k \\ k \end{bmatrix} = M_{intrinsic} \cdot M_{extrinsic} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} .$$

For the left virtual camera, analogously, we have:

$$A_{LeftVirtual} \cdot X_{LeftVirtual} = B_{LeftVirtual} \quad . \quad (3.20)$$

By obtaining $n>6$ data points $[X_{w,1}, Y_{w,1}, Z_{w,1}] \dots \dots [X_{w,n}, Y_{w,n}, Z_{w,n}]$ from the calibration board in World Coordinates and the corresponding image points $[X_{lf,1}, Y_{lf,1}] \dots \dots [X_{lf,n}, Y_{lf,n}]$ of the left virtual camera, the 13 calibration parameters for the left virtual camera can be obtained as:

$$\begin{bmatrix} X_{lf} k_l \\ Y_{lf} k_l \\ k_l \end{bmatrix} = M_{intrinsic}^L \cdot M_{extrinsic}^L \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M^L \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad . \quad (3.21)$$

Similarly, for the right virtual camera, Eq. (3.22) holds,

$$A_{RightVirtual} \cdot X_{RightVirtual} = B_{RightVirtual}$$

$$\begin{bmatrix} X_{rf} k_r \\ Y_{rf} k_r \\ k_r \end{bmatrix} = M_{intrinsic}^R M_{extrinsic}^R \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = M^R \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (3.22)$$

Here M^L and M^R are the 3 by 4 projection matrices of the left and right virtual camera.

3.2 Stereo correspondence of the single-lens bi-prism based stereovision system through camera calibration

To recover a 3D scene, stereovision technique is employed to capture two views of the scene at different viewpoints. However, in this process, we must ensure that a same scene point is visible in two or more stereo images/views. The process of recovering these two points in the two or more views is known as stereo correspondence.

Ideally, the linear relationship between the 2D image point and its corresponding 3D world point always exists for the pin-hole camera model in the form of projection matrix as discussed in Section 3.1.2. All the variables in the projection matrix are the combination of the intrinsic and extrinsic parameters of the stereovision system and are obtainable through camera calibrations. The real camera and the two virtual camera models of our single-lens binocular stereovision system are derived in Eq. (3.12), (3.21) and (3.22), respectively.

We can infer from Eq. (3.21) and Eq. (3.22) that it is possible to establish the correspondence between the right and left images due to the common $(X_w, Y_w, Z_w, 1)$ in the same Homogenous Coordinates. For point $P_w(X_w, Y_w, Z_w)$ in the World Coordinates, its corresponding image point $P_{lf}(X_{lf}, Y_{lf})$ in the left virtual camera and $P_{rf}(X_{rf}, Y_{rf})$ in the right virtual camera has the following relationship by substituting Eq. (3.21) to Eq. (3.22):

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = M^R (M^L)^{-1} \begin{bmatrix} X_{lf}k_l \\ Y_{lf}k_l \\ k_l \end{bmatrix}. \quad (3.23)$$

Here, all the parameters of M^L and M^R are the calibrated parameters of the left and right virtual cameras. They are known parameters and their inverse form $(M^L)^{-1}$ and $(M^R)^{-1}$ are also obtainable. We can write them in the form of:

$$M^L = \begin{bmatrix} m_{1,1} & m_{1,2} & m_{1,3} & m_{1,4} \\ m_{2,1} & m_{2,2} & m_{2,3} & m_{2,4} \\ m_{3,1} & m_{3,2} & m_{3,3} & m_{3,4} \end{bmatrix} = \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix}, \quad (3.24)$$

$$M^R = \begin{bmatrix} m'_{1,1} & m'_{1,2} & m'_{1,3} & m'_{1,4} \\ m'_{2,1} & m'_{2,2} & m'_{2,3} & m'_{2,4} \\ m'_{3,1} & m'_{3,2} & m'_{3,3} & m'_{3,4} \end{bmatrix} = \begin{bmatrix} m'_1 \\ m'_2 \\ m'_3 \end{bmatrix}, \quad (3.25)$$

$$(M^L)^{-1} = \begin{bmatrix} n_{1,1} & n_{1,2} & n_{1,3} \\ n_{2,1} & n_{2,2} & n_{2,3} \\ n_{3,1} & n_{3,2} & n_{3,3} \\ n_{4,1} & n_{4,2} & n_{4,3} \end{bmatrix} = [n_1, n_2, n_3], \quad (3.26)$$

$$(M^R)^{-1} = \begin{bmatrix} n'_{1,1} & n'_{1,2} & n'_{1,3} \\ n'_{2,1} & n'_{2,2} & n'_{2,3} \\ n'_{3,1} & n'_{3,2} & n'_{3,3} \\ n'_{4,1} & n'_{4,2} & n'_{4,3} \end{bmatrix} = [n'_1, n'_2, n'_3]. \quad (3.27)$$

Thus,

$$X_{rf} = (m'_1 n_1 X_{lf} + m'_1 n_2 Y_{lf} + m'_1 n_3 k_l) / k_r,$$

$$Y_{rf} = (m'_2 n_1 X_{lf} + m'_2 n_2 Y_{lf} + m'_2 n_3 k_l) / k_r,$$

$$k_r = m'_3 n_1 X_{lf} + m'_3 n_2 Y_{lf} + m'_3 n_3 k_l. \quad (3.28)$$

Using Eq. (3.28), we can obtain the correspondence image points in the right image plane. Note that k_l and k_r are scale factors. At this juncture, we can conclude that,

through the camera calibration, the stereo correspondence of our single-lens prism based stereovision system is established.

This camera calibration based method is easy and straightforward to apply in the solving of the stereo correspondence once the projection matrices are determined. However, the projection matrix can be generated only under the assumptions that 1) the stereovision system is distortion-free and 2) the pin-hole camera model is valid. These two assumptions are hardly valid in practice. Once the distortion model is considered in the calibration technique, this linear relationship between the 2D image and 3D object will be replaced by a nonlinear one, and the solution can only be obtained by an iterative process. It is certain that the accuracy of the stereo correspondence would be improved if the system distortion is taken into account in the camera calibration. This would be a good future research direction to refine the prism based stereovision system.

3.3 Depth recovery of single-lens bi-prism stereovision system

Mathematically speaking, to recover the 3D coordinates of a point on an object, there must be at least 3 equations due to the fact that there are three unknowns: X_w, Y_w and Z_w . Obviously, a single camera capturing only one single image without any other cues can only provide two equations through the knowledge of the image coordinates (X_f, Y_f) , resulting in an under-determined problem. In contrast, with the setup of a single-lens prism based stereovision system, the two virtual cameras can provide us with two sets of coordinates, (X_{lf}, Y_{lf}) and (X_{rf}, Y_{rf}) , of the two corresponding points in the two images, thus giving us sufficient conditions to solve the three unknowns (the coordinates of a 3D point). In this regards, however, it is necessary to calibrate the virtual cameras to provide us the correspondence relationship between (X_{lf}, Y_{lf}) and (X_{rf}, Y_{rf}) together with

(X_w, Y_w, Z_w) . Once this is done, the depth recovery of a 3D scene could theoretically be achieved.

With the obtained stereo correspondence information discussed in Section 3.2 and both real and virtual camera calibration in Section 3.1, 3D information of our stereovision system can be computed as follows.

Recalling Eq. (3.9) in Section 3.1.2,

$$\begin{cases} X_f k = (f s_x d'_x{}^{-1} r_{1,1} + C_x r_{3,1}) X_w + (f s_x d'_x{}^{-1} r_{1,2} + C_x r_{3,2}) Y_w + (f s_x d'_x{}^{-1} r_{1,3} + C_x r_{3,3}) Z_w + (f s_x d'_x{}^{-1} T_x + C_x T_z) \\ Y_f k = (f d'_y{}^{-1} r_{2,1} + C_y r_{3,1}) X_w + (f d'_y{}^{-1} r_{2,2} + C_y r_{3,2}) Y_w + (f d'_y{}^{-1} r_{2,3} + C_y r_{3,3}) Z_w + (f d'_y{}^{-1} T_y + C_y T_z) \end{cases}, \quad (3.9)$$

With $k = r_{3,1} X_w + r_{3,2} Y_w + r_{3,3} Z_w + T_z$

Rearranging Eq. (3.9), we have:

$$\begin{cases} (X_f r_{3,1} - \frac{f s_x}{d'_x} r_{1,1} - C_x r_{3,1}) X_w + (X_f r_{3,2} - \frac{f s_x}{d'_x} r_{1,2} - C_x r_{3,2}) Y_w + (X_f r_{3,3} - \frac{f s_x}{d'_x} r_{1,3} + C_x r_{3,3}) Z_w, \\ = \frac{f s_x}{d'_x} T_x + C_x T_z - X_f T_z \end{cases}, \quad (3.29)$$

Re-write as,

$$\begin{bmatrix} X_f r_{3,1} - \frac{f s_x}{d'_x} r_{1,1} - C_x r_{3,1} \\ X_f r_{3,2} - \frac{f s_x}{d'_x} r_{1,2} - C_x r_{3,2} \\ X_f r_{3,3} - \frac{f s_x}{d'_x} r_{1,3} + C_x r_{3,3} \end{bmatrix}^T \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \frac{f s_x}{d'_x} T_x + C_x T_z - X_f T_z, \quad (3.30)$$

And

$$\begin{cases} (Y_f r_{3,1} - \frac{f}{d'_y} r_{2,1} - C_y r_{3,1}) X_w + (Y_f r_{3,2} - \frac{f}{d'_y} r_{2,2} - C_x r_{3,2}) Y_w + (Y_f r_{3,3} - \frac{f}{d'_y} r_{2,3} - C_x r_{3,3}) Z_w, \\ = \frac{f}{d'_y} T_y + C_y T_z - Y_f T_z \end{cases} \quad (3.31)$$

Re-write as,

$$\begin{bmatrix} Y_f r_{3,1} - \frac{f}{d'_y} r_{2,1} - C_y r_{3,1} \\ Y_f r_{3,2} - \frac{f}{d'_y} r_{2,2} - C_x r_{3,2} \\ Y_f r_{3,3} - \frac{f}{d'_y} r_{2,3} - C_x r_{3,3} \end{bmatrix}^T \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = \frac{f}{d'_y} T_y + C_y T_z - Y_f T_z \quad (3.32)$$

Combining Eq. (3.31) and Eq. (3.32), we can express the results in the form of matrix:

$$HI = G \quad (3.33)$$

With

$$H = \begin{bmatrix} (X_{lf} r_{3,1} - \frac{fs_x}{d'_x} r_{1,1} - C_x r_{3,1})_{LEFT} & (Y_{lf} r_{3,1} - \frac{f}{d'_y} r_{1,2} - C_y r_{3,2})_{LEFT} & (X_{lf} r_{3,3} - \frac{fs_x}{d'_x} r_{1,3} + C_x r_{3,3})_{LEFT} \\ (Y_{lf} r_{3,1} - \frac{f}{d'_y} r_{2,1} - C_y r_{3,1})_{LEFT} & (Y_{lf} r_{3,2} - \frac{f}{d'_y} r_{2,2} - C_x r_{3,2})_{LEFT} & (Y_{lf} r_{3,3} - \frac{f}{d'_y} r_{2,3} - C_x r_{3,3})_{LEFT} \\ (X_{rf} r_{3,1} - \frac{fs_x}{d'_x} r_{1,1} - C_x r_{3,1})_{RIGHT} & (Y_{rf} r_{3,1} - \frac{f}{d'_y} r_{1,2} - C_y r_{3,2})_{RIGHT} & (X_{rf} r_{3,3} - \frac{fs_x}{d'_x} r_{1,3} + C_x r_{3,3})_{RIGHT} \\ (Y_{rf} r_{3,1} - \frac{f}{d'_y} r_{2,1} - C_y r_{3,1})_{RIGHT} & (Y_{rf} r_{3,2} - \frac{f}{d'_y} r_{2,2} - C_x r_{3,2})_{RIGHT} & (Y_{rf} r_{3,3} - \frac{f}{d'_y} r_{2,3} - C_x r_{3,3})_{RIGHT} \end{bmatrix};$$

$$I = \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix};$$

$$G = \begin{bmatrix} (\frac{fs_x}{d'_x} T_x + C_x T_z - X_{lf} T_z)_{LEFT} \\ (\frac{f}{d'_y} T_y + C_y T_z - Y_{lf} T_z)_{LEFT} \\ (\frac{fs_x}{d'_x} T_x + C_x T_z - X_{rf} T_z)_{RIGHT} \\ (\frac{f}{d'_y} T_y + C_y T_z - Y_{rf} T_z)_{RIGHT} \end{bmatrix}.$$

Note that the subscripts *LEFT* and *RIGHT* represent the parameters (recovered using the calibration matrix) for the left and right virtual cameras, respectively. (X_{lf}, Y_{lf}) and (X_{rf}, Y_{rf}) are the known corresponding image pairs. Other parameters in G and H are the two virtual camera intrinsic and extrinsic parameters. They are all known values after the calibration step. The 3D recovery of the point $P_w (X_w, Y_w, Z_w)$ in the World Coordinates is thus solved by:

$$I = [X_w, Y_w, Z_w]^T = (H^T * H)^{-1} * H^T * G . \quad (3.34)$$

Subsequently, using the calibration results of the real camera, we can recover the 3D information of the points/regions of interest in the Camera Coordinates.

$$\begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} = \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix},$$

$$\begin{aligned} \Rightarrow X_c &= r_{1,1} * X_w + r_{1,2} * Y_w + r_{1,3} * Z_w + T_x, \\ \Rightarrow Y_c &= r_{2,1} * X_w + r_{2,2} * Y_w + r_{2,3} * Z_w + T_y, \\ \Rightarrow Z_c &= r_{3,1} * X_w + r_{3,2} * Y_w + r_{3,3} * Z_w + T_z. \end{aligned} \quad (3.35)$$

3.4 Summary

The algorithm proposed above, based on the virtual and real camera calibration allows direct reconstruction of the 3D scene, which is different from the methods proposed in the existing literature. Lim and Xiao [35] achieved their results based on averaging the depths recovered with an extra constraint of a Y coordinate, which requires that the image planes of the two virtual cameras are coplanar. While, our developed method is still valid even the two virtual camera image planes are subtending at an angle. Minor misalignments in the prism may also be tolerated due to the nature of the recovery equations.

Camera calibration based approach is straightforward to achieve the objectives of the stereovision system - stereo correspondence and depth recovery. This method is explored fully in this project. However, it is normally tedious and time-consuming for real-time applications. Whatever changes in the camera, 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. In the next section, an approach of determining stereovision system by using the geometrical analyses of ray sketching is proposed and discussed, which can achieve the objectives without camera calibration process and hence resulting in an easier system implementation process.

Chapter IV Ray sketching based approach for stereo correspondence and depth recovery of single-lens bi-prism based stereovision system

4.1 Introduction of epipolar geometry

A concept in stereovision, known as epipolar geometry, presented by Grew and Kak [11] is popularly used to ease the difficulties and reduce the mismatches of stereo correspondence. Epipole, epipolar line and epipolar plane are the three major concepts in the epipolar geometry.

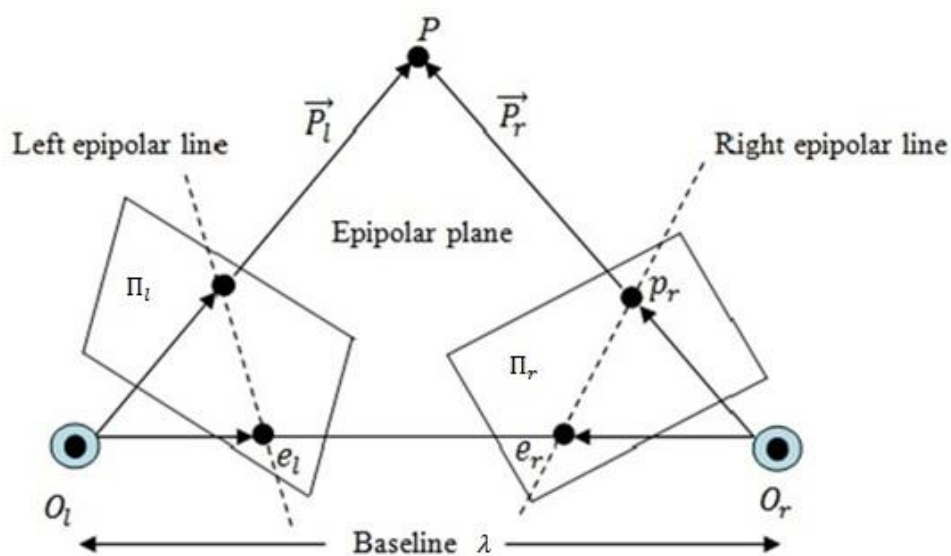


Fig. 4.1: Illustration of the epipolar geometry

O_l and O_r are the optical centers of the left and right cameras and the distance λ between them is called the baseline. Π_l and Π_r are the left and right camera image planes, respectively. For an arbitrary object point P , the points p_l and p_r are the two projection image points located on the left and right image plane, respectively. Epipolar geometry

defines the intersection point between camera optical axis and camera image plane as *epipole*, which is indicated in Fig. 4.1 as the left epipole e_l and right epipole e_r . The object point P and the two cameras optical centers 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*. As shown by the dash line in Fig. 4.1, line $e_l p_l$ is the left epipolar line and line $e_r p_r$ is the right epipolar line. Referring to Fig. 4.1, the concept of the epipolar geometry can be stated 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 we establish the epipolar geometry, the correspondence search will be reduced from the entire image to a two-dimensional epipolar line. Alternatively, the same knowledge can be used to verify whether a candidate match lies on the corresponding epipolar line. This is usually the most effective procedures to reject false matches due to occlusions.

As we know, all the global matching algorithms for stereo correspondence have a pre-requisite condition: the epipolar line is essentially horizontal. This can be achieved by setting the cameras' optical axes parallel to each other, and aligning the two image planes along a straight line, usually along the x -axis (Fig. 4.2). Epipolar geometry under this configuration is called non-verged geometry.

a) The non-verged geometry of stereovision system

In the non-verged geometry of a stereovision system, the baseline is parallel to the camera x axis (Fig. 4.2). Epipolar geometry implies that a point in space is projected to two locations on the same scan line in the left and right camera images. The resulting displacement of the two projected points is termed as disparity and it only has x component ($x_r - x_l$). Therefore, the stereo correspondence search for non-verged

stereovision system is only a 1D search along the x -axis. Due to its simplification, this configuration is commonly used to solve the stereo correspondence.

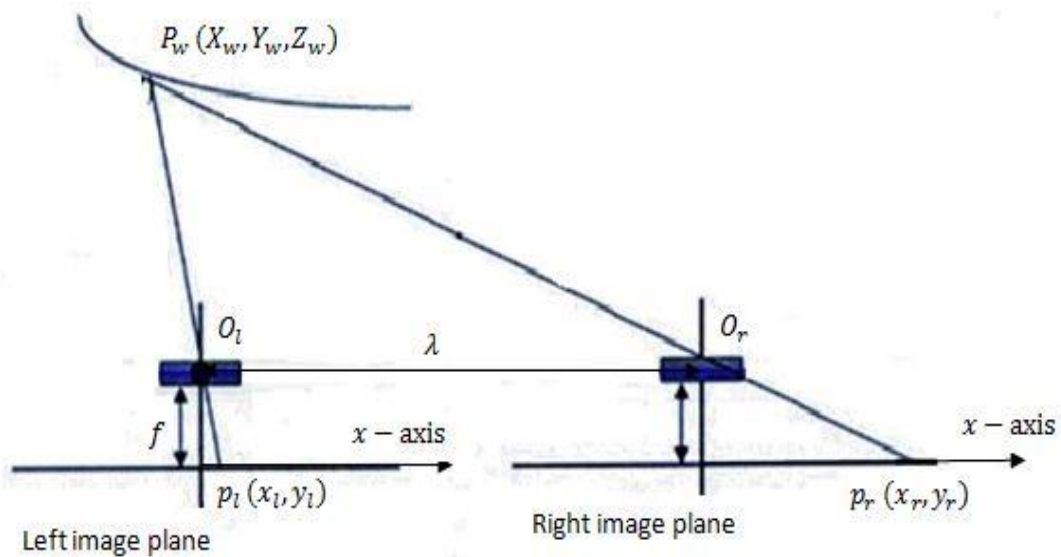


Fig. 4.2: The illustration of non-merged geometry of stereovision system

b) The merged geometry of stereovision system

Although most standard stereo image pairs or stereo image databases are taken carefully by arranging the cameras in non-merged geometry, it is hard to establish this stereo configuration in reality. We often, use the geometry of the merged stereo shown in Fig. 4.3. Under this configuration, the epipolar line is no longer a horizontal scan line.

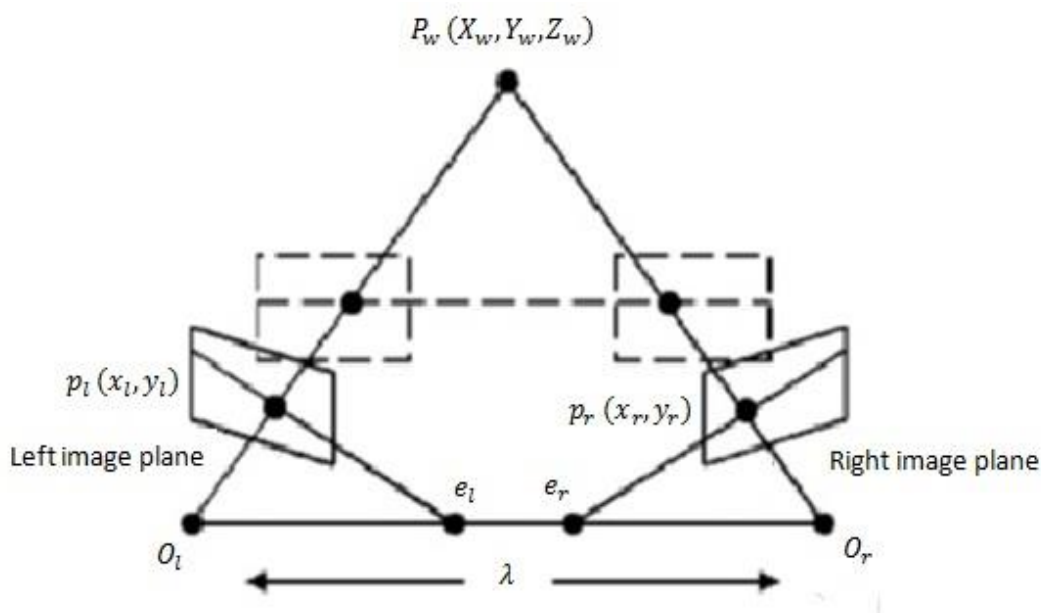


Fig. 4.3: The geometry of verged stereo with the epipolar line (solid) and the collinear scan-lines (dashed) after rectification

In order to have an efficient stereo correspondence algorithm, the epipolar line must be determined with high accuracy. One direct way to get the epipolar line is to compute the line equation without rectification (Fig. 4.3). Trifocal Tensor, presented by Lamdan et al. [91] and Cheong [92] is such an approach to determine the epipolar lines from camera geometry. Boyer et al. [93] also demonstrated a technique to compute the epipolar line from relative orientation without camera calibration. In this thesis, we propose a novel ray sketching based method to determine the epipolar line, which will be more accurate and straightforward.

4.2 Stereo correspondence by ray sketching based method

4.2.1 Theoretical basis of the novel ray sketching based method

The setup of the single-lens bi-prism stereovision system is shown schematically in Fig. 4.4 with the following descriptions:

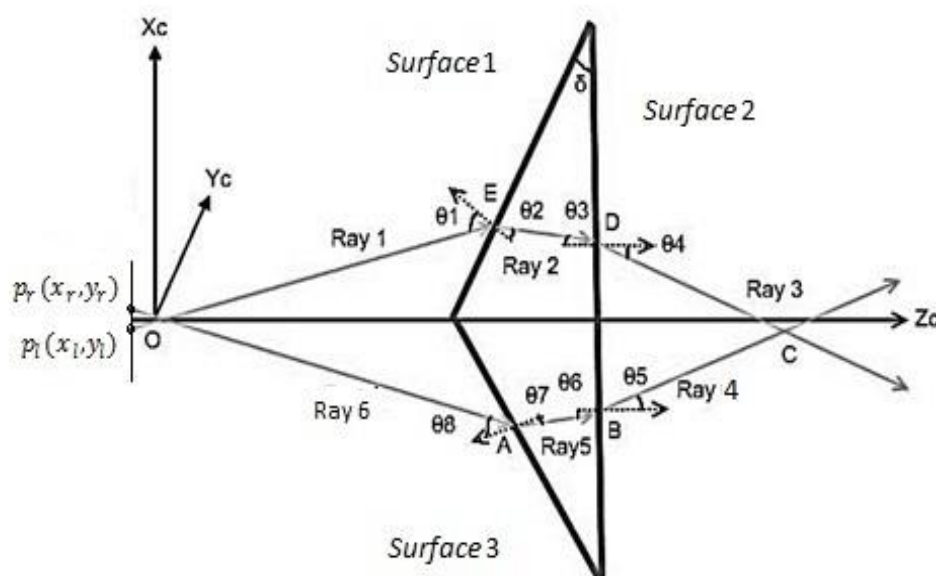


Fig. 4.4: Ray sketching based stereo correspondence configuration (top view)

- Surface 1, Surface 2, Surface 3* : the 3 surfaces of the bi-prism (top view);
- $O[X_c, Y_c, Z_c]$: Camera Coordinates, named as U ;
- (x_l, y_l) (x_r, y_r) : the corresponding left and right image coordinates appearing in the camera image plane;
- O : the camera optical center.

In the current work, the following assumptions are made:

- 1) *Surface 3* of the bi-prism (Fig. 4.4) is parallel to the image plane of the camera;
- 2) the apex line of the prism is perpendicular to the camera optical axis;
- 3) the centre point of the prism is the intersection point of the apex line and the optical axis;
- 4) the refraction of light is perfect, i.e., the diffusion of light is not considered;
- 5) the property of the optical prism is uniformly consistent.

To illustrate the basis of the ray sketching based approach, we shall consider an arbitrary point C in 3D space, as shown in Fig. 4.4. The point C will appear on the two-

halves of the image planes as two correspondence points $p_l(x_l, y_l)$ and $p_r(x_r, y_r)$. It is clear that if we trace the rays originating from two (stereo) correspondence points on the image plane, the rays, each after two-refraction through the prism, will intersect at the corresponding 3D point C (Ray 3). In other words, there is clearly some relationship between the two correspondence points on the image plane with a unique, particular 3D point in the scene. It is not difficult to establish this relationship, if the geometrical setup of the system and the optical properties of the prism are known. In the latter section, we shall attempt to obtain this relationship.

Generally speaking, we can say that the ray sketching based approach is basically to mathematically describe all the pertinent points, line, planes/surfaces, and their relationship, of a single-lens prism based stereovision system with respect to the Camera Coordinates U . And thence to construct the epipolar geometry, thereby to facilitate the subsequent stereo correspondence and depth recovery/3D reconstruction works.

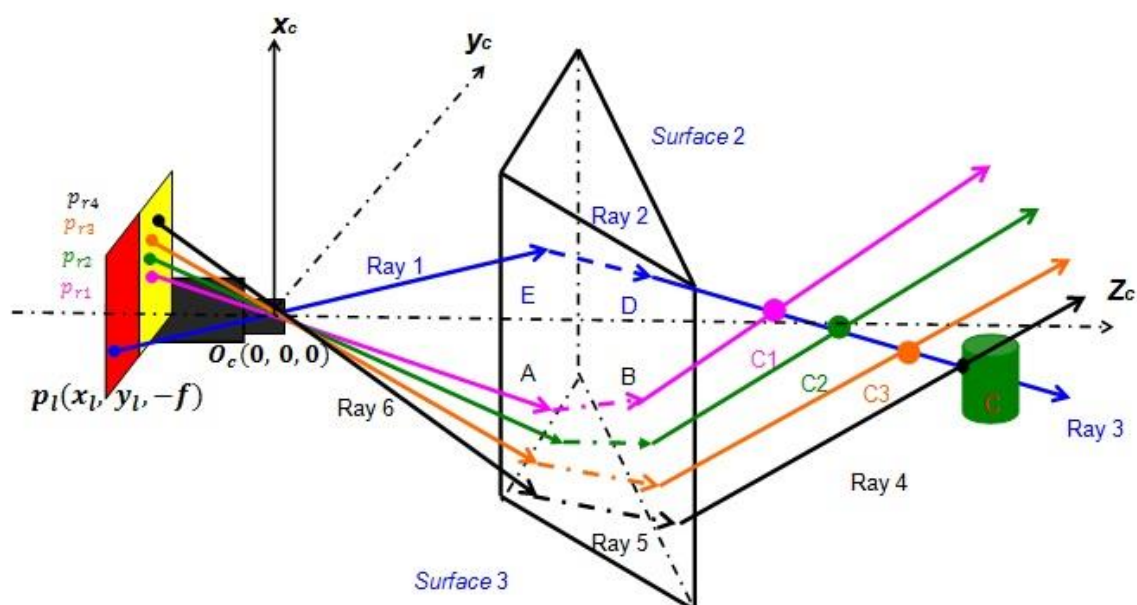


Fig. 4.5: Demonstration of ray sketching based approach (isometric view)

In Fig. 4.5, a set of points, such as C_1 , and C_2 on Ray 3, will appear as (x_{r1}, y_{r1}) and (x_{r2}, y_{r2}) , respectively, in the right image plane of the real camera. These image points (x_{r1}, y_{r1}) and (x_{r2}, y_{r2}) are naturally the stereo correspondence point candidates for the left image point (x_l, y_l) . This is because these corresponding 3D points (C_1 and C_2) are all located on somewhere along the refracted light ray from the left image point (x_l, y_l) (Ray 3). In this connection, this set of points on the right image plane can be used to construct the epipolar line based on the definition of the epipolar geometry. The correspondence point (x_r, y_r) of the left image (x_l, y_l) should lie on the epipolar line if error due to the system noise is assumed to be negligible. For simplicity and clarity, the above discussion only shows the formation of the epipolar line using only two points on the right image plane. In practice, increasing the number of right image points will improve the accuracy of the epipolar line.

4.2.2 Stereo correspondence by ray sketching based approach

Fig. 4.6 shows the system setup in the isometric view based on which we shall proceed to describe the theoretical development of our proposed ray sketching based method. The known system parameters are listed below.

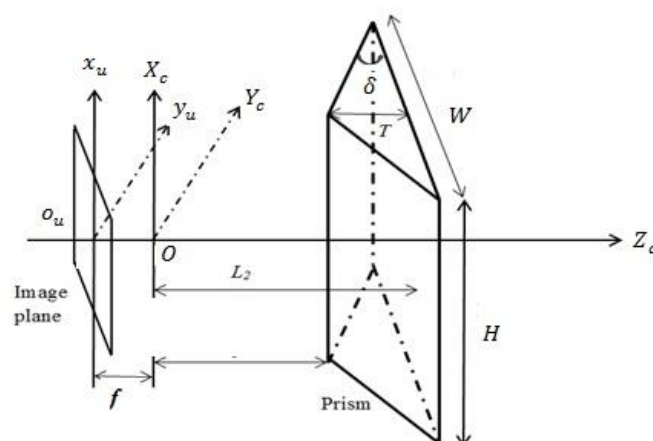


Fig. 4.6: Parameters of the bi-prism single-lens stereovision system (isometric view)

f : the focal length of the camera;

L_1 : the shortest distance from optical center of camera to the apex line of the prism;

δ : the corner angle of the prism (manufacturer's data);

W : the width of the prism;

T : the thickness of the prism;

H : the height of the prism;

n_r : the refraction index of the optic material of the prism;

L_2 : the shortest distance from optical center of camera to the prism back plane,

which is the sum of T and L_1 .

According to the epipolar geometry, any known left image point $p_l(x_l, y_l)$ has its corresponding right image point $p_r(x_r, y_r)$ locating on the right epipolar line, and vice versa. We shall discuss the construction of the right epipolar line with the given left image point in this thesis. Furthermore, to describe all the pertinent points, lines, planes/surfaces, and their relationship of our stereovision system for the solution of the stereo correspondence, the Snell's Law and some knowledge of 3D geometrical analyses (Appendix A) are employed in our development. The stereo correspondence by the ray sketching based approach can be divided into three steps.

Step 1: Forward ray sketching with an arbitrary left image point

Forward ray sketching is divided into three sub-steps: a) determination of Ray 1; b) determination of Ray 2 and c) determination of Ray 3.

a) Determination of Ray 1 in Camera Coordinates U

In the Camera Coordinates U , Ray 1 can be represented by two points: the arbitrary left image point $p_l(x_l, y_l, -f)$ and the camera optical center $O(0, 0, 0)$. The image point, P_l are represented on the Screen Coordinates (X_{lf}, Y_{lf}) . By applying Eq. (3.5),

$$x_l = (X_{lf} - C_x) \frac{d_x N_{cx}}{s_x N_{fx}}, \quad (4.1)$$

$$y_l = (Y_{lf} - C_y) \frac{d_y N_{cy}}{N_{fy}}. \quad (4.2)$$

Substituting x_l and y_l into the general straight line expression (c.f. Eq. (A.1)), Ray 1 in U is given below:

$$\frac{X}{-x_l} = \frac{Y}{-y_l} = \frac{Z}{f} = k_1. \quad (4.3)$$

Where k_1 is a scale factor and the direction vector of Ray1 is shown below.

$$R_{Ray1}^U = \begin{bmatrix} -x_l \\ -y_l \\ f \end{bmatrix}$$

R_{Ray1}^U is the direction vector of Ray 1 with respect to the Camera Coordinates U .

Therefore, for any arbitrary image point $p_l(x_l, y_l, -f)$ on the left image, Ray 1 is uniquely determined.

b) Determination of Ray 2 in Camera Coordinates U

In order to obtain Ray 2, the intersection between Ray 1 and *Surface* 1 of the bi-prism, point E in Fig. 4.7, must first be determined. Ray 2 will be represented by the direction

vector $R_{Ray2}^U = \begin{bmatrix} R_{ray2x}^U \\ R_{ray2y}^U \\ R_{ray2z}^U \end{bmatrix}$ in the form of Eq. (A.2) as:

$$\frac{X - E_x}{R_{ray2x}^U} = \frac{Y - E_y}{R_{ray2y}^U} = \frac{Z - E_z}{R_{ray2z}^U} = k_2. \quad (4.4)$$

Note that the subscripts $()_x$, $()_y$ and $()_z$ indicate the x -, y -, and z - components, respectively, of the variable in question.

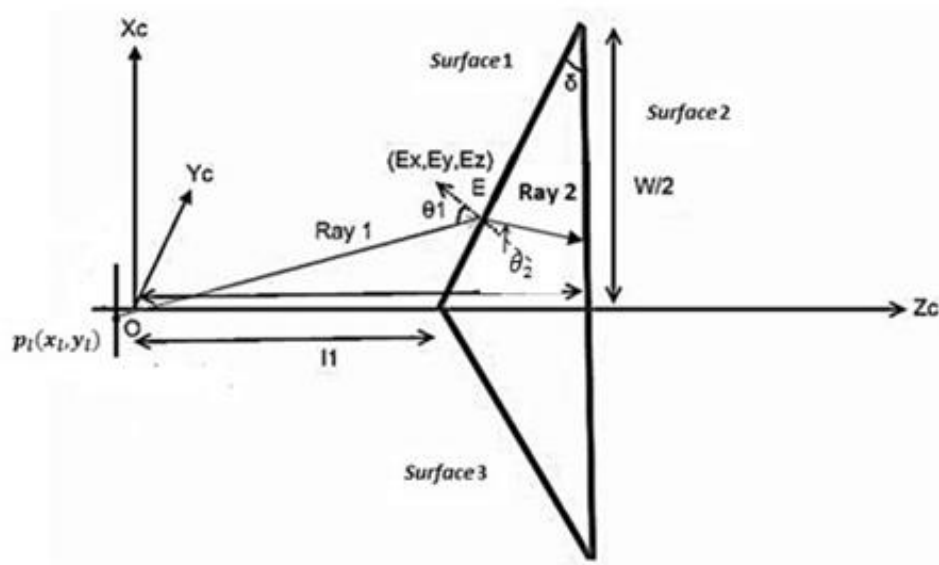


Fig. 4.7: Derivation of the intersection point E (top view)

As shown in Fig. 4.7, point E lies on the *Surface 1*. We can write the equation of *Surface 1* as: (c.f. Eq. (A.4))

$$R_{Surface1_x}^U (X - X_0) + R_{Surface1_y}^U (Y - Y_0) + R_{Surface1_z}^U (Z - Z_0) = 0 . \quad (4.5)$$

Where $R_{Surface1_x}^U$, $R_{Surface1_y}^U$ and $R_{Surface1_z}^U$ are the (x, y, z) components of the normal vector of *Surface 1*, respectively. (X_0, Y_0, Z_0) is an arbitrary point on the *Surface 1*.

There are three known information that can be used to derive $R_{Surface1}^U$.

- (1) Center point of the apex line of the prism with the coordinates of $(0, 0, L_1)$ lies on the *Surface 1*.

$$R_{Surface1_x}^U X + R_{Surface1_y}^U Y + R_{Surface1_z}^U (Z - L_1) = 0 . \quad (4.6)$$

- (2) *Surface 1* subtends an angle δ with *Surface 2* of the prism. Note that the normal vector of *Surface 2* is equal to $(0, 0, 1)^T$. Referring to Eq. (A.5), we have:

$$\cos \delta = \frac{R_{Surface1_z}^U}{\sqrt{R_{Surface1_x}^U{}^2 + R_{Surface1_y}^U{}^2 + R_{Surface1_z}^U{}^2}} \quad (4.7)$$

(3) *Surface 1* is parallel to the Y_c -axis of the Camera Coordinates U , thus,

$$R_{Surface1_y}^U = 0 \quad (4.8)$$

By substituting the solution from Eq. (4.6) ~ Eq. (4.8) into Eq. (4.5), *Surface 1* is given by:

$$X - \cot \delta * Z + \cot \delta * L_1 = 0, \quad (4.9)$$

and its normal vector as

$$R_{surface1}^U = \begin{bmatrix} 1 \\ 0 \\ -\cot \delta \end{bmatrix}.$$

With the known Ray 1 and *Surface 1*, we can then solve the 3D coordinates of the intersection point E as:

$$E = [Ex, Ey, Ez] = \left[\frac{-\cot \delta \cdot L_1 \cdot x_l}{f \cdot \cot \delta + x_l}, \frac{-\cot \delta \cdot L_1 \cdot y_l}{f \cdot \cot \delta + x_l}, \frac{\cot \delta \cdot L_1 \cdot f}{f \cdot \cot \delta + x_l} \right].$$

Referring to Fig. 4.7, the angle θ_1 is the angle between Ray 1 and the normal vector of *Surface 1*. And angle θ_2 is the angle between Ray 2 and the normal vector of *Surface 1*. They fulfil: (c.f. (Eq. (A.3)))

$$\cos \theta_1 = \frac{\cot \delta \cdot f + x_l}{\sqrt{1 + \cot^2 \delta} * \sqrt{x_l^2 + y_l^2 + f^2}}, \quad (4.10)$$

$$\cos \theta_2 = \frac{-R_{ray2x}^U + \cot \delta \cdot R_{ray2z}^U}{\sqrt{1 + \cot^2 \delta} * \sqrt{R_{ray2x}^U{}^2 + R_{ray2y}^U{}^2 + R_{ray2z}^U{}^2}} \quad (4.11)$$

Based on the Snell's Law and 3D geometrical analyses, Ray 1, normal vector of *Surface 1*, and Ray 2 must be coplanar. Eq. (4.12) and (4.13) are then established as:

$$\begin{aligned} \cos(\theta_1 - \theta_2) &= \cos \theta_1 * \cos \theta_2 + \sin^2 \theta_1 / n_r \\ &= \frac{-x_l * R_{ray2x}^U - y_l * R_{ray2y}^U + f * R_{ray2z}^U}{\sqrt{x_l^2 + y_l^2 + f^2} * \sqrt{R_{ray2x}^U{}^2 + R_{ray2y}^U{}^2 + R_{ray2z}^U{}^2}}, \end{aligned} \quad (4.12)$$

$$\begin{vmatrix} E_x & E_y & E_z \\ -x_l & -y_l & f \\ R_{ray2x}^U & R_{ray2y}^U & R_{ray2z}^U \end{vmatrix} = 0 \quad . \quad (4.13)$$

The relationship between θ_1 and θ_2 can be written as follows,

$$\sin \theta_1 / \sin \theta_2 = n_r \Rightarrow \frac{1 - \cos^2 \theta_1}{1 - \cos^2 \theta_2} = n_r^2 \quad . \quad (4.14)$$

As the parameters $x_l, y_l, f, E_x, E_y, E_z, \delta$ and n_r are known, the direction vector of Ray 2 can be uniquely determined by solving Eq. (4.13) and (4.14). It can be expressed as the function of (x_l, y_l) as follows:

$$R_{ray2x}^U = f_1(x_l, y_l),$$

$$R_{ray2y}^U = f_2(x_l, y_l),$$

$$R_{ray2z}^U = f_3(x_l, y_l).$$

Thus, Ray 2 can then be represented as:

$$\frac{X + \frac{\cot \delta \cdot L_1 \cdot x_l}{f \cdot \cot \delta + x_l}}{f_1(x_l, y_l)} = \frac{Y + \frac{\cot \delta \cdot L_1 \cdot y_l}{f \cdot \cot \delta + x_l}}{f_2(x_l, y_l)} = \frac{Z - \frac{\cot \delta \cdot L_1 \cdot f}{f \cdot \cot \delta + x_l}}{f_3(x_l, y_l)} = k_2.$$

c) Determination of Ray 3 in Camera Coordinates U

In Fig. 4.8, the refracted Ray 3, leaving *Surface 2* at point $D(D_x, D_y, D_z)$ and going towards the 3D point $P(X_w, Y_w, Z_w)$, forms the angle θ_4 with the normal of *Surface 2*. We

can express Ray 3 in the following form with its direction vector $R_{Ray3}^U = \begin{bmatrix} R_{ray3x}^U \\ R_{ray3y}^U \\ R_{ray3z}^U \end{bmatrix}$ and scale

factor k_3 as: (c.f. Eq. (A.2))

$$\frac{X - D_x}{R_{ray3x}^U} = \frac{Y - D_y}{R_{ray3y}^U} = \frac{Z - D_z}{R_{ray3z}^U} = k_3. \quad (4.15)$$

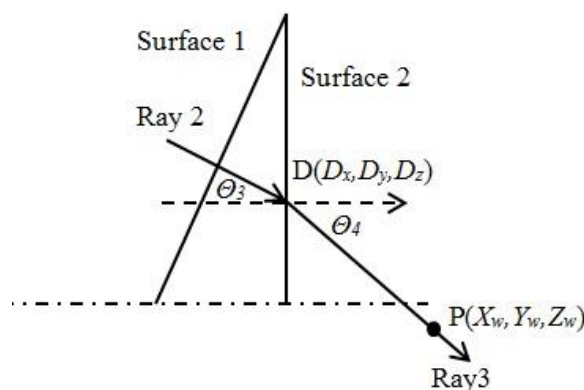


Fig. 4.8: Illustration of Ray 3 derivation (top view)

Point D is the intersection between Ray 2 and *Surface 2* and it can be solved by constructing the following two equations:

$$\frac{X + \frac{\cot \delta \cdot L_1 \cdot x_l}{f \cdot \cot \delta + x_l}}{f_1(x_l, y_l)} = \frac{Y + \frac{\cot \delta \cdot L_1 \cdot y_l}{f \cdot \cot \delta + x_l}}{f_2(x_l, y_l)} = \frac{Z - \frac{\cot \delta \cdot L_1 \cdot f}{f \cdot \cot \delta + x_l}}{f_3(x_l, y_l)} = k_2, \quad (4.16)$$

$$Z = L_2. \quad (4.17)$$

Hence,

$$D = \left[\frac{f_1(x_l, y_l) \cdot (L_2 - E_z)}{f_3(x_l, y_l)} + E_x, \frac{f_2(x_l, y_l) \cdot (L_2 - E_z)}{f_3(x_l, y_l)} + E_y, L_2 \right].$$

Writing the expression in more details,

$$D = \left[\frac{f_1(x_l, y_l) \cdot (L_2 - \frac{\cot \delta \cdot L_1 \cdot f}{f \cdot \cot \delta + x_l})}{f_3(x_l, y_l)} + \frac{\cot \delta \cdot L_1 \cdot x_l}{f \cdot \cot \delta + x_l}, \frac{f_2(x_l, y_l) \cdot (L_2 - \frac{\cot \delta \cdot L_1 \cdot f}{f \cdot \cot \delta + x_l})}{f_3(x_l, y_l)} + \frac{\cot \delta \cdot L_1 \cdot y_l}{f \cdot \cot \delta + x_l}, L_2 \right].$$

Referring to Fig. 4.8, the angles θ_3 and θ_4 are the angles between Ray 2 and Ray 3 with the normal of *Surface 2*, respectively. As the normal vector of the *Surface 2* is set to be $(0, 0, 1)^T$, angle θ_3 and θ_4 can be written as: (c.f. Eq. (A.3))

$$\cos \theta_3 = \frac{f_3(x_l, y_l)}{\sqrt{f_1(x_l, y_l)^2 + f_2(x_l, y_l)^2 + f_3(x_l, y_l)^2}}, \quad (4.18)$$

$$\cos \theta_4 = \frac{R_{ray3z}^U}{\sqrt{R_{ray3x}^U{}^2 + R_{ray3y}^U{}^2 + R_{ray3z}^U{}^2}}. \quad (4.19)$$

Similarly, Ray 3 is then derived through the Snell's Law and 3D geometrical analyses. Based on Snell's Law, Ray 2, the normal vector of *Surface 2*, and Ray 3 are coplanar. Thus, Eq. (4.20) and (4.21) can be generated:

$$\begin{aligned} \cos(\theta_4 - \theta_3) &= \cos \theta_4 * \cos \theta_3 + \sin^2 \theta_3 * n_r \\ &= \frac{f_1(x_l, y_l) * R_{ray3x}^U - f_2(x_l, y_l) * R_{ray3y}^U + f_3(x_l, y_l) * R_{ray3z}^U}{\sqrt{f_1^2(x_l, y_l) + f_2^2(x_l, y_l) + f_3^2(x_l, y_l)} * \sqrt{R_{ray3x}^U{}^2 + R_{ray3y}^U{}^2 + R_{ray3z}^U{}^2}}, \end{aligned} \quad (4.20)$$

$$\begin{vmatrix} D_x - E_x & D_y - E_y & L_2 - E_z \\ f_1(x_l, y_l) & f_2(x_l, y_l) & f_3(x_l, y_l) \\ R_{ray3x}^U & R_{ray3y}^U & R_{ray3z}^U \end{vmatrix} = 0. \quad (4.21)$$

Meanwhile, θ_3 and θ_4 are related according to Snell's Law,

$$\sin \theta_4 / \sin \theta_3 = n_r \Rightarrow \frac{1 - \cos^2 \theta_4}{1 - \cos^2 \theta_3} = n_r^2. \quad (4.22)$$

We can calculate the direction vector of Ray 3 with the known parameters of

$x_l, y_l, f, E_x, E_y, E_z, \delta, n_r, D_x, D_y, D_z, L_1$ and L_2 , whereby $E_x, E_y, E_z, D_x, D_y, D_z$ can be

uniquely determined by the left image point $p_l(x_l, y_l)$ and f, δ, n_r, L_1, L_2 are the known system parameters. Therefore, with any image point $p_l(x_l, y_l)$ in the left image taken by the bi-prism stereovision system, the direction vector of Ray 3 can be expressed as:

$$R_{ray3x}^U = g_1(x_l, y_l),$$

$$R_{ray3y}^U = g_2(x_l, y_l),$$

$$R_{ray3z}^U = g_3(x_l, y_l).$$

Ray 3 can thus be written as:

$$\frac{X - D_x}{g_1(x_l, y_l)} = \frac{Y - D_y}{g_2(x_l, y_l)} = \frac{Z - D_z}{g_3(x_l, y_l)} = k_3. \quad (4.23)$$

Hence, any 3D point $P (X_w, Y_w, Z_w)$ on the Ray 3 can be represented by the following equation:

$$\frac{X_w - D_x}{g_1(x_l, y_l)} = \frac{Y_w - D_y}{g_2(x_l, y_l)} = \frac{Z_w - L_2}{g_3(x_l, y_l)} = k_3. \quad (4.24)$$

Where k_3 is the scale factor within the range of $[-\infty, \infty]$. Any specific value of k_3 uniquely corresponds to a specific 3D point on Ray 3.

In summary, all the forward light rays (Rays 1, 2 and 3) are all directly attainable by using Snell's Law and 3D geometrical analyses with the known left image point $p_l(x_l, y_l)$ and the system parameters f, δ, n_r, L_1, L_2 . The experimental results, which will be shown in Chapter V, verify the above development.

Step 2: Backward ray sketching through Spatial Coordinate transformation

Step 2 computes three rays (Rays 4, 5 and 6) in the backward path as shown in Fig. 4.4. Similarly, we shall divide it into three sub-steps: a) determination of Ray 4; b) determination of Ray 5 and c) determination of Ray 6.

During the process of determining the forward rays (Rays 1, 2 and 3), we have two kinds of parameters: direct parameters $x_l, y_l, f, \delta, n_r, L_1, L_2$ and indirect parameters $E_x, E_y, E_z, D_x, D_y, D_z$, which are determined by the coordinates of $P_l(x_l, y_l)$. Although most of the equations to solve the indirect parameters and direction vectors of the light rays are nonlinear in nature, it is not difficult to obtain the numerical solutions with the known values of x_l and y_l .

However, in the case of the backward light rays (Rays 4, 5 and 6), their determination would be more complex even though we know the coordinates of the left image point (x_l, y_l) (i.e., Ray 3 is known) and the world point P on Ray 3 (with a chosen value of the scale factor k_3). The main reason is that we do not know the direction vector of Ray 4. While, in contrast, the determination of the direction vector of Ray 1 is straightforward. It can be determined by the coordinates of the left image points and the camera optical center easily. Referring to Fig. 4.4, the intersection point A , the direction vectors of Rays 4, 5 and 6 could be expressed in terms of the coordinates of point $B (B_x, B_y, L_2)$. However, the relevant equations will later be shown, to be highly nonlinear, which impose challenges for the initial value selection when solving the problems iteratively. Thanks to the unique geometrical properties of the single-lens bi-prism based stereovision system, we propose to use the Spatial Coordinate transformation approach to derive the direction vectors of all the backward rays. The details are as follows.

a) Determination of Ray 4 in Camera Coordinates U

As shown in Fig. 4.4, $C (C_x, C_y, C_z)$ is an arbitrary point that lies on Ray 3. In order to make sure that C is certain to have a corresponding right image point, it must fulfil the condition of $C_z > L_2$. Otherwise, C will be out of the two camera's field of view and no

corresponding image points are projected onto the image plane. The coordinates of the intersection point C can be derived from the Eq. (4.15) as:

$$C_z \in (L_2, +\infty) , \quad (4.25)$$

$$C_x = (C_z - L_2) * \frac{R_{ray3x}^U}{R_{ray3z}^U} + D_x , \quad (4.26)$$

$$C_y = (C_z - L_2) * \frac{R_{ray3y}^U}{R_{ray3z}^U} + D_y . \quad (4.27)$$

We define the intersection point B between Ray 4 and the *Surface 2* of the prism with coordinates $[B_x, B_y, L_2]$. Thus, the unit direction vector of Ray 4 is:

$$R_{Ray4}^U = \begin{bmatrix} R_{ray4x}^U \\ R_{ray4y}^U \\ R_{ray4z}^U \end{bmatrix} = \begin{bmatrix} (C_x - B_x) / \gamma \\ (C_y - B_y) / \gamma \\ (C_z - L_2) / \gamma \end{bmatrix} . \quad (4.28)$$

Where γ is the norm of the direction vector of Ray 4,

$$\gamma = \|Ray4\| = \sqrt{(B_x - C_x)^2 + (B_y - C_y)^2 + (L_2 - C_z)^2} .$$

Therefore, the first backward ray of Ray 4 can be expressed as:

$$\frac{X - C_x}{(C_x - B_x) / \gamma} = \frac{Y - C_y}{(C_y - B_y) / \gamma} = \frac{Z - C_z}{(C_z - L_2) / \gamma} = k_4 . \quad (4.29)$$

b) Determination of Ray 5 in Camera Coordinates U

Referring to Fig. 4.9, Ray 5 is determined by the intersection point B and its direction

vector in U ($R_{Ray5}^U = \begin{pmatrix} R_{ray5x}^U \\ R_{ray5y}^U \\ R_{ray5z}^U \end{pmatrix}$) as: (c.f. Eq. (A.2))

$$\frac{X - B_x}{R_{ray5x}^U} = \frac{Y - B_y}{R_{ray5y}^U} = \frac{Z - B_z}{R_{ray5z}^U} = k_5 .$$

Where $B_z = L_2$ and k_5 is the scale factor.

In order to obtain the direction of the Ray 5 in the Camera Coordinates U , we would apply the concept of the Spatial Coordinate transformation (c.f. Apex A) and carry out the following procedures:

- (i) To determine the direction of the Ray 5 in the local coordinates H ;
- (ii) To obtain the rotational relationship between H and U ;
- (iii) To derive the direction vector of the Ray 5 in U by using the relationship obtained in step (ii).

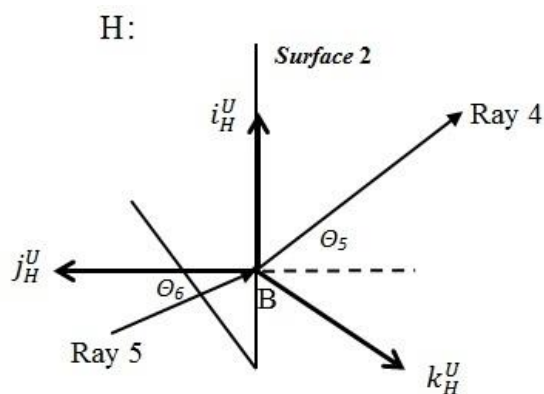


Fig. 4.9: Local Coordinates H attached at point B related with Ray 4 and Ray 5 (top view)

As shown in Fig. 4.9, a Local Coordinates H is attached to the intersection point B . Its orientation in the Camera Coordinates U can be written in the form of:

$$R_H^U = (i_H^U, j_H^U, k_H^U).$$

Where i_H^U, j_H^U and k_H^U are the directional vectors of the x, y and z axes of the local coordinates H expressed in U . We set the normal vector of *Surface 2* as the y -axis of H

and it is given by $j_H^U = \begin{bmatrix} j_{Hx}^U \\ j_{Hy}^U \\ j_{Hz}^U \end{bmatrix}$.

j_{Hx}^U and j_{Hy}^U are null vectors as j_H^U is parallel to the z -axis of U . Thus, we can write

j_H^U as:

$$j_H^U = \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}. \quad (4.30)$$

And meanwhile we set:

$$k_H^U = \begin{bmatrix} k_{Hx}^U \\ k_{Hy}^U \\ k_{Hz}^U \end{bmatrix} = j_H^U \times R_{Ray4}^U = \begin{bmatrix} (C_y - B_y)/\gamma \\ (B_x - C_x)/\gamma \\ 0 \end{bmatrix}, \quad (4.31)$$

$$i_H^U = \begin{bmatrix} i_{Hx}^U \\ i_{Hy}^U \\ i_{Hz}^U \end{bmatrix} = j_H^U \times k_H^U = \begin{bmatrix} (B_x - C_x)/\gamma \\ (B_y - C_y)/\gamma \\ 0 \end{bmatrix}, \quad (4.32)$$

Therefore,

$$R_H^U = \begin{bmatrix} (B_x - C_x)/\gamma & 0 & (C_y - B_y)/\gamma \\ (B_y - C_y)/\gamma & 0 & (B_x - C_x)/\gamma \\ 0 & -1 & 0 \end{bmatrix}.$$

In Fig. 4.9, θ_5 and θ_6 are the angles between Ray 4, Ray 5 and $-j_H^U$, respectively.

They fulfil: (c.f. Eq. (A.3))

$$\begin{aligned} \theta_5 &= \cos^{-1} \frac{R_{ray4x}^U \cdot j_{Hx}^U + R_{ray4y}^U \cdot j_{Hy}^U + R_{ray4z}^U \cdot j_{Hz}^U}{\sqrt{R_{ray4x}^U{}^2 + R_{ray4y}^U{}^2 + R_{ray4z}^U{}^2} \cdot \sqrt{j_{Hx}^U{}^2 + j_{Hy}^U{}^2 + j_{Hz}^U{}^2}}, \\ &= \cos^{-1}(-R_{ray4z}^U) \\ &= \cos^{-1}((L_2 - C_z)/\gamma), \end{aligned} \quad (4.33)$$

$$\theta_6 = \sin^{-1} \frac{\sin \theta_5}{n_r}. \quad (4.34)$$

Note that once a point C on the Ray 3 has been chosen, the values of C_x, C_y and C_z are known and therefore θ_5 is only determined by γ ($\gamma = \sqrt{(B_x - C_x)^2 + (B_y - C_y)^2 + (L_2 - C_z)^2}$).

In other words, θ_5 is determined by B_x and B_y .

The orientation of the Ray 5 in the local coordinates H can be expressed as:

$$R_{Ray5}^H = \begin{bmatrix} \sin \theta_6 \\ -\cos \theta_6 \\ 0 \end{bmatrix}. \quad (4.35)$$

Applying the Spatial Coordinates transformation, the orientation of the Ray 5 in Camera Coordinates U is determined as:

$$\begin{aligned} R_{Ray5}^U &= R_H^U \cdot R_{Ray5}^H = \begin{bmatrix} i_{Hx}^U & j_{Hx}^U & k_{Hx}^U \\ i_{Hy}^U & j_{Hy}^U & k_{Hy}^U \\ i_{Hz}^U & j_{Hz}^U & k_{Hz}^U \end{bmatrix} \begin{bmatrix} \sin \theta_6 \\ -\cos \theta_6 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} i_{Hx}^U \sin \theta_6 - j_{Hx}^U \cos \theta_6 \\ i_{Hy}^U \sin \theta_6 - j_{Hy}^U \cos \theta_6 \\ i_{Hz}^U \sin \theta_6 - j_{Hz}^U \cos \theta_6 \end{bmatrix} \\ &= \begin{bmatrix} (B_x - C_x) \sin \theta_6 / \gamma \\ (B_y - C_y) \sin \theta_6 / \gamma \\ \cos \theta_6 \end{bmatrix}. \end{aligned} \quad (4.36)$$

In summary, the equation representing Ray 5 is given below: (c.f. Eq. (A.2))

$$\frac{X - B_x}{(B_x - C_x) \sin \theta_6 / \gamma} = \frac{Y - B_y}{(B_y - C_y) \sin \theta_6 / \gamma} = \frac{Z - L_2}{\cos \theta_6} = k_5. \quad (4.37)$$

And meanwhile, Ray 5 can also be expressed by the two intersection points A and B as: (c.f. Eq. (A.1))

$$\frac{X - B_x}{B_x - A_x} = \frac{Y - B_y}{B_y - A_y} = \frac{Z - B_z}{(B_z - A_z)}. \quad (4.38)$$

c) Determination of Ray 6 in Camera Coordinates U

As shown in Fig. 4.4, A is the intersection point between *Surface 3* of the bi-prism and Ray 5. The coordinates of point $A (A_x, A_y, A_z)$ can be derived through the equations stated below with the known corner angle δ of the prism.

Similar to the determination of the *Surface 1*, the angle between *Surface 3* and *Surface 2* is the corner angle of the prism - δ and thus can be directly obtained. (c.f. Eq. (4.5) ~ (4.8))

$$-(X + \frac{W}{2}) - \cot \delta (Z - L_2) = 0 \quad (4.39)$$

Where the point $(-\frac{W}{2}, 0, L_2)$ is the prism edge point located on *Surface 3* and $(-1, 0, -\cot \delta)$ is the normal vector of *Surface 3*.

Combining Eq. (4.37), (4.38) and (4.39) together, we can obtain the coordinates of point A as:

$$A_x = \frac{\frac{R_{ray5z}^U}{R_{ray5x}^U} \cdot B_x - \frac{W}{2} \tan \delta}{\tan \delta + \frac{R_{ray5z}^U}{R_{ray5x}^U}} = \frac{B_x \cdot \gamma \cdot \cot \theta_6 - \frac{W}{2} (B_x - C_x) \tan \delta}{\gamma \cot \theta_6 + (B_x - C_x) \tan \delta} \quad (4.40)$$

$$A_y = (A_x - B_x) \cdot \frac{R_{ray5y}^U}{R_{ray5x}^U} + B_y = \frac{(B_y - C_y) \cdot (A_x - B_x)}{(B_x - C_x)} + B_y \quad (4.41)$$

$$A_z = (A_x - B_x) \cdot \frac{R_{ray5z}^U}{R_{ray5x}^U} + B_z = \frac{\gamma \cot \theta_6 (A_x - B_x)}{(B_x - C_x)} + L_2 \quad (4.42)$$

To derive the direction vector of Ray 6, we follow similar steps as in the case of Ray 5. A local coordinates G is attached at the intersection point A as shown in Fig. 4.10. Its rotational relationship with respect to the Camera Coordinates U is $R_G^U = (i_G^U, j_G^U, k_G^U)$.

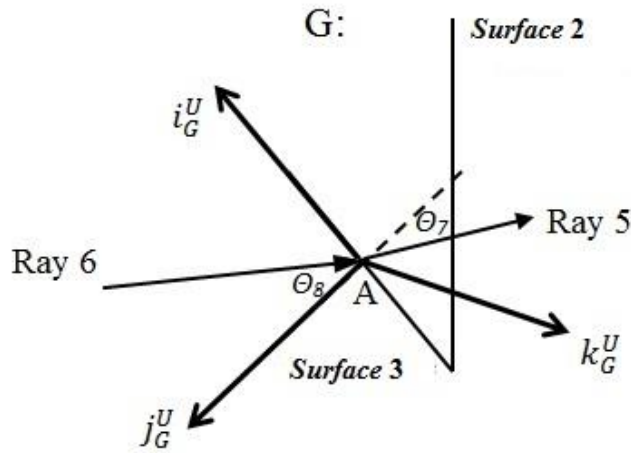


Fig. 4.10: Local Coordinate G attached at point A related with Ray 5 and Ray 6 (top view)

The normal to the *Surface 3* of the bi-prism is set as the y -axis of G , the origin of

which is attached to the point A . The y -axis of the G in U is expressed as $j_G^U = \begin{bmatrix} j_{Gx}^U \\ j_{Gy}^U \\ j_{Gz}^U \end{bmatrix}$. The

y - component j_{Gy}^U of the j_G^U is equal to zero because the normal of the *Surface 3* is perpendicular to the y -axis of U . If j_{Gx}^U is set as -1 , j_{Gz}^U is then equal to $-\cot \delta$. Thus, the final form of j_G^U in the form of unit vector is

$$j_G^U = \begin{bmatrix} -\sin \delta \\ 0 \\ -\cos \delta \end{bmatrix}. \quad (4.43)$$

The x and z axes of local coordinates G are represented in U .

$$\begin{aligned} k_G^U = \begin{bmatrix} k_{Gx}^U \\ k_{Gy}^U \\ k_{Gz}^U \end{bmatrix} &= R_{Ray5}^U \times j_G^U = \begin{bmatrix} (B_x - C_x) \sin \theta_6 / \gamma \\ (B_y - C_y) \sin \theta_6 / \gamma \\ \cos \theta_6 \end{bmatrix} \times \begin{bmatrix} -\sin \delta \\ 0 \\ -\cos \delta \end{bmatrix} \\ &= \begin{bmatrix} (C_y - B_y) \sin \theta_6 \cos \delta / \gamma \\ (B_x - C_x) \sin \theta_6 \cos \delta / \gamma - \cos \theta_6 \sin \delta \\ (B_y - C_y) \sin \theta_6 \sin \delta / \gamma \end{bmatrix}, \end{aligned} \quad (4.44)$$

$$i_G^U = \begin{bmatrix} i_{Gx}^U \\ i_{Gy}^U \\ i_{Gz}^U \end{bmatrix} = j_G^U \times k_G^U = \begin{bmatrix} \frac{(B_x - C_x) \sin \theta_6 \cos^2 \delta}{\gamma} - \cos \theta_6 \sin 2\delta / 2 \\ \frac{\sin \theta_6 (C_y - B_y)}{\gamma} \\ \frac{\sin \theta_6 \sin 2\delta (C_x - B_x)}{2\gamma} + \cos \theta_6 \sin^2 \delta \end{bmatrix}. \quad (4.45)$$

Putting Eq. (4.43), (4.44) and (4.45) together, the rotational matrix of G with respect to U is

$$R_G^U = \begin{bmatrix} i_{Gx}^U & j_{Gx}^U & k_{Gx}^U \\ i_{Gy}^U & j_{Gy}^U & k_{Gy}^U \\ i_{Gz}^U & j_{Gz}^U & k_{Gz}^U \end{bmatrix} = \begin{bmatrix} \frac{2(B_x - C_x) \sin \theta_6 \cos^2 \delta - \cos \theta_6 \sin 2\delta}{2\gamma} & -\sin \delta & (C_y - B_y) \sin \theta_6 \cos \delta / \gamma \\ \frac{\sin \theta_6 (C_y - B_y)}{\gamma} & 0 & (B_x - C_x) \sin \theta_6 \cos \delta / \gamma - \cos \theta_6 \sin \delta \\ \frac{\sin \theta_6 \sin 2\delta (C_x - B_x) + 2\gamma \cos \theta_6 \sin^2 \delta}{2\gamma} & -\cos \delta & (B_y - C_y) \sin \theta_6 \sin \delta / \gamma \end{bmatrix}. \quad (4.46)$$

From Fig. 4.10, we can observe that θ_7 is the angle between Ray 5 and $-j_G^U$. It can be expressed as:

$$\begin{aligned} \theta_7 &= \cos^{-1} \frac{R_{ray5x}^U \cdot j_{Gx}^U + R_{ray5y}^U \cdot j_{Gy}^U + R_{ray5z}^U \cdot j_{Gz}^U}{\sqrt{R_{ray5x}^U{}^2 + R_{ray5y}^U{}^2 + R_{ray5z}^U{}^2} \cdot \sqrt{j_{Gx}^U{}^2 + j_{Gy}^U{}^2 + j_{Gz}^U{}^2}} \\ &= \cos^{-1} \frac{(C_x - B_x) \sin \theta_6 \sin \delta - \cos \theta_6 \cos \delta}{\sqrt{1 - (L_2 - C_z)^2 \sin^2 \theta_6 / \gamma^2}}. \end{aligned} \quad (4.47)$$

From Snell's Law,

$$\theta_8 = \sin^{-1}(n_r \sin \theta_7). \quad (4.48)$$

Ray 6 is located in the plane determined by x and y axes of the local coordinate G .

Its orientation in G (refer to Fig. 4.10) is thus given by:

$$R_{Ray6}^G = \begin{bmatrix} -\sin \theta_8 \\ -\cos \theta_8 \\ 0 \end{bmatrix}. \quad (4.49)$$

By applying Spatial Coordinate Transformation on Eq. (4.46) and (4.49), the direction vector of Ray 6 expressed in U is then given below.

$$\begin{aligned} R_{Ray6}^U &= R_G^U \cdot R_{Ray6}^G = \begin{bmatrix} i_{Gx}^U & j_{Gx}^U & k_{Gx}^U \\ i_{Gy}^U & j_{Gy}^U & k_{Gy}^U \\ i_{Gz}^U & j_{Gz}^U & k_{Gz}^U \end{bmatrix} \cdot \begin{bmatrix} -\sin \theta_8 \\ -\cos \theta_8 \\ 0 \end{bmatrix} \\ &= \begin{bmatrix} -i_{Gx}^U \sin \theta_8 - j_{Gx}^U \cos \theta_8 \\ -i_{Gy}^U \sin \theta_8 - j_{Gy}^U \cos \theta_8 \\ -i_{Gz}^U \sin \theta_8 - j_{Gz}^U \cos \theta_8 \end{bmatrix}. \end{aligned} \quad (4.50)$$

The Ray 6 is therefore expressed as:

$$\frac{A_x}{R_{ray6x}^U} = \frac{A_y}{R_{ray6y}^U} = \frac{A_z}{R_{ray6z}^U} = k_6. \quad (4.51)$$

Where k_6 is a scale factor.

The coordinates of the right image point $p_r(x_r, y_r, -f)$ to 3D object point C on the Ray 4 are given as:

$$\begin{aligned} z_r &= -f \\ x_r &= \frac{R_{ray6x}^U}{R_{ray6z}^U} \cdot (-f) \\ y_r &= \frac{R_{ray6y}^U}{R_{ray6z}^U} \cdot (-f) \end{aligned} \quad (4.52)$$

We can then determine the Screen Coordinates of the right image point using the following transformations, (c. f. Eq. (3.5))

$$X_{rf} = x_r \frac{s_x \cdot N_{fx}}{d_x \cdot N_{cx}} + C_x, \quad (4.53)$$

$$Y_{rf} = y_r \frac{N_{fx}}{d_y \cdot N_{cy}} + c_y. \quad (4.54)$$

Step 3: Stereo corresponding search

With forward and backward ray sketching based approach, we are certain that with any known arbitrary point in the left half of the real camera image plane, we can determine an unique refracted ray (Ray 3 in Fig. 4.4). From this refracted ray leaving *Surface 3*, we can obtain an unique correspondence point on the right half of the real camera image plane. Eq. (4.25) to (4.54) developed above would enable us to determine the coordinates of this correspondence point. However, this is only possible when the intersection point B is known. This is because we would need to know the coordinates of point B to proceed to determine Rays 4, 5 and 6, and thence the correspondence point in the right image. In practice, it is not possible to fix the coordinates of point B . However, it is possible to set the range of the components of the coordinates of the point B .

$$-\frac{H}{2} \leq B_x \leq \frac{H}{2} \quad : \text{ where } H \text{ is the height of the bi-prism;}$$

$$0 \leq B_y \leq \frac{W}{2} \quad : \text{ where } W \text{ is the width of the bi-prism;}$$

$$B_z = L_2 \quad : \text{ which is determined by system setup and thickness of the bi-prism.}$$

In this research, iterative algorithms will be implemented to find the correspondence candidate right image points (x_r, y_r) . The iterative process will be carried out on B_x and B_y with appropriate initial values and intervals. The terminating condition for the iterative

process is when the values of B_x and B_y lead Ray 6 to pass through the optical center of the real camera. That means Eq. (4.55) and (4.56) are fulfilled (c.f. Eq. (4.51)).

Mathematically,

$$\frac{A_x}{R_{ray6x}^U} = \frac{A_y}{R_{ray6y}^U} = \frac{A_z}{R_{ray6z}^U} = k_6 ,$$

$$\text{Threshold 1} = \left| \frac{\frac{A_y}{R_{Ray6y}^U} - \frac{A_x}{R_{Ray6x}^U}}{\sqrt{\frac{A_y^2}{R_{Ray6y}^U} + \frac{A_x^2}{R_{Ray6x}^U}}} \right| = 0 , \quad (4.55)$$

$$\text{Threshold 2} = \left| \frac{\frac{A_z}{R_{Ray6z}^U} - \frac{A_x}{R_{Ray6x}^U}}{\sqrt{\frac{A_z^2}{R_{Ray6z}^U} + \frac{A_x^2}{R_{Ray6x}^U}}} \right| = 0 . \quad (4.56)$$

In practice, it would not be possible to satisfy the conditions (Eq. (4.55) and (4.56)) due to the existence of the errors in the system, such as lens distortion and digitizing errors. Therefore, the values of the thresholds are set to some small values (determined experimentally). With a threshold value of 0.0005, our iterative algorithm took approximately 3 seconds to offer a converged solution, whereas with the threshold value sets to 0.0001, 10 seconds were needed. The lower value 0.0001 requires significantly longer computational time, but yields results which differ from that of 0.0005 by less than 0.1%. Hence, 0.0005 is chosen as the threshold value in our process to determine the coordinates of the point in the right-half image plane of the real camera.

At the juncture, it is appropriate to re-iterate the ray sketching based method for stereo correspondence. Referring to Fig. 4.4, with a known image point $p_l(x_l, y_l, -f)$ on the left half image plane of the real camera, we can determine Rays 1, 2 and 3. Ray 3 is the

“output” ray in this forward ray sketching process. The corresponding 3D point in the scene should lie on Ray 3 in this case. Supposing that the point C in Fig. 4.4 is indeed the point of interest, by the backward ray sketching process described above, we would be able to obtain Rays 6, 5 and 4. With the knowledge of Ray 6, we can then determine the stereo correspondence point $p_r(x_r, y_r, -f)$ in the right half image plane of the real camera. The knowledge of these two corresponding image points will allow us to determine the depth of the point C in the Camera Coordinates U .

Based on the above discussion, the real 3D point of interest must lie on Ray 3. Therefore, any point on Ray 3 would be a probable candidate stereo correspondence point of $p_l(x_l, y_l, -f)$. In this regards, we can pick several points on Ray 3, and then by using the backward ray sketching process, we would be able to obtain a series of stereo correspondence points $p_{ri}(x_{ri}, y_{ri}, -f) \ i = 1, 2, \dots, n$. A line can then be constructed using these series of $p_{ri}(x_{ri}, y_{ri}, -f)$ by the least square method. This line is the epipolar line indeed, corresponding to a given $p_l(x_l, y_l, -f)$. In other words, the search of the stereo correspondence of a given $p_l(x_l, y_l, -f)$ can be done along the epipolar line thus formed.

Once the right epipolar line is determined through the ray sketching based approach with an arbitrary given left image point, the corresponding right image point is obtainable through searching on the 2D epipolar line rather than the whole right image. We employ the local correlation-based algorithm for the searching process. The searching process is illustrated in detail as follows.

Fig. 4.11(a) shows the left image with an image point $P_{lf}(X_{lf}, Y_{lf})$. Fig. 4.11(b) shows the right image with the epipolar line obtained above. Note that there are candidate correspondence image points on the epipolar lines. The search region $R(P_l)$ is confined to the epipolar line.

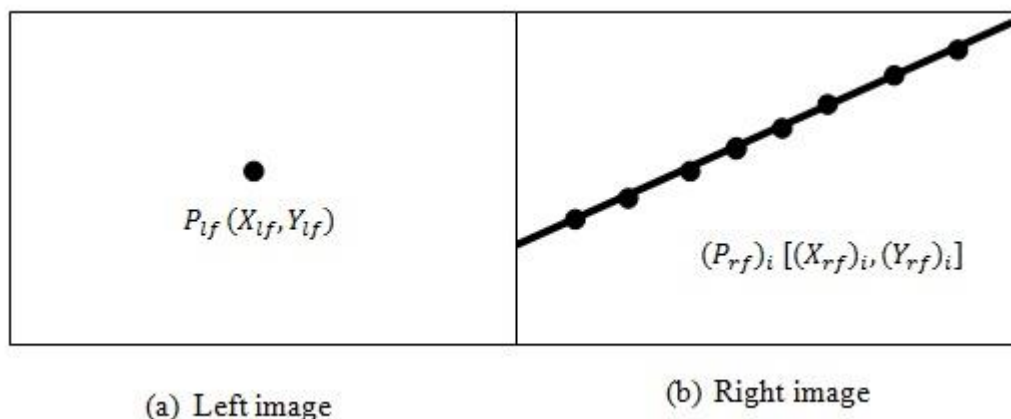


Fig. 4.11: Illustration of the stereo corresponding search

The correlation is evaluated by defining a window, the center of which is the point of interest. The size of the windows is $(2W + 1)$ by $(2W + 1)$, $W = 1, 2, \dots$. All the calculations are confined to the pixels within the windows. The procedure is illustrated below.

For a given $P_{lf}(X_{lf}, Y_{lf})$, if there are N numbers of candidate correspondence points in the right image, or $(P_{rf})_i = [(X_{rf})_i, (Y_{rf})_i]$ for $i = 1, 2, \dots, N$.

For $i = 1, 2, \dots, N$

The disparity: $(d)_i = (dx_i, dy_i)$

Where $dx_i = (X_{rf})_i - X_{lf}$, $dy_i = (Y_{rf})_i - Y_{lf}$

The correlation is evaluated as follows:

$$C(d_i) = \sum_{m=-W}^W \sum_{n=-W}^W \Psi \{I_l(X_{lf} + m, Y_{lf} + n), I_r(X_{lf} + dx_i + m, Y_{lf} + dy_i + n)\}. \quad (4.57)$$

Where I_l and I_r are the intensities of the left and right stereo image pair, respectively.

$$\psi(u, v) = -(u - v)^2. \quad (4.58)$$

Here u and v are the intensities of the pixels in the left and right images, respectively.

The correct corresponding point is one when $C(d)_i$ is the maximum.

If, $i = c$, $C(d)_i$ is the maximum, the corresponding right image point is:

$$[(X_{lf} + dx_c), (Y_{lf} + dy_c)].$$

4.3 Depth recovery of the single-lens bi-prism based stereovision system

The depth of a scene can be recovered by using a stereovision system if two or multiple images of the same scene are taken from two or more different camera viewpoints. The significance of the stereo correspondence algorithm in Section 4.2 is to find the correspondence points from the captured images of a 3D scene, which is projected onto the cameras. If such correspondence is identified for each pixel in the images, the depth of the points in 3D space can be recovered. In this section, 3D depth recovery by using the triangulation method is introduced. Section 4.3.1 introduces the triangulation method of general stereo image pairs. Section 4.3.2 combines the ray sketching based approach with the triangulation method to solve the depth recovery problem of the single-lens prism based stereovision system.

4.3.1 Triangulation of general stereo image pairs

Referring to Fig. 4.12, the left camera coordinate frame is chosen to be the world coordinates. The point P' is projected to a pair of corresponding points p_l and p_r on the two image planes, lying at the intersection of the two rays from O_l through p_l and from O_r through p_r . In other words, the coordinates of point P' can be computed by finding the intersection of these two rays. However, the two rays are usually two skew lines in 3D space, which may not intersect with each other at a physical point. Therefore, the mid-point theorem is employed to estimate the coordinates of point P' (see Fig. 4.12).

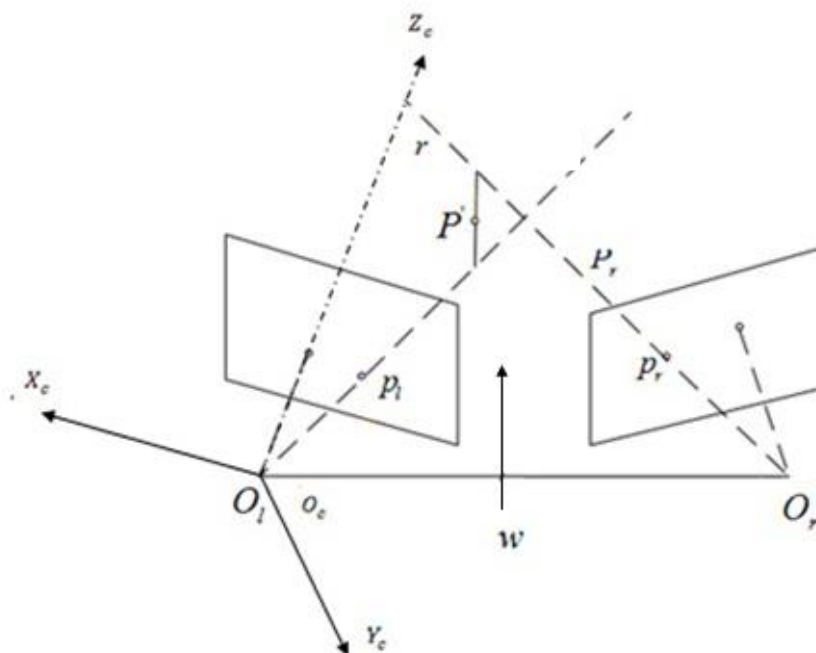


Fig. 4.12: Triangulation with nonintersecting

Set the Left Camera Coordinates as World Coordinates. Let ap_l ($a \in \mathbb{R}$) be the Ray l through O_l and p_l . Let $T + bR^T p_r$ be the Ray r through O_r and p_r , expressed in the left Camera Coordinates (which is taken to be the world coordinates). R and T denote the rotational and translational transformation matrices of the right camera image with respect to the left camera image, respectively. Let w be a vector orthogonal to both l and r . Our problem reduced to determine the midpoint P' of the segment ab , which joins l and r . (ab is parallel to w)

Using the midpoint theorem, the endpoints of the segment, $a_0 p_l$ and $T + b_0 R^T p_r$, can be first computed by solving the linear system of Eq. (4.59):

$$ap_l + c(p_l \times R^T p_r) = T + bR^T p_r. \quad (4.59)$$

Where a_0 , b_0 and c_0 are the solutions for a , b and c of Eq. (4.59), respectively. The two endpoints of the segment ($a_0 p_l$ and $T + b_0 R^T p_r$) can be computed from a_0 and b_0 . Once

the two endpoints are computed, the midpoint P' can be obtained by computing the average coordinates of the two endpoints.

4.3.2 Triangulation of single-lens bi-prism based stereovision system

Our proposed ray sketching based approach and the triangulation method discussed are combining to solve the depth recovery problem of the single-lens bi-prism based stereovision system. The same steps discussed in the forward ray sketching (the detailed derivations are presented in Section 4.2.2) are used to compute Rays 4, 5 and 6 in order. Theoretically, Rays 3 and 4 should intersect with each other and the intersection point is the 3D point we are looking for. However, due to noises, hardware imperfections, Rays 3 and 4 are often two skew lines and do not intersect at one point in the space. Thus, we will employ the midpoint theorem in Section 4.3.1 to compute the point P , which is located at the midpoint of the shortest segment between Ray 3 and Ray 4 as shown in Fig 4.13. The detail depth recovery results will be presented and discussed in Chapter V.

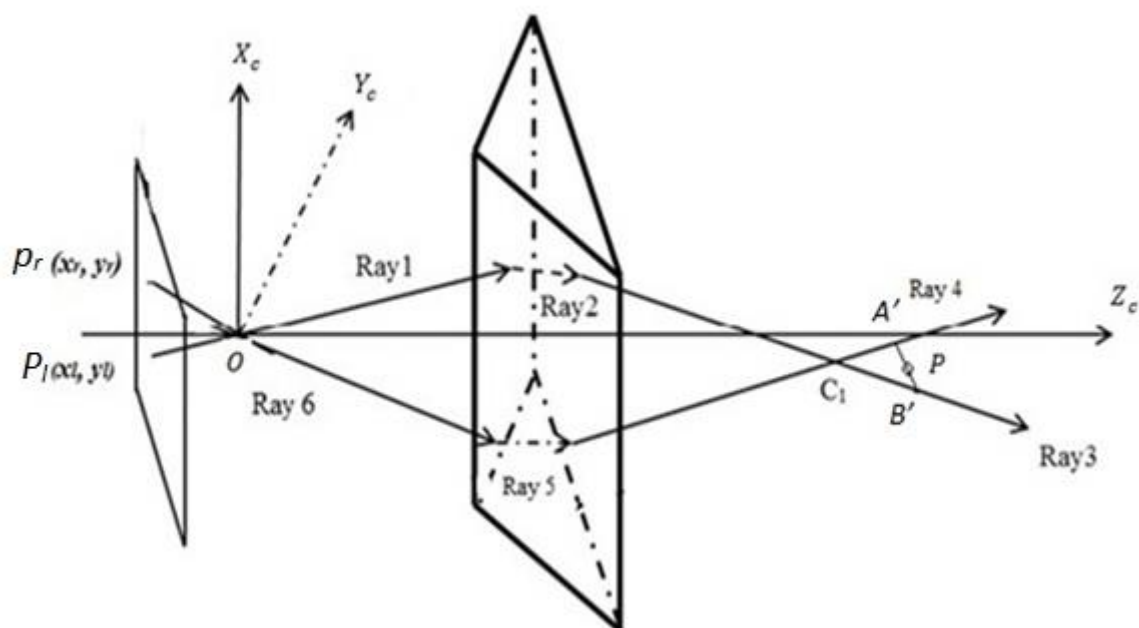


Fig 4.13: Object point determination when Ray 3 and Ray 4 are not intersected in space

4.4 Summary

We have presented the ray sketching based method, which is capable of solving the stereo correspondence and depth recovery of the single-lens bi-prism based stereovision system. It is a camera calibration-free approach, thereby reduces the fastidious calibration process compared to the camera calibration based method. Moreover, by landing on the uniquely geometrical properties of the prism-based imaging system, all the development work done on the binocular stereovision system can be easily generalized to tri-ocular and even multi-ocular stereovision systems. The experimental results and performance of the novel ray sketching based approach will be demonstrated in Chapter V.

Once we have a known left image point, the expressions for the forward light Rays 1, 2 and 3 in the 3D space can be determined. Then we can obtain the corresponding right image point using the algorithm presented in Section 4.2.

With a known corresponding right image point, we apply the same steps discussed in the forward ray sketching to obtain Rays 4, 5 and 6. Theoretically, Rays 3 and 4 should intersect with each other and the intersection point is the 3D point we are looking for. However, due to the noises and setup errors, Rays 3 and 4 do not intersect at one point. Thus, we use the conventional midpoint theorem to determine the 3D point by locating the midpoint of the shortest segment between the Rays 3 and 4 (Fig. 4.13).

Chapter V Experiment and experimental results

5.1 Setup of the single-lens prism based stereovision system

Fig. 5.1 shows the experimental setup of our stereovision system, which includes:



Fig. 5.1: Experimental setup of the single-lens prism based stereovision system

- A tailor-made mechanical stand: 1) vernier calipers are installed in X , Y , and Z directions; and 2) a stage that can rotate in 360° , holds camera, prism and laser distance meter (Fig. 5.2);



Rotational stage



Vernier caliper

Fig. 5.2: Vernier calipers and rotational stage

- Custom made optical prism: it is mounted in front of the camera with its backplane made to be parallel to the camera image plane. Its center point lies on the camera optical axis. The refraction index is 1.48 and the corner angle is 21.8° . One of the prisms used in our system is shown in Fig. 5.3.

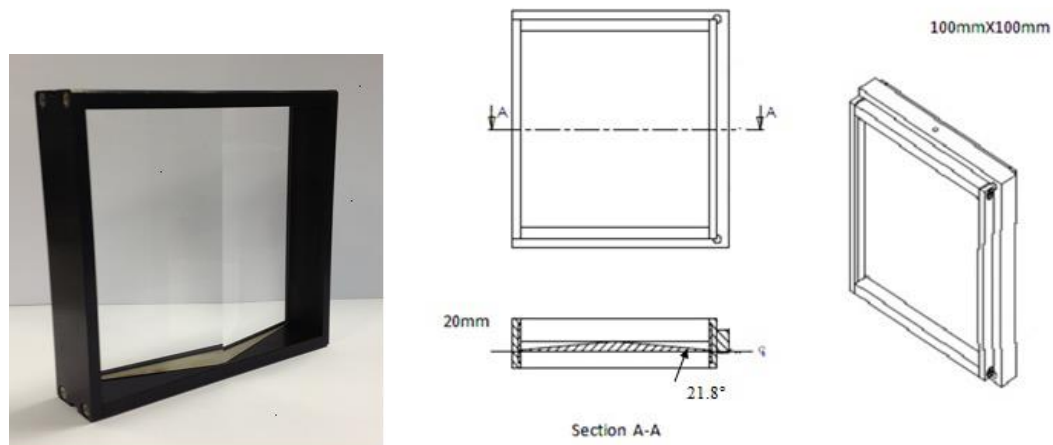


Fig. 5.3: Optical bi-prism used for our experiment

- JAI-CV-M9CL CCD Camera: it is set behind the optical prism. The distance between the camera and the prism can be adjusted on the mechanical stand. The specifications of the camera are provided in Table 5.1. (c.f. (Eq. (3.5) for the definition of the parameters used)

Table 5.1: Specification of the JAI CV-M9CL camera

Camera parameters	JAI-CV-M9CL
N_{fx} (pixels)	1024
N_{cx} (pixels)	768
Focal length f (mm)	8
d_x (mm)	0.00465
d_y (mm)	0.00465
Centre of the computer image (C_x, C_y)	(512, 384)

- The calibration board: a high-precision calibration board was manufactured with black-dot pattern as shown in Fig. 5.4. The diameter of the dots is 20mm. The center-to-center distance in the horizontal and vertical directions is 25mm;

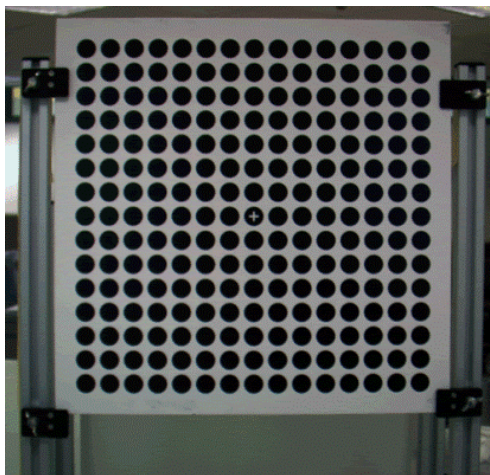


Fig. 5.4: Customized calibration board

- Leica DISTO D5: a laser distance measurement system with an accuracy of +/- 1.5mm. It is fixed on the mechanical stand with its origin aligns with the camera optical centre;
- Software: Matlab 7.0.

5.2 Experimental results by camera calibration based approach

5.2.1 Experimental procedures of the camera calibration based approach

The procedures to calibrate the real camera and two virtual cameras are as follows.

- 1) Fix the JAI-CV-M9CL camera and Leica DISTO D5 distance meter at vernier calliper's 60mm reading positions on the mechanical stand;
- 2) Place the 7 by 15 dotted pattern of calibration board (Fig. 5.5) at position A ($d_1 = 1000mm$) as shown in Fig. 5.6. The calibration board is then moved and rotated at 5 random orientations around position A. Record all the coordinate values of the 105

(7×15) image points. Throughout the whole experiments, the centre of the circle at top-left corner in Fig. 5.5 is set as the origin (0, 0, 0) of the World Coordinates;

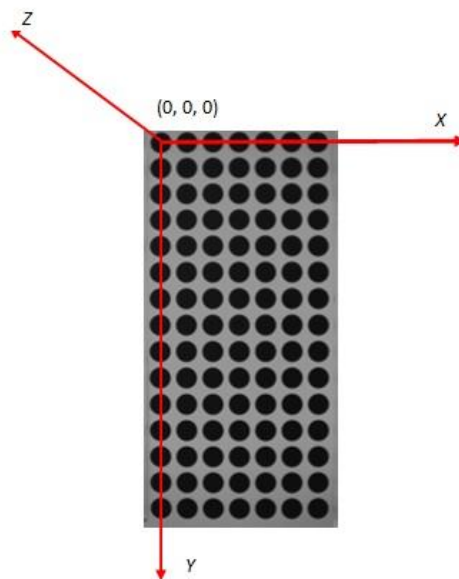


Fig. 5.5: The 7 by 15 calibration pattern for camera calibration

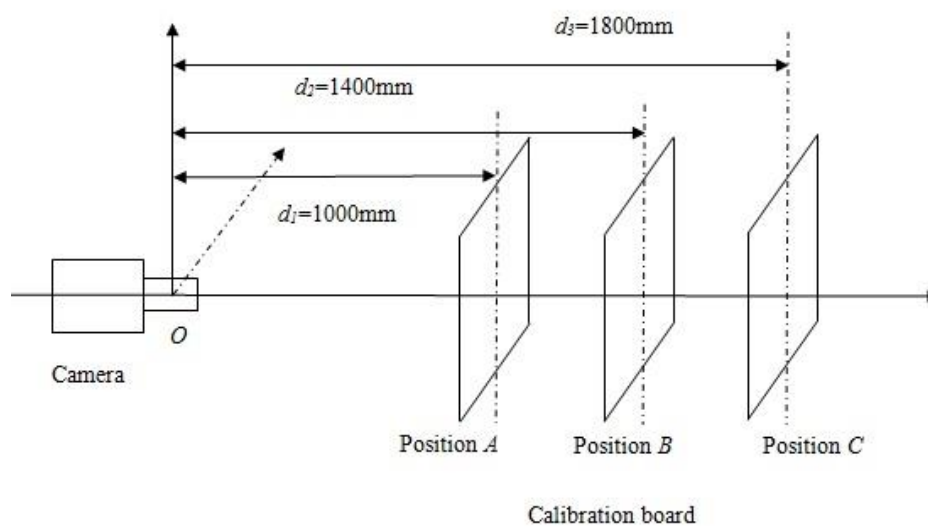


Fig. 5.6: The setup of the positions of calibration board in the experiment

- 3) The minimum number of the data points to be chosen is 7, as there are 13 unknown parameters to be calibrated. (c. f. Section 3.1). By using the recorded coordinate

values in step 2), the intrinsic parameters ($f, s_x, d_x^{-1}, d_y^{-1}, C_x$ and C_y) and extrinsic parameters (R, T) of the real camera are computed;

- 4) Mount the bi-prism vertically in front of the camera (the distance between camera optical centre and apex line of the prism is 170mm). The centre point of the prism must lie on the real camera's optical axis (the z -axis). Steps 2) and 3) are repeated and 105 image points are recorded to calibrate both the left and right virtual cameras; (c.f. Eq. (3.21) and Eq. (3.22))
- 5) Place the calibration board at positions B ($d_2=1400\text{mm}$) and C ($d_3=1800\text{mm}$). Repeat the steps 2) to 4) to obtain the calibration of both the real camera and virtual cameras.

5.2.2 Results of stereo correspondence by camera calibration based approach

By following the experimental procedures in section 5.2.1, the camera calibration results at position A are shown.

Refer to Eq. (3.10) and Eq. (3.11), for the real camera, we have:

$$\begin{bmatrix} X_f k \\ Y_f k \\ k \end{bmatrix} = M \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} fd_x^{-1} & 0 & C_x \\ 0 & fd_y^{-1} & C_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & T_x \\ r_{2,1} & r_{2,2} & r_{2,3} & T_y \\ r_{3,1} & r_{3,2} & r_{3,3} & T_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (5.1)$$

Substituting the recorded coordinates values from the step 2) of the experiment into Eq. (3.18), the intrinsic and extrinsic parameters at position A can be recovered.

Overall, for the real camera, we have the calibration results as:

$$\begin{bmatrix} X_f k \\ Y_f k \\ k \end{bmatrix} = \begin{bmatrix} 1972 & 0 & 284 & 54143 \\ -7 & 1808 & 38 & 53788 \\ 0 & 0 & 0 & 967 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (5.2)$$

The similar procedures are repeated for the left and right virtual camera with the calibration board at position A.

For the left virtual camera:

$$\begin{bmatrix} X_{lf}k_l \\ Y_{lf}k_l \\ k_l \end{bmatrix} = M^L \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_l d'_{xl}{}^{-1} & 0 & C_{xl} \\ 0 & f_l d'_{yl}{}^{-1} & C_{yl} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & T_x \\ r_{2,1} & r_{2,2} & r_{2,3} & T_y \\ r_{3,1} & r_{3,2} & r_{3,3} & T_z \end{bmatrix}^L \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (5.3)$$

Substituting all the calibrated parameters, we have:

$$\begin{bmatrix} X_{lf}k_l \\ Y_{lf}k_l \\ k_l \end{bmatrix} = \begin{bmatrix} 1976 & 0 & 261 & 54242 \\ -58 & 1800 & 388 & 53922 \\ 0 & 0 & 0 & 967 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (5.4)$$

For the right virtual camera:

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = M^R \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} f_r d'_{xr}{}^{-1} & 0 & C_{xr} \\ 0 & f_r d'_{yr}{}^{-1} & C_{yr} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} & T_x \\ r_{2,1} & r_{2,2} & r_{2,3} & T_y \\ r_{3,1} & r_{3,2} & r_{3,3} & T_z \end{bmatrix}^R \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \quad (5.5)$$

And

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = \begin{bmatrix} 2000 & -10 & 370 & 129820 \\ 40 & 1780 & 390 & 43230 \\ 0 & 0 & 0 & 930 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}. \quad (5.6)$$

Once the left and right virtual cameras are known, the stereo correspondence between the left and right virtual cameras at position A is established in Eq. (5.7).

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = \begin{bmatrix} 1.0190 & -0.0012 & 77.1695 \\ 0.0478 & 0.9873 & -13.0337 \\ 0.0001 & 0 & 0.9573 \end{bmatrix} \begin{bmatrix} X_{lf}k_l \\ Y_{lf}k_l \\ k_l \end{bmatrix} \quad (5.7)$$

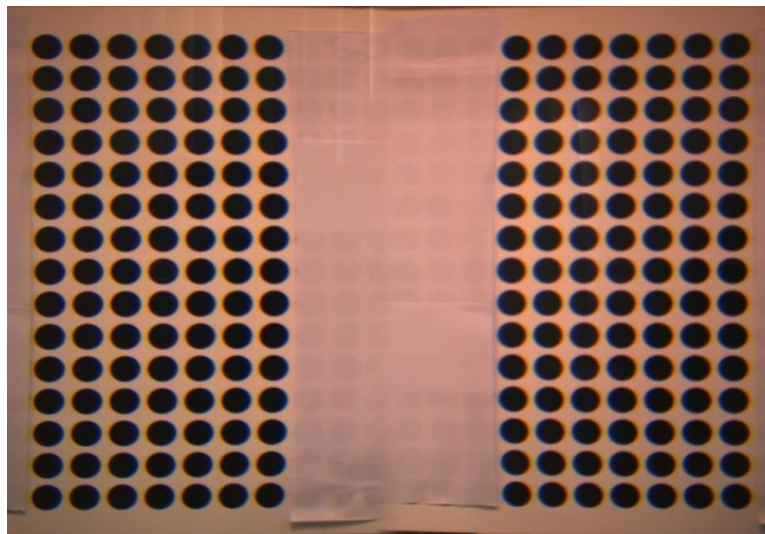


Fig. 5.7: Stereo image pair taken by our single-lens stereovision system at position A

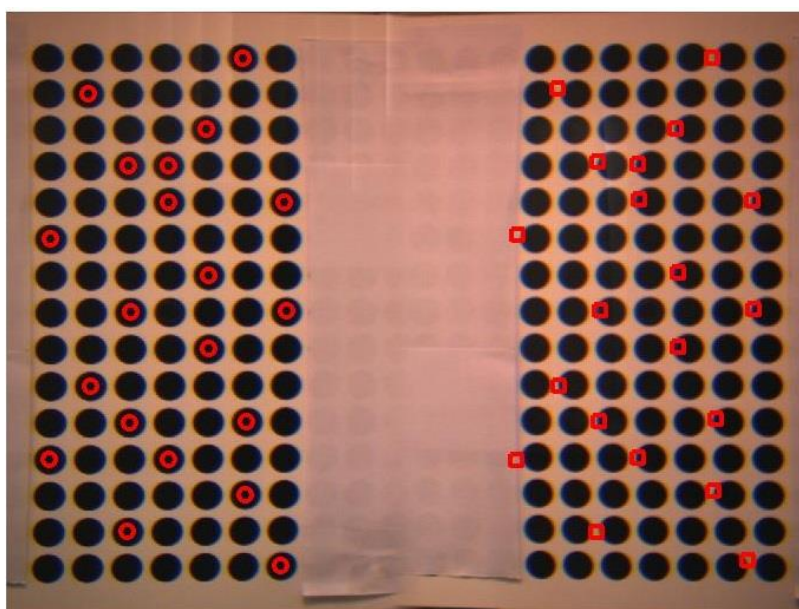


Fig. 5.8: Display of the stereo corresponding points (position A)

Fig. 5.7 is the stereo image pairs taken at position A ($d_l=1000\text{mm}$). We randomly select 20 dots and extract their centres as our left image point (marked by Red Cross). The

results of the corresponding points are obtained using Eq. (5.7) as shown in the Fig. 5.8 (marked by Red Square). Their values are presented in Table 5.2, which are compared with the real corresponding image points.

Table 5.2: Results of stereo correspondence at position A

	X_{lf}	Y_{lf}	X_{rf}	X_{rf}^c	$ X_{rf} - X_{rf}^c $	Y_{rf}	Y_{rf}^c	$ Y_{rf} - Y_{rf}^c $
1	301	59	920	898	22	57	60	3
2	104	104	724	701	23	104	98	6
3	254	150	868	851	17	149	149	0
4	156	196	770	753	17	195	192	3
5	206	196	818	804	14	195	194	1
6	207	242	818	805	13	241	240	1
7	354	242	966	949	17	241	241	0
8	56	289	672	651	21	288	285	3
9	257	335	866	854	12	334	333	1
10	158	382	768	755	13	381	380	1
11	356	381	966	950	16	381	379	2
12	257	429	866	854	12	428	427	1
13	107	476	720	703	17	474	476	2
14	157	523	769	754	15	521	522	1
15	306	522	917	902	15	521	519	2
16	55	569	674	649	15	566	571	5
17	206	569	820	803	17	567	568	1
18	304	614	919	900	19	614	610	4
19	154	661	773	751	22	659	662	3
20	349	705	971	943	28	707	699	8

X_{lf}, Y_{lf} : Coordinates of the randomly selected left image points on the screen;

X_{rf}, Y_{rf} : Coordinates the actual right corresponding image points on the screen;

X_{rf}^c, Y_{rf}^c : the calculated corresponding right image point coordinates based on the camera calibration parameters.

Same calibration procedures are repeated at positions B and C (Fig. 5.6). The relationship of the stereo correspondence between the left and right virtual cameras is obtained, respectively.

At position B ($d_2=1400\text{mm}$),

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = \begin{bmatrix} 1.0235 & -0.0014 & 108.6622 \\ 0.0433 & 0.9873 & -11.3399 \\ 0.0001 & 0 & 0.9616 \end{bmatrix} \begin{bmatrix} X_{lf}k_l \\ Y_{lf}k_l \\ k_l \end{bmatrix}. \quad (5.8)$$

At position C ($d_3=1800\text{mm}$),

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = \begin{bmatrix} 1.0103 & 0.0021 & 126.5333 \\ 0.0358 & 0.9881 & -9.8732 \\ 0.0001 & 0 & 0.9650 \end{bmatrix} \begin{bmatrix} X_{lf}k_l \\ Y_{lf}k_l \\ k_l \end{bmatrix}. \quad (5.9)$$

With the Eq. (5.8) and Eq. (5.9), the detail results of the stereo correspondence at positions B and C can be found in Appendix B .

5.2.3 Results of the depth recovery by camera calibration based approach

Referring to Section 3.3, the recovered depth information of the 20 points in the Camera Coordinates can be calculated with the knowledge of the results of the camera calibration and stereo correspondence (c.f. Eq. (3.33) and Eq. (3.34)). Again,

$$H = \begin{bmatrix} (X_{lf}r_{3,1} - \frac{f_l s_{xl}}{d'_{xl}}r_{1,1} - C_{xl}r_{3,1})_{LEFT} & (Y_{lf}r_{3,1} - \frac{f_l}{d'_{yl}}r_{2,1} - C_{yl}r_{3,1})_{LEFT} & (X_{lf}r_{3,3} - \frac{f_l s_{xl}}{d'_{xl}}r_{1,3} + C_{xl}r_{3,3})_{LEFT} \\ (Y_{lf}r_{3,1} - \frac{f_l}{d'_{yl}}r_{2,1} - C_{yl}r_{3,1})_{LEFT} & (Y_{lf}r_{3,2} - \frac{f_l}{d'_{yl}}r_{2,2} - C_{xl}r_{3,2})_{LEFT} & (Y_{lf}r_{3,3} - \frac{f_l}{d'_{yl}}r_{2,3} - C_{xl}r_{3,3})_{LEFT} \\ (X_{rf}r_{3,1} - \frac{f_r s_{xr}}{d'_{xr}}r_{1,1} - C_{xr}r_{3,1})_{RIGHT} & (Y_{rf}r_{3,1} - \frac{f_r}{d'_{yf}}r_{2,1} - C_{yf}r_{3,1})_{RIGHT} & (X_{rf}r_{3,3} - \frac{f_r s_{xr}}{d'_{xr}}r_{1,3} + C_{xr}r_{3,3})_{RIGHT} \\ (Y_{rf}r_{3,1} - \frac{f_r}{d'_{yr}}r_{2,1} - C_{yr}r_{3,1})_{RIGHT} & (Y_{rf}r_{3,2} - \frac{f_r}{d'_{yr}}r_{2,2} - C_{xr}r_{3,2})_{RIGHT} & (Y_{rf}r_{3,3} - \frac{f_r}{d'_{yf}}r_{2,3} - C_{yf}r_{3,3})_{RIGHT} \end{bmatrix},$$

$$G = \begin{bmatrix} (\frac{f_l s_{xl}}{d'_{xl}}T_x + C_{xl}T_z - X_{lf}T_z)_{LEFT} \\ (\frac{f_l}{d'_{yl}}T_y + C_{yl}T_z - Y_{lf}T_z)_{LEFT} \\ (\frac{f_r s_{xr}}{d'_{xr}}T_x + C_{xr}T_z - X_{rf}T_z)_{RIGHT} \\ (\frac{f_r s_{xr}}{d'_{xr}}T_x + C_{xr}T_z - X_{rf}T_z)_{RIGHT} \end{bmatrix}.$$

We have:

$$I = \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = (A^T * A)^{-1} A^T * B \quad . \quad (5.10)$$

All the parameters in the matrices H and G have been recovered in section 5.2.2. To be consistent, we set Z_w as the recovered depth in the World Coordinates and Z_c^c as the recovered depth in the Camera Coordinates. The unit for Z_w and Z_c^c is in mm. The recovered depth Z_c^c is: (c.f. Eq. (3.35))

$$Z_c^c = r_{3,1} * X_w + r_{3,2} * Y_w + r_{3,3} * Z_w + T_z, \quad (5.11)$$

Where $r_{3,1}, r_{3,2}, r_{3,3}, T_z$ are the real camera rotational and translational parameters.

Table 5.3 and Fig. 5.9 show the results of the actual depth recovery Z_c^c at position A.

Table 5.3: Results of depth recovery by calibration based approach at position A

	X_{lf}	Y_{lf}	X_{rf}^c	Y_{rf}^c	Z_c	Z_c^c	$ Z_c - Z_c^c $
1	301	59	898	57	1000	934	66
2	104	104	701	104	1000	994	6
3	254	150	851	149	1000	963	27
4	156	196	753	195	1000	955	45
5	206	196	804	195	1000	971	29
6	207	242	805	241	1000	969	31
7	354	242	949	241	1000	931	69
8	56	289	651	288	1000	1026	26
9	257	335	854	334	1000	985	15
10	158	382	755	381	1000	998	2
11	356	381	950	381	1000	927	73
12	257	429	854	428	1000	981	19
13	107	476	703	474	1000	979	21
14	157	523	754	521	1000	991	9
15	306	522	902	521	1000	948	52
16	55	569	649	566	1000	1012	12
17	206	569	803	567	1000	966	34
18	304	614	900	614	1000	945	55
19	154	661	751	659	1000	984	16
20	349	705	943	707	1000	915	85

Z_c : Actual depth in Camera Coordinates;

The error in percentage by camera calibration based approach at position A is:

$$\left(\frac{\sum |Z_c - Z_c^c|}{20} \right) / Z_c * 100\% = 3.46\%$$

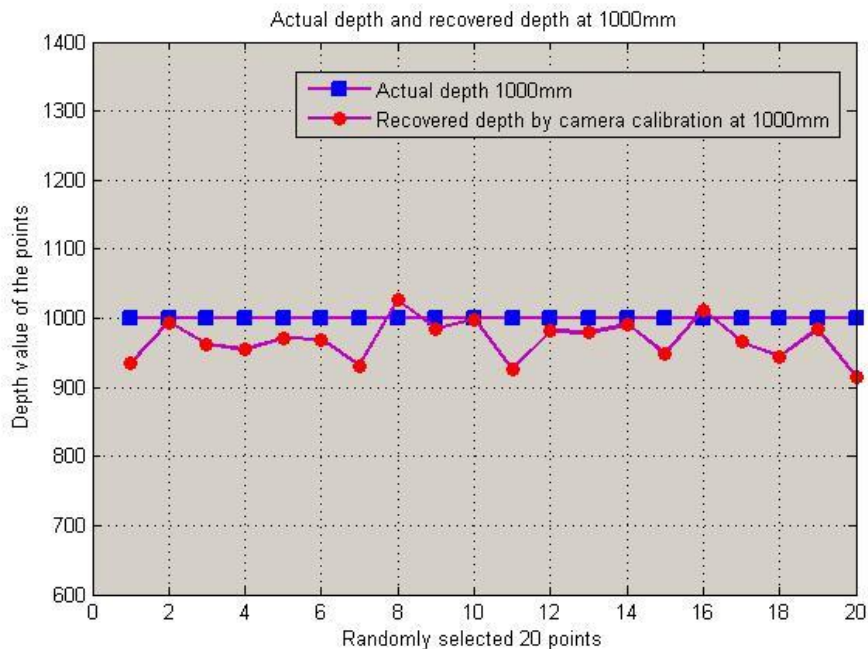


Fig. 5.9: Actual depth and recovered depth by calibration based approach (position A)

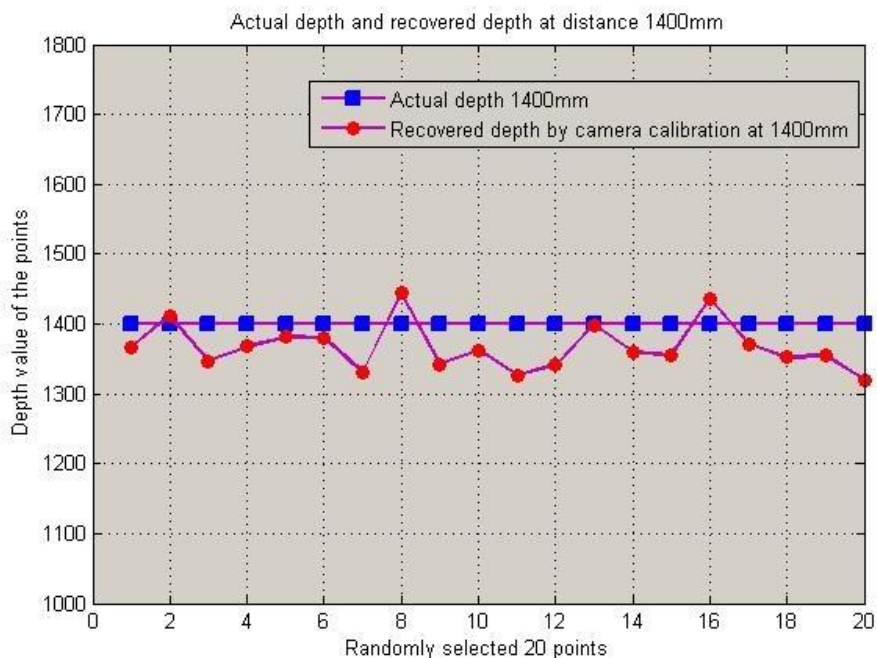


Fig. 5.10: Actual depth and recovered depth by calibration based approach (position B)

$$\text{Error in Percentage} = \left(\frac{\sum |Z_c - Z_c^c|}{20} \right) / Z_c * 100\% = 2.98\%$$

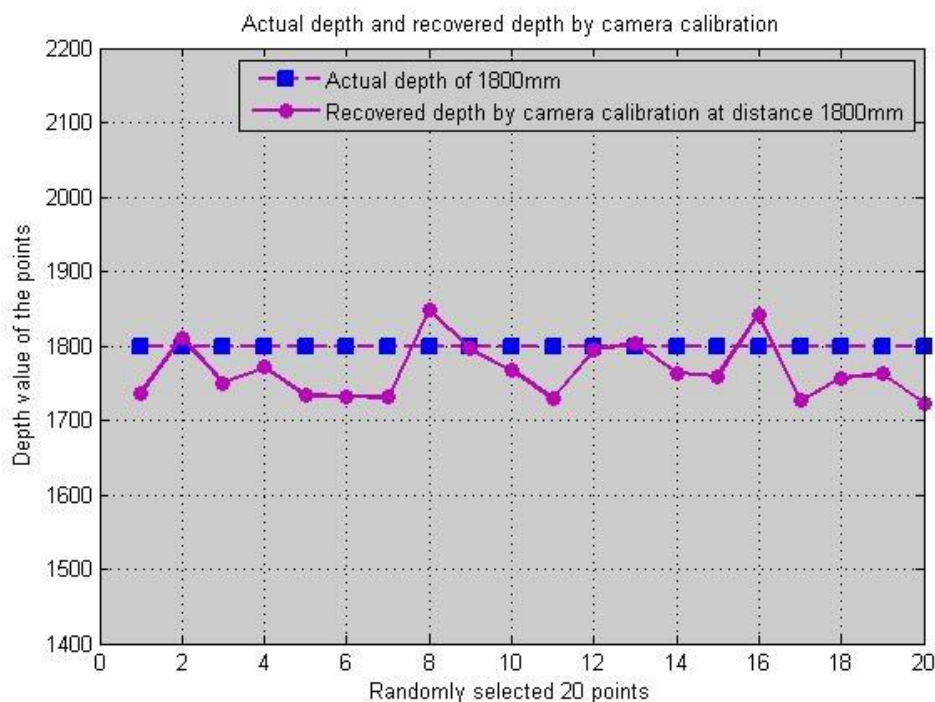


Fig. 5.11: Actual depth and recovered depth by calibration based approach (position C)

$$\text{Error in Percentage} = \left(\frac{\sum |Z_c - Z_c^c|}{20} \right) / Z_c * 100\% = 2.42\%$$

Repeating the same procedures, depth recovery results in the Camera Coordinates at position *B* and *C* are shown in Fig. 5.10 and Fig. 5.11, respectively. The more detailed quantitative results can be found in Appendix *B* (B.1.3 and B.1.4).

5.3 Experimental results by ray sketching based approach

As we know, the bi-prism has a corner angle of $\delta=21.8^\circ$, refraction index $n_r = 1.48$, width $W=100 \text{ mm}$ and thickness $T=20 \text{ mm}$ (Refer to Fig. 4.6). Moreover, it is fixed at a distance of 170 mm away from the camera optical centre ($L_I=170 \text{ mm}$) with the camera parameters are given in Table 5.1.

5.3.1 Results of stereo correspondence by ray sketching based approach

By following the ray sketching based approach discussed in Chapter IV, we perform the epipolar line construction and then the stereo correspondence for the stereo image pair shown in Fig. 5.7.

For the randomly selected 20 image points (marked by Red Circle) on the left half of the image in Fig. 5.7, their corresponding epipolar lines (blue line) are constructed as shown in Fig. 5.12. By using the correspondence search algorithm (Section 4.2.2) along the constructed epipolar lines, the corresponding right image points (marked by Red Square) are determined as shown in Fig. 5.12. The quantitative results of the stereo correspondence at position *A* are presented in Table 5.4.

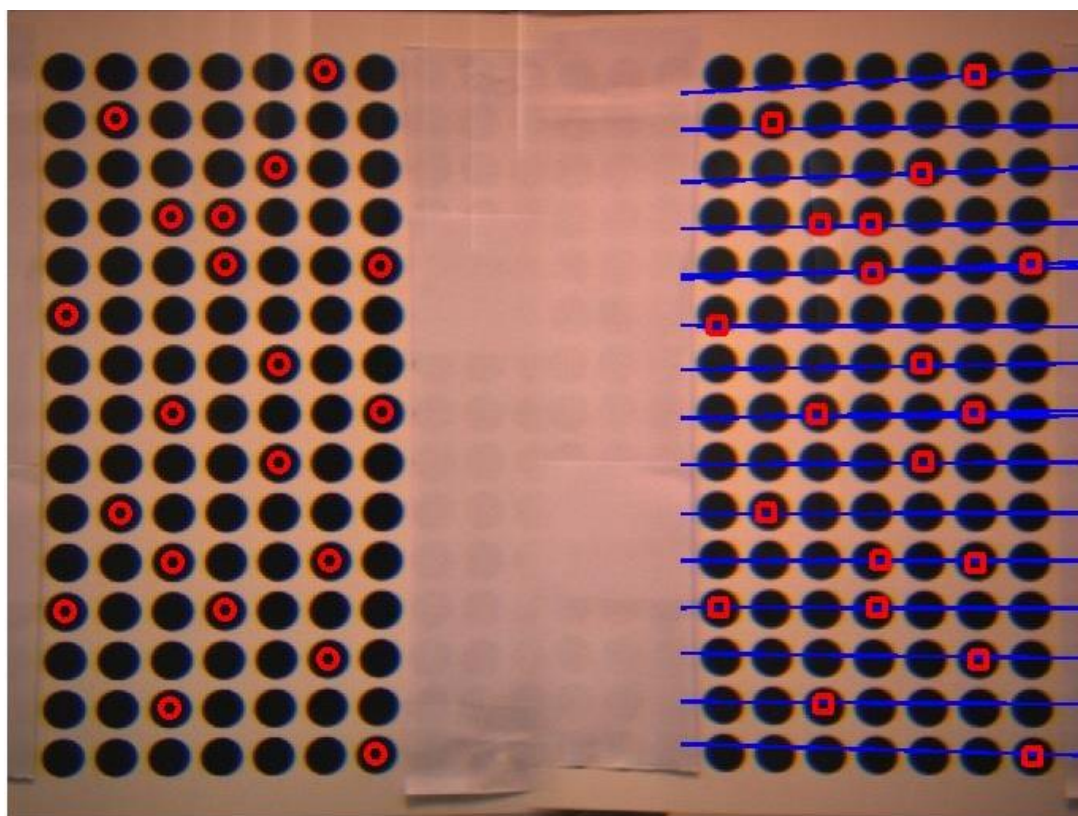


Fig. 5.12: Epipolar line and stereo correspondence by ray sketching based approach (position *A*)

Table 5.4: Result of stereo correspondence by ray sketching based approach (position A)

	X_{lf}	Y_{lf}	X_{rf}	X_{rf}^c	$ X_{rf} - X_{rf}^c $	Y_{rf}	Y_{rf}^c	$ Y_{rf} - Y_{rf}^c $	
1	301	59	920	918	2	57	62	5	
2	104	104	724	725	1	104	108	4	
3	254	150	868	867	1	149	155	6	
4	156	196	770	771	1	195	204	9	
5	206	196	818	818	0	195	203	8	
6	207	242	818	820	2	241	250	9	
7	354	242	966	970	4	241	251	1	
8	56	289	672	673	1	288	300	12	
9	257	335	866	866	0	334	335	1	
10	158	382	768	767	1	381	383	2	
11	356	381	966	916	50	381	382	1	x
12	257	429	866	868	2	428	428	0	
13	107	476	720	720	0	474	476	2	
14	157	523	769	828	59	521	522	1	x
15	306	522	917	917	0	521	524	3	
16	55	569	674	675	1	566	566	0	
17	206	569	820	825	5	567	567	0	
18	304	614	919	921	2	614	616	2	
19	154	661	773	773	0	659	658	1	
20	349	705	971	971	0	707	707	0	

Note: x indicates mismatch points.

Similarly, the epipolar line construction and stereo correspondence results by ray sketching based method at positions *B* and *C* are shown Fig. 5.13 and Fig. 5.14, respectively. The quantitative results can be found in Appendix *B*.

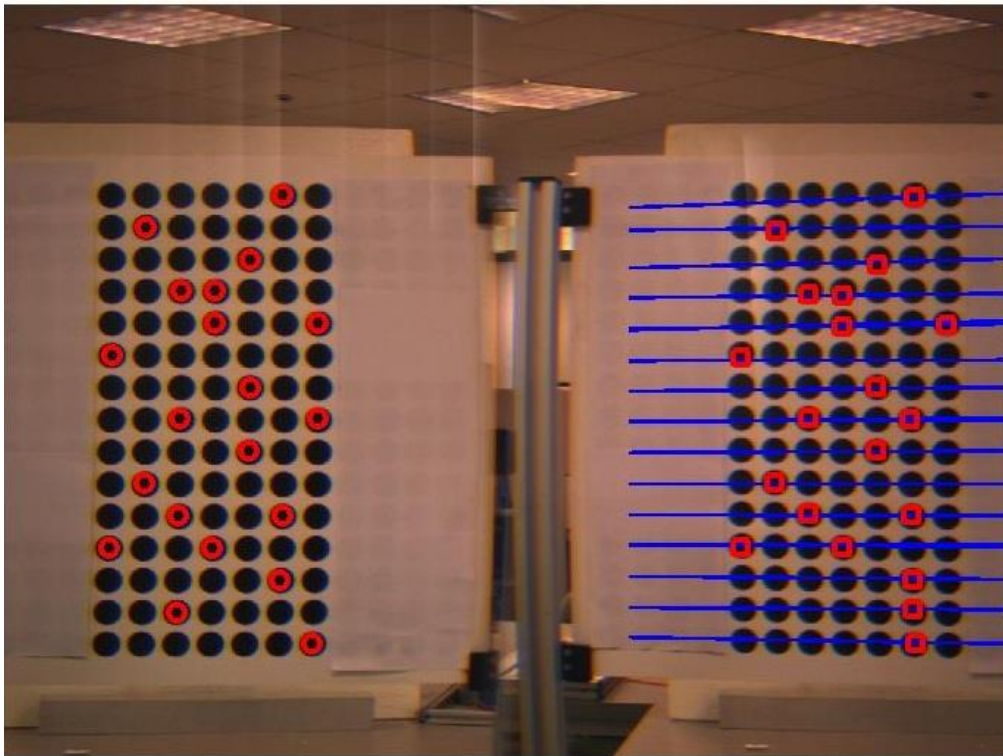


Fig. 5.13: Epipolar line and stereo correspondence by ray sketching based approach (position *B*)

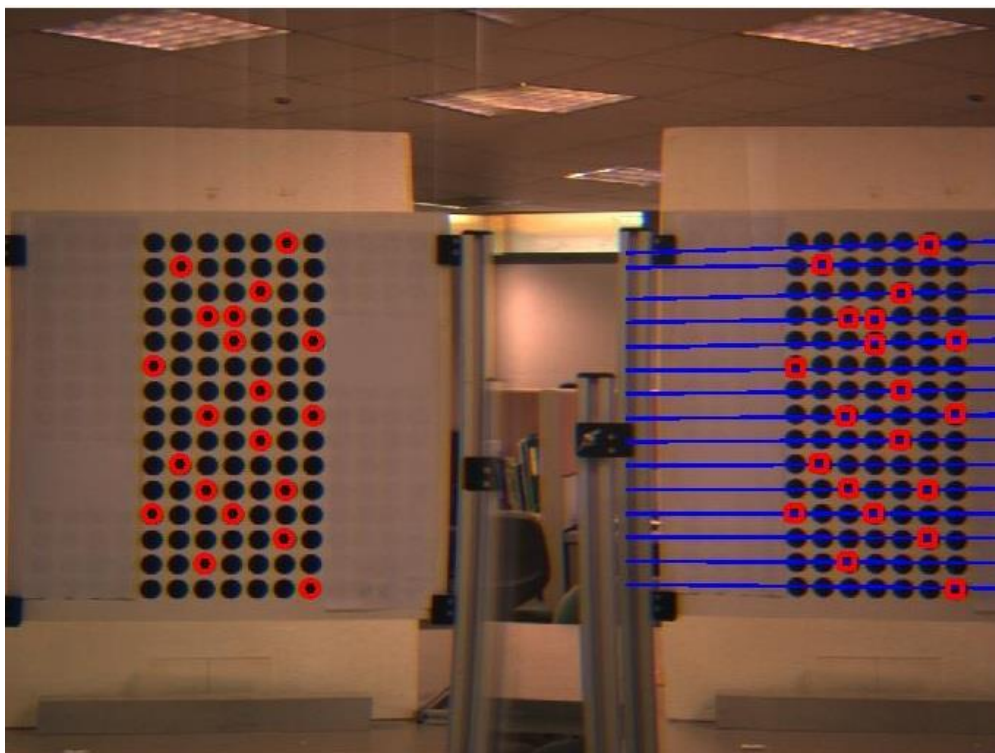


Fig. 5.14: Epipolar line and stereo correspondence by ray sketching based approach (positon *C*)

5.3.2 Results of depth recovery by ray sketching based approach

After solving the stereo correspondence problem, we obtain the 20 correspondence points for the stereo image. Refer to Fig. 4.13, for each pair of the correspondence points, (x_{li}, y_{li}) and (x_{ri}, y_{ri}) , we can determine Rays 1 to 6 (c.f. Section 4.2.2). Rays 3 and 4 normally do not intersect each other due to the digitization errors and noise. We calculate the shortest distance (segment $A'B'$) between the two rays in Camera Coordinates and set the midpoint of the $A'B'$ as our reconstructed 3D points (c.f. Section 4.3).

Fig. 5.15 and Table 5.5 show the depth recovery results by using ray sketching based method at position A .

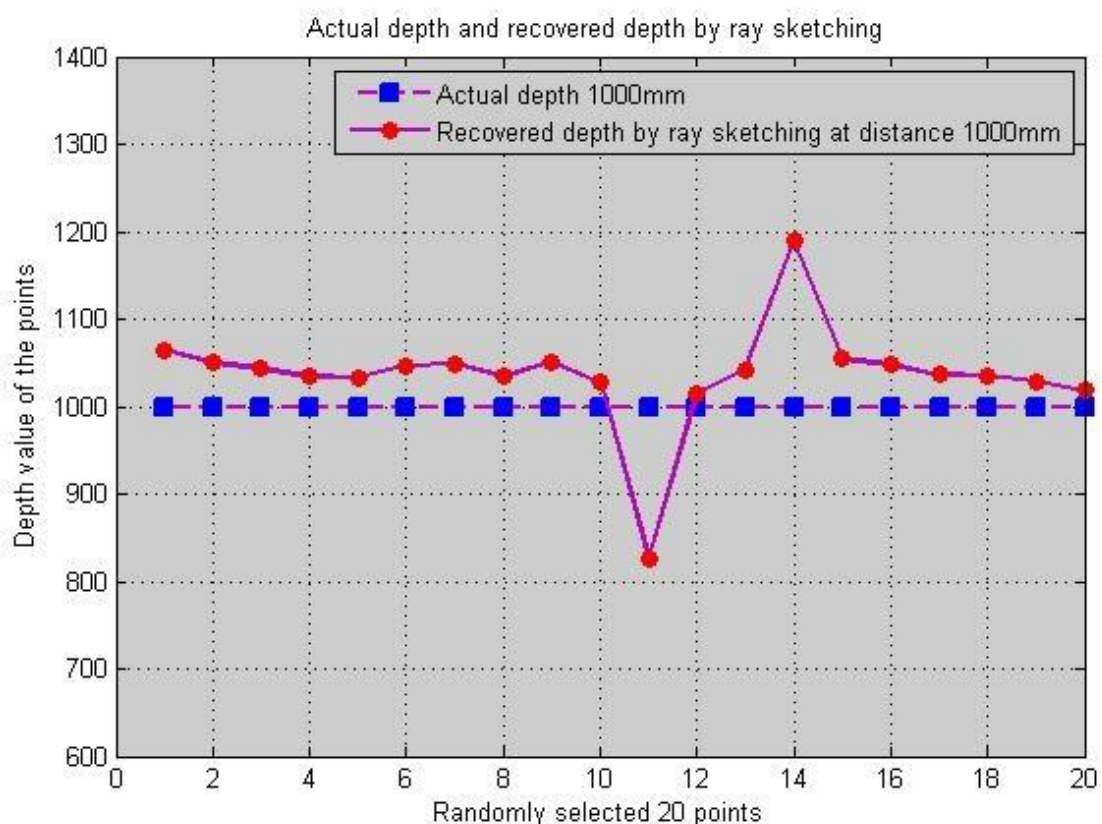


Fig. 5.15: Actual and recovered depth of points by ray sketching based approach (position A)

Table 5.5: The recovered depth of points by ray sketching based approach (position A)

	X_{lf}	Y_{lf}	X_{rf}^c	Y_{rf}^c	Z_c	Z_c^c	$ Z_c - Z_c^c $
1	301	59	918	62	1000	1065	65
2	104	104	725	108	1000	1050	50
3	254	150	867	155	1000	1043	43
4	156	196	771	204	1000	1034	34
5	206	196	818	203	1000	1033	33
6	207	242	820	250	1000	1047	47
7	354	242	970	241	1000	1049	49
8	56	289	673	300	1000	1034	34
9	257	335	866	335	1000	1051	51
10	158	382	767	383	1000	1028	28
11	356	381	916	382	1000	825	175
12	257	429	868	428	1000	1015	15
13	107	476	720	476	1000	1042	42
14	157	523	828	522	1000	1190	190
15	306	522	917	524	1000	1055	55
16	55	569	675	566	1000	1048	48
17	206	569	825	567	1000	1036	36
18	304	614	921	616	1000	1035	35
19	154	661	773	658	1000	1029	29
20	349	705	971	707	1000	1018	18

$$\text{Error in Percentage} = \left(\frac{\sum |Z_c - Z_c^c|}{20} \right) / Z_c * 100\% = 5.39\%$$

For matching points, error in percentage = $\left(\frac{\sum |Z_c - Z_c^c|}{18} \right) / Z_c * 100\% = 4.46\%$ as

there are two mis-matched points.

Fig. 5.16 and Fig. 5.17 present the recovered depth Z_c^c at positions B and C , respectively. The quantitative results can be found in Appendix B (B.2.3 and B.2.4).

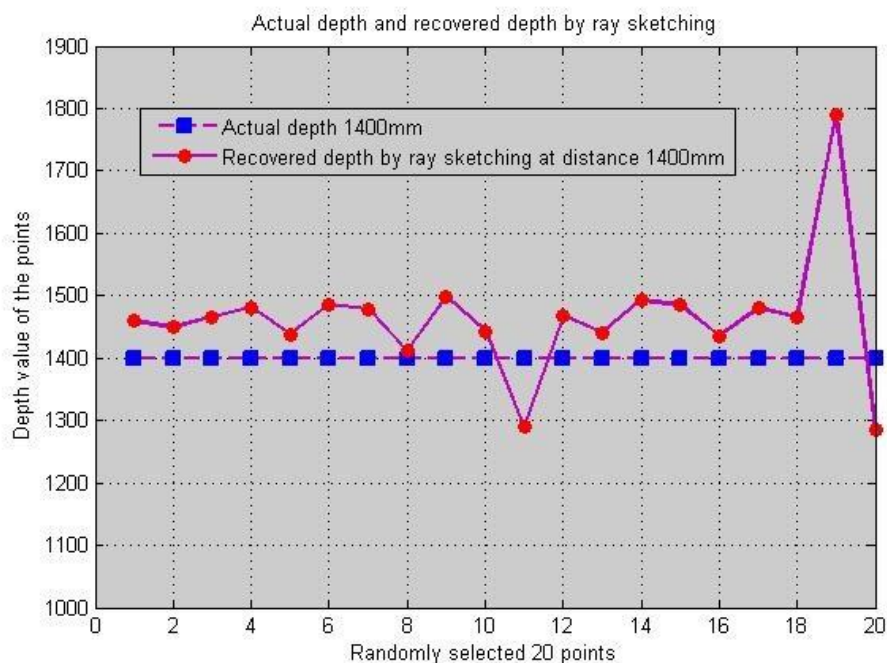


Fig. 5.16: Actual and recovered depth of points by ray sketching based approach (position B)

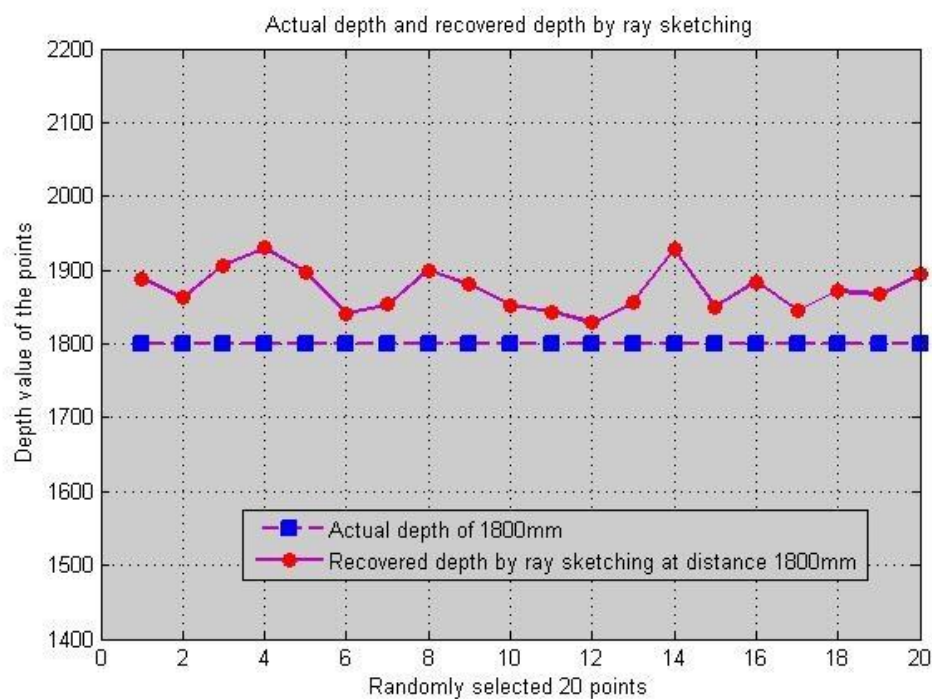


Fig. 5.17: Actual and recovered depth of points by ray sketching based approach (position C)

5.4 Evaluation and discussion of the experimental results

The depth recovery results for both the calibration based approach and ray sketching based approach are plotted for comparison as follows.

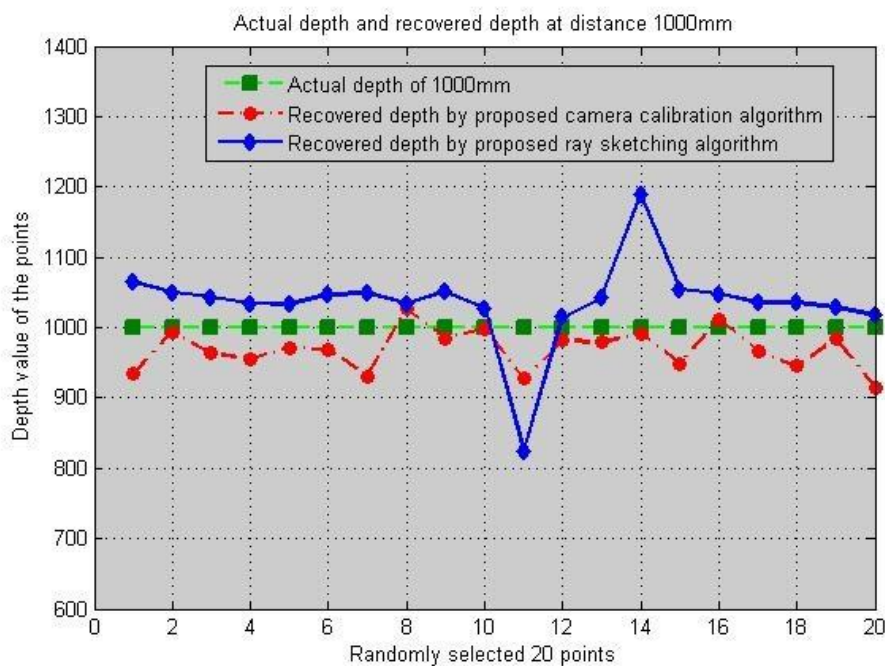


Fig. 5.18: Depth recovery of the 20 points by two approaches at position A

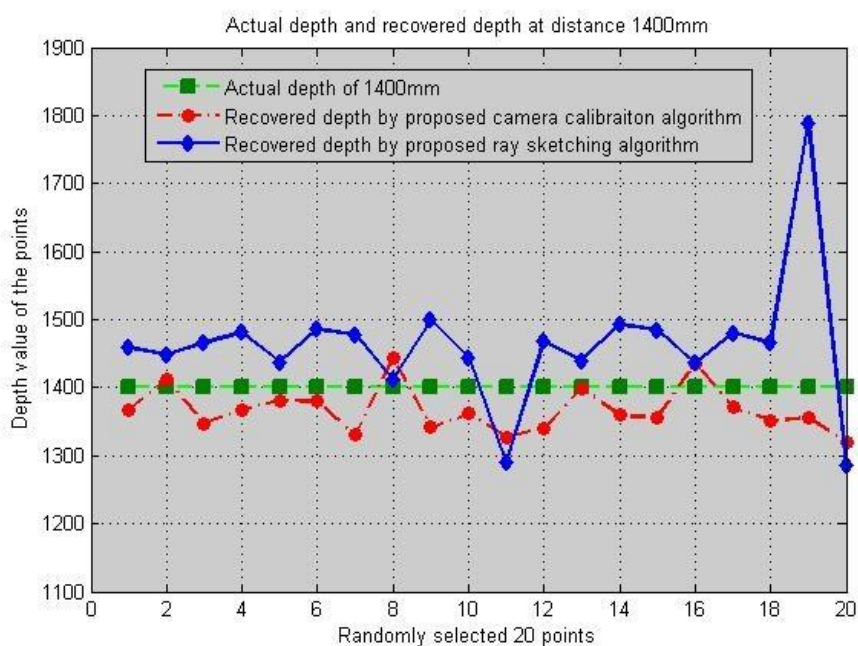


Fig. 5.19: Depth recovery of the 20 points by two approaches at position B

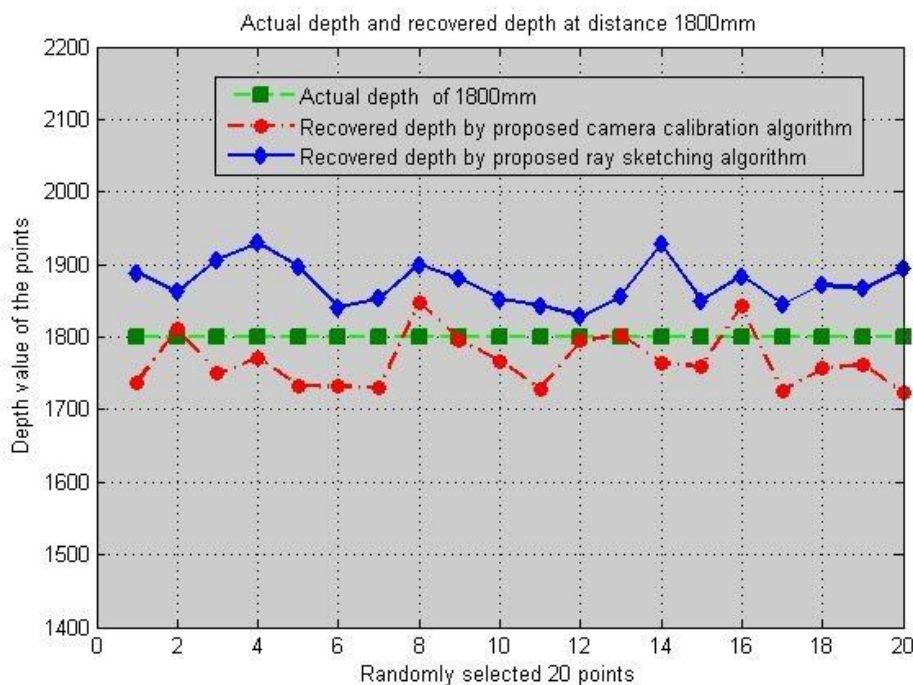


Fig. 5.20: Depth recovery of the 20 points by two approaches at position C

5.4.1 Evaluation and discussion on the camera calibration based approach

In the ensuing evaluation and discussion of results, we compare our results with those presented by Lim and Xiao [35], as their works are close to that presented in this thesis. In addition, they have published the most comprehensive theoretical development and results for the first time in recent years. By using the optical prism with corner angle δ equal to 3.1° in the experiments, Lim and Xiao [35] performed the depth recovery by their camera calibration based approach within the range of 1.3m to 2.5m and achieved an absolute recovery error of 1%. In contrast, we use large corner angle of prism ($\delta = 21.8^\circ$) in our experiments and the average depth recovery error of 3% is obtained for the range from 1m to 1.8m.

Within the same depth range, our depth recovery error is larger than Lim and Xiao's. Reviewing their methods and experimental setup, the following observations are listed:

- 1) The different optical prisms used in the experiment

The prism used in [35] has the corner angle δ of to 3.1° , which leads to less error to the system than our prism (corner angle δ of to 21.8°). This can be attributed to the fact that the larger the prism corner angle is, the larger would be the error caused by the thin prism distortions.

- 2) The assumption of distortion-free of the camera lens is made in the camera calibration based approach

In the discussion on the 4-step transformation of the camera calibration based approach, we have made the assumption that there are no distortions, such as lens radial distortion and prism distortion, in the stereovision system (c.f. Eq. (3.3)). With this assumption, we have $x_d = x_u, y_d = y_u$ and $D_x = D_y = 0$.

In fact, distortions always exist in the stereovision system. Although Lim and Xiao have made this same assumption, our calibration parameters for both the real camera and virtual cameras will be less accurate due to the larger prism distortion being neglected.

- 3) The different approaches for the solutions of the stereo correspondence

Referring to Eq. (3.33) and Eq. (3.34), the coordinates of the corresponding image points, (X_{lf}, Y_{lf}) and (X_{rf}, Y_{rf}) , need to be determined for the purpose of the depth recovery.

In Lim and Xiao's work, the stereo correspondence was not investigated and solved. The coordinates of the corresponding points (X_{lf}, Y_{lf}) and (X_{rf}, Y_{rf}) were read directly from the computer screen by putting the colour markers on the interested points.

However, in our novel approach, we calculate the corresponding points based on the results of the virtual camera calibrations (c.f. Eq. (3.23) and Eq. (3.28)). No markers are needed and the image points can be any points in the common view of the two virtual cameras. Obviously, the errors of our stereo correspondence results (through calibration)

could also be the factors contributing to the errors of the recovered depth.

4) The different algorithms for the depth recovery

As we know, corresponding points (X_{lf}, Y_{lf}) and (X_{rf}, Y_{rf}) (with markers) was read directly from the computer screen in Lim and Xiao's work. They have also enforced a constraint on the Y coordinates, which is, $Y_{lf} = Y_{rf}$, when performing the depth recovery. In another words, they assumed that the two virtual camera image planes are coplanar and, it can then be assumed that the epipolar lines are horizontally positioned.

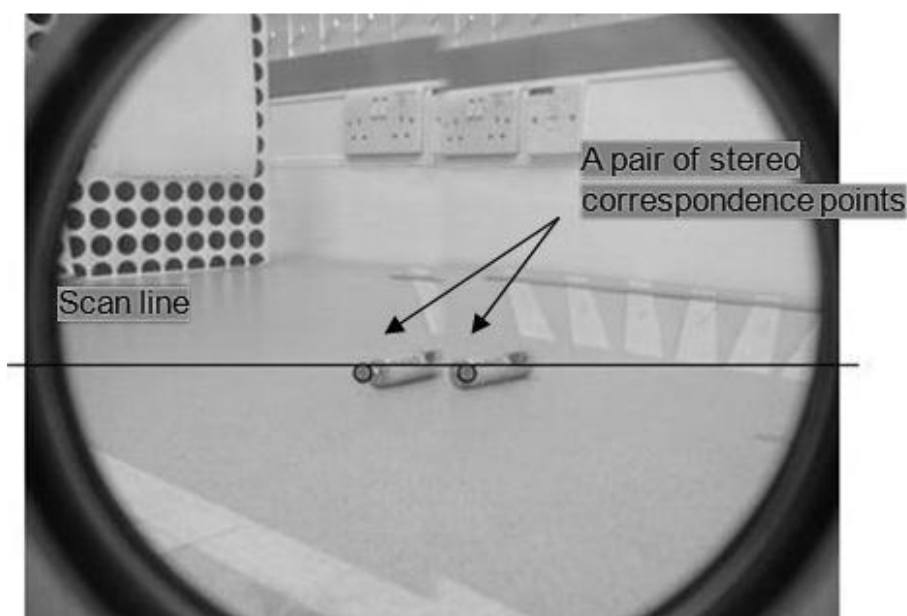


Fig. 5.21: Illustration of the assumption of $Y_{lf} = Y_{rf}$

Under these circumstances, their depth recovery can be obtained through the simple triangulation method shown below: (c.f. Eq. (2.1))

$$Z_c^c = \lambda \frac{f}{x_r - x_l} \quad . \quad (5.12)$$

However, this constraint may only be enforced when the prism corner angle is very small. For the corner angle $\delta=21.8^\circ$ of our experimental prism, their constraint cannot be applied. The left and right virtual camera image planes are not coplanar anymore. In reality, there are rotational and translational relationships between the two virtual camera

image planes and therefore, the epipolar lines are not horizontal.

We proposed a new approach for the depth recovery in Section 3.3. Except the parameters of the two virtual cameras, we also introduce the real camera calibration parameters in the depth recovery. (c.f. Eq. (3.35)) The direct transformation from Camera Coordinates to the World Coordinates not only avoids any loss of information but also makes our approach more versatile. Whereas, the disadvantage of the new approach is: the imperfect real camera calibration parameters which enlarge our depth recovery errors. However, it is unavoidable.

Therefore, our camera calibration based approach is less accurate in depth recovery than the one presented in [35]. But the advantages of the algorithm itself over the method used in [35] are obvious.

- i) Knowledge on the dimensions of the prism is not required, so as to the length of the baseline λ of the two virtual cameras. (c.f. Eq. (5.12))
- ii) The proposed algorithm is versatile and applicable for both the small and large angles of a prism used in the single-lens stereovision system.
- iii) The algorithms for the stereo correspondence do not require any external markers for the identification of contentious points on the screen.
- iv) The novel depth recovery algorithm proposed in this thesis utilizes the calibration parameters of the real camera and the two virtual cameras. No constraints on the displacement of the two virtual cameras are enforced.

By observing Fig. 5.8, we notice that there is no mismatching issue appeared between the corresponding images using the camera calibration based method. By scanning the left (right) image pixel by pixel, the corresponding right (left) image point can be computed through the two virtual camera calibration parameters (c.f. Eq. (3.23)),

regardless of the qualities of the images. It indicates that the calibration based method is insensitive to the image occlusions and noises. Obviously, the accuracy of the stereo correspondence is highly dependent on the accuracy of the camera calibration parameters.

Figures (5.18), (5.19) and (5.20) show the depth recovery results by the camera calibration based approach at positions *A*, *B* and *C*. We can observe the trend that the recovered depth is mostly smaller than the actual depth. With a bias added in the recovered depth, we can expect that the results would be accurate enough for many applications, such as autonomous robot indoor navigation.

5.4.2 Evaluation and discussion on ray sketching based method

The proposed ray sketching based method for the stereo correspondence and depth recovery of the single-lens prism based stereovision system developed in this thesis is novel. Comparing to the ray sketching based method reported in [34] which described light rays in two-dimensional space, our method is developed in the three-dimensional space without the boundary identification. All the points, lines and surfaces are described based on geometrical analyses. We are able to construct the epipolar lines and observe that:

- 1) The epipolar lines are all well-determined with an acceptable level of error.
- 2) Theoretically, the epipolar lines should pass through the centres of the dots on the test board used in our experiments where the corresponding right images are located. We can observe visually the clear image distortion in the three figures due to the lens and prism distortions. Therefore, our constructed epipolar lines have small errors from the desired positions.
- 3) The stereo correspondence results by ray sketching are presented in the same figures as epipolar lines (Fig. (5.12), (5.13) and (5.14)). Searching along the epipolar lines,

there are few mismatched correspondence points (2 mismatched at position *A* and 3 mismatched at position *B*). The mismatch could be due to the variations of the image brightness, which causes inconsistency in intensity values for the same image points.

- 4) There is no meaningful comparison with our ray sketching based method as this method is novel. Nevertheless, the experimental results indicate that the method is effective in solving the stereo correspondence of the stereovision system, regardless the size or corner angle of the prism being used.

The mismatched correspondence points clearly result in considerable error in the depth recovery. The two findings of the depth recovery by the ray sketching based approach are:

Finding 1: The error of the recovered depth is around 4%;

The 4% depth recovery error is caused by:

- a) The assumption of the properties of the prism, such as 100% light refraction (no light diffusion), smooth glass surface and the assumption that n_r is a constant.
- b) Intrinsic distortions of the system due to camera lens and prism distortion.

Our intrinsic distortion includes radial distortion and de-centring distortions which are caused by the characteristics of the camera. In addition, thin prism distortion, chromatic aberrations, weak reflections and blurring caused by an optical prism, can contribute to the error of the system [85-86]. Normally, the effect of radial distortion is relative higher than the other distortions. It is caused by the flawed shape of the lens, giving the symmetric distortion along the optical axis and we can clearly observe its effect on the image of Fig. 5.12. The model of radial distortion is represented in Eq. (5.13).

$$\begin{aligned} D_{xr} &= k_r x_d (x_d^2 + y_d^2) + O(x_d, y_d)^5 \\ D_{yr} &= k_r y_d (x_d^2 + y_d^2) + O(x_d, y_d)^5 \end{aligned} \quad (5.13)$$

Where k_r is the coefficient of radial distortion; x_d and y_d are the (x, y) coordinates of the distorted image point. (c.f. Eq. (3.3))

c) Extrinsic distortion of the system

In the derivation of the ray sketching based method, the setup of the optical prism is required to fulfil:

- 1) surface 2 of the prism is parallel to the camera image plane, i.e. $\sigma=0$;
- 2) the center point of the prism lies on the optical axis of the camera ($d=0$).

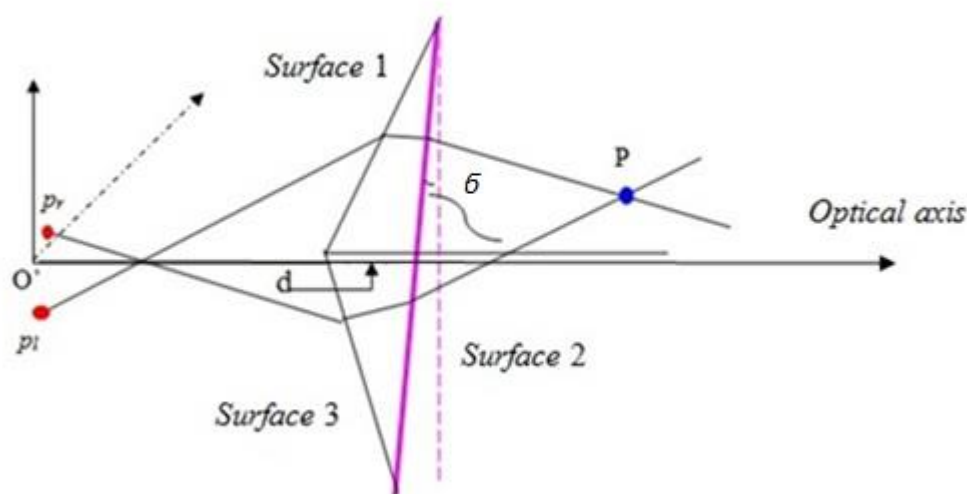


Fig. 5.22: Prism misalignment in the stereovision system setup

However, in practice, these two assumptions are difficult to satisfy. The non-zero values of d and δ will change the equations of the prism Surface 1, 2 and 3. And eventually will affect the results of the epipolar line, stereo correspondence and then, the depth recovery.

Finding 2: All the recovered depths are larger than the actual depth

Figures (5.18), (5.19) and (5.20) show the depth recovery results by the ray sketching based approach at positions *A*, *B* and *C*. We can infer the trend that the recovered depth is larger than the actual depth within a small error range. This also demonstrates that the novel method is working. Again, a bias could be determined and then be added into the recovered depth to account for the errors.

5.4.3 Summary

From the above discussion, we can summarize that:

- 1) The two proposed methods are valid. The ray sketching based algorithm is new and the calibration based method is more versatile.
- 2) Errors exist for both approaches. We have identified the sources of error, which should be addressed in the future works of this thesis.
- 3) For camera calibration based approach, the error is around 3%, whereas, it is around 4% error for the ray sketching based approach. The latter is preferred because of its better computational efficiency the avoidance of complicated calibration procedures.

Chapter VI Conclusions

A single-lens binocular stereovision system using an optical prism has been developed, analyzed, implemented and tested. A systematic and in depth investigation on this system including image capturing, stereo correspondence and depth recovery has been carried out. In this system, with an optical prism placed in front of the camera and its back facing the scene of interest, each captured image is split into two sub-images. The images are considered to have been captured simultaneously by two virtual cameras generated by the prism.

The aim of this research is to establish faithful model of the system and to develop efficient methods to solve the stereo correspondence issue and to achieve depth recovery of a 3D scene. Two different approaches were presented for this purpose: one is based on camera calibration and the other is based on a novel ray sketching technique developed in this thesis. The experimental work proves that the latter is easier in implementation and yields results of comparable accuracy.

A comprehensive literature review has been carried out, followed by the system design and theoretical analyses. Mathematical basics of the two said approaches are then established. Subsequently, specially designed hardware setup is designed, the components of which are fabricated to very high accuracy with tight mechanical tolerances. This is to alleviate errors introduced due to hardware setup, so that we can concentrate our efforts in locating other sources of errors affecting the experimental results. Software, coded mainly in MATLAB, is then developed for the system, followed by experimental design. The final results showed that the two approaches are both effective and with sufficient accuracy to enable the system to work as a binocular stereovision system.

As a stereo image pairs can be captured simultaneously, this system presents several advantages over a conventional binocular stereovision system. In addition, the system also has the following advantages:

- 1) No inter-camera synchronization is needed since only one camera is used. This is especially important when dealing with dynamic scenes;
- 2) The use of one filter/prism to replace one or more cameras will significantly reduce the cost and space of building a multi-camera stereovision system;
- 3) This single-lens system has fewer system parameters, which renders it easier in its implementation. This is especially so for the ray sketching based approach.

However, 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 filter size and shape, and hence, at the present moment, this system possibly can only find its main applications for short-range stereovision (0.8mm-3mm).

As mentioned earlier, two approaches have been developed for the stereo correspondence and depth recovery of the single-lens prism based stereovision system. They are summarized below.

(1) The camera calibration based approach

The nature of this approach is to establish the relationship between the projected digital images and the scene by a four-step transformation - from the 3D World Coordinates of the scene to the 3D Camera Coordinates, to 2D Image Plane and finally to 2D Digital Image Plane where the digital image locates. We have successfully derived the

linear relationship for the real camera and its two associated virtual cameras. It is a 3 by 4 projective matrix in Homogenous Coordinates, consisting of the system intrinsic and extrinsic parameters. The solutions of the stereo correspondence and depth recovery are therefore quite straightforward by matrix calculation. Experimental results are convincing, which show that this approach is valid. However, this method is applicable only when the camera system is well-calibrated and the camera system model is distortion free.

(2) The ray sketching based approach

This approach is camera calibration free, thereby reduces the fastidious calibration process. The idea of the approach is to trace the light paths of the system in three-dimensional space by following the Snell's Law to revert the imaging process. Its basis is the ability to express the pertinent points, lines (light paths) and planes in 3D Camera Coordinates landing on the unique geometrical property of our single-lens stereovision system. 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. This is because the candidate points form the epipolar line and one can search for the correct corresponding point without much difficulty along the epipolar line. The intersection point of the two rays from the located corresponding pair would be the three-dimensional point.

The major advantage of the two aforesaid proposed approaches is: they can be generalized from handling the binocular stereovision system to the tri-ocular and even multi-ocular stereovision systems. The difference between binocular and tri-ocular systems in the case of the calibration based technique is the addition of only one more set of processing data. As for the ray sketching based approach, we need only to derive

different equations of the prism surfaces. However, it is worthwhile to note that the increased number of virtual cameras will reduce the view zone for each virtual camera as they share only one CCD array. This will reduce the accuracy of calibration

The experimental and simulation results validate the reliability and practicality of the two proposed approaches. The average depth recovery error for the two methods is about 3% to 4% for the recovery range of 1m to 2 m. In addition, the ray sketching performs slightly better in this regard. The usefulness of this will be for the indoor robot navigation, object detection, small size hand-held stereovision system for dynamic scene, 3D feature checker and stereoscopic endoscope, etc.

The major contributions of this thesis include:

- 1) We are the first research group proposed “virtual camera concept” and carried out the “virtual camera research”. The preliminary work done by the group peers focused on the “virtual camera calibration technique”. Our work is the first attempt to solve the stereo correspondence and depth recovery using single-lens prism based stereovision system.
- 2) A novel ray sketching based approach is investigated in this research for the solution of the prism based stereovision system. It builds the foundation for the future research among the research group. This developed approach avoids tedious and time-consuming camera calibration process and generates competitive results. Furthermore, it can be easily generalized from binocular stereovision system to tri-ocular and even multi-ocular stereovision systems for the solution of stereo correspondence and depth recovery/3D reconstruction.
- 3) We developed a new depth recovery algorithm based on the camera calibration results, with this algorithm (compare to Lim and Xiao):

- Knowledge of the optical prism is not required;
- No constraints on the displacement of the two virtual cameras are enforced;
- It is versatile and applicable for both the small and large angle of prism used in the stereovision system.

Three future work directions are recommended from this thesis.

Firstly, to introduce the consideration on lens distortion in the model, with which the accuracy of the system in depth recovery will be improved. The effect of the distortion will be even more severe when the system is used in endoscopic applications when the camera lens is smaller in diameter.

Secondly, as the accuracy of the position and orientation of the system setup greatly affects the system accuracy, future studies should include the investigation on the effect of mis-alignments of the prism and camera, etc. This will help to develop measure to counter the problem.

Last but not the least, the range of 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 geometries should be used to further investigate the system performance. This will extend the recoverable depth of the system.

BIBLIOGRAPHY

- [1] P. Zhou, Y. Liu, and Y. Wang, "Pipeline Architecture and Parallel Computation-Based Real-Time Stereovision Tracking System for Surgical Navigation", *IEEE Transactions on Instrumentation and Measurement*, vol. 59(5), pp. 1240-1250, 2010.
- [2] J. X. Yang, et al., "Research on Computer Aided Surgical Navigation Based on Binocular Stereovision", in *Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation*, Luoyang, Henan, 2006, pp. 1532-1536.
- [3] A. Naoulou, et al., "A 3D Real-time Vision System Based on Passive Stereovision Algorithms: Application to Laparoscopic Surgical Manipulations", in *Proceedings of Information and Communication Technologies*, Damascus, 2006, pp. 1068-1073.
- [4] D. S. Kim, S. S. Lee, and B. H. Choi, "A Real-time Stereo Depth Extraction Hardware for Intelligent Home Assistant Robot", *IEEE Transactions on Consumer Electronics*, vol. 56(3), pp. 1782-1788, 2010.
- [5] T. P. Pachidis and J. N. Lygouras, "Pseudo Stereo-vision System: A Monocular Stereo-vision System As a Sensor for Real-time Robot Applications", *IEEE Transactions on Instrumentation and Measurement*, vol. 56(6), pp. 2547-2560, 2007.
- [6] R. Racca and J. Dewey, "A Method for Automatic Particle Tracking In a Three-dimensional Flow Field", *Experiments In Fluids*, vol. 6(1), pp. 25-32, 1988.
- [7] M. Virant and T. Dracos, "3D PTV and Its Application on Lagrangian Motion", *Measurement Science and Technology*, vol. 8(12), pp. 1539, 1999.
- [8] S. T. Barnard and M. A. Fischler, "Computational Stereo", *ACM Computing Surveys (CSUR)*, vol. 14(4), pp. 553-572, 1982.
- [9] D. Marr, "Vision - A Computational Investigation Into the Human Representation and Processing of Visual information", *Journal of Mathematical Psychology*, vol. 27(1), pp. 107-110, 1983.

- [10] K. B. Lim and Y. Xiao, "Virtual Stereovision System: New Understanding on Single-lens Stereovision Using a Biprism", *Journal of Electronic Imaging*, vol. 14(4), pp. 043020-043020-11, 2005.
- [11] L. Grewe and A. Kak, "Handbook of Pattern Recognition and Image Processing: Computer Vision", Academic Press Inc., vol. 2, pp. 239-317, 1994.
- [12] R. Tsai, "A Versatile Camera Calibration Technique for High-accuracy 3D Machine Vision Metrology Using Off-the-shelf TV Cameras and Lenses", *IEEE Journal of Robotics and Automation*, vol. 3(4), pp. 323-344, 1987.
- [13] Z. Zhang, "A Flexible New Technique for Camera Calibration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22(11), pp. 1330-1334, 2000.
- [14] Z. Zhang, "Camera Calibration with One-Dimensional Objects", in *Proceedings of the 7th European Conference on Computer Vision-Part IV*, London, UK, 2002, pp. 161-174.
- [15] E. Trucco and A. Verri, "Introductory Techniques for 3-D Computer Vision", ed: Prentice Hall, 1998.
- [16] J. Gluckman and S. K. Nayar, "Catadioptric Stereo Using Planar Mirrors", *International Journal of Computer Vision*, vol. 44(1), pp. 65-79, 2001.
- [17] J. Segen and S. Kumar, "Shadow Gestures: 3D Hand Pose Estimation Using a Single Camera", in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Fort Collins, CO, 1999, pp. 479-485.
- [18] R. LeGrand and R. C. Luo, "Position Estimation of Selected Targets", in *Proceedings of IEEE International Conference on Robotics and Automation*, Minneapolis, MN, 1996, pp. 1714-1719.
- [19] D. Moore and M. Hayes, "Tracking 3d Position and Orientation from 2D Sequences Using Simple Geometry", in *Proceedings of Conference Record of the Thirtieth*

- Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, 1996, pp. 125-129.
- [20] Y. Suzuki and M. Takano, "Estimation of the Position of a Rigid Body with a Single Camera", in Proceedings of the International Conference on Industrial Electronics, Control and Instrumentation, Kobe, 1991, pp. 1309-1311.
- [21] E. H. Adelson and J. Y. A. Wang, "Single Lens Stereo with a Plenoptic Camera", Ieee Transactions on Pattern Analysis and Machine Intelligence, vol. 14(2), pp. 99-106, 1992.
- [22] J. Cardillo, M. Sid-Ahmed, and J. Soltis, "3-D Position Sensing Using a Single Camera Approach", in Proceedings of the 32nd Midwest Symposium on Circuits and Systems, Champaign, IL, 1989, pp. 325-328.
- [23] G. Lester, J. Watts, and D. Wilmington, "Ferroelectric Liquid Crystal Device for a Single Camera Stereoscopic Endoscope System", Electronics Letters, vol. 33(10), pp. 857-858, 1997.
- [24] G. Lester, J. Watts, and D. Wilmington, "Single Camera Three-dimensional Endoscope System Using a Ferroelectric Liquid Crystal Device", IEE Proceedings of Science, Measurement and Technology, vol. 145(2), pp. 49-51, 1998.
- [25] D. W. Murray, "Recovering Range Using Virtual Multicamera Stereo", Computer Vision and Image Understanding, vol. 61(2), pp. 285-291, 1995.
- [26] Y. Nishimoto and Y. Shirai, "A Feature-based Stereo Model Using Small Disparities", in Proceedings of Computer Vision and Pattern Recognition, 1987, pp. 192-196.
- [27] W. Teoh and X. Zhang, "An Inexpensive Stereoscopic Vision System for Robots", in Proceedings of IEEE International Conference on Robotics and Automation, 1984, pp. 186-189.

- [28] A. Goshtasby and W. A. Gruver, "Design of a Single-lens Stereo Camera System", *Pattern Recognition*, vol. 26(6), pp. 923-937, 1993.
- [29] S. A. Nene and S. K. Nayar, "Stereo with Mirrors", in *Proceedings of the Sixth International Conference on Computer Vision, Bombay*, 1998, pp. 1087-1094.
- [30] V. P. Lee, "Stereovision Using A Single CCD Camera", Master Thesis, Department of Mechanical Engineering, NUS, 2001.
- [31] L. C. Ng, "Stereo-vision Using Single CCD Camera", Technical Report, Control and Mechantronics Laboratory, Department of Mechanical Engineering, NUS, 2001.
- [32] D. H. Lee and I. S. Kweon, "A Novel Stereo Camera System by a Biprism", *IEEE Transactions on Robotics and Automation*, vol. 16(5), pp. 528-541, 2000.
- [33] Y. Xiao and K. B. Lim, "A Single-lens Trinocular Stereovision System Using a 3F Filter", in *Proceedings of 2004 IEEE Conference on Robotics, Automation and Mechatronics*, 2004, pp. 396-400.
- [34] Y. Xiao and K. B. Lim, "A Prism-based Single-lens Stereovision System: From Trinocular to Multi-ocular", *Image and Vision Computing*, vol. 25(11), pp. 1725-1736, 2007.
- [35] Y. Xiao, "Stereovision Using Single CCD Camera", PhD Thesis, Control and Mechantronics Laboratory, Department of Mechanical Engineering, NUS, 2006.
- [36] U. R. Dhond and J. K. Aggarwal, "Structure from Stereo-A Review", *IEEE Transactions on Systems, Man and Cybernetics*, vol. 19(6), pp. 1489-1510, 1989.
- [37] J. E. W. Mayhew and J. P. Frisby, "Psychophysical and Computational Studies Towards a Theory of Human Stereopsis", *Artificial Intelligence*, vol. 17(1), pp. 349-385, 1981.
- [38] P. M. Hans, "Towards Automatic Visual Obstacle Avoidance", in *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, 1977, pp. 584-584.

- [39] Y. Kim and J. Aggarwal, "Positioning 3-D Objects Using Stereo Images", *IEEE Journal of Robotics and Automation*, vol. 3(4), pp. 361-373, 1987.
- [40] D. Marr and T. Poggio, "Cooperative Computation of Stereo Disparity", *Science (New York, NY)*, vol. 194(4262), pp. 283, 1976.
- [41] G. Medioni and R. Nevatia, "Segment-based Stereo Matching", *Computer Vision, Graphics, and Image Processing*, vol. 31(1), pp. 2-18, 1985.
- [42] S. T. Barnard, "Stochastic Stereo Matching over Scale", *International Journal of Computer Vision*, vol. 3(1), pp. 17-32, 1989.
- [43] D. Terzopoulos, "Concurrent Multilevel Relaxation", in *Proceedings of Image Understanding Workshop*, 1985, pp. 156-161.
- [44] D. Terzopoulos, "Multilevel Computational Processes for Visual Surface Reconstruction", *Computer Vision, Graphics, and Image Processing*, vol. 24(1), pp. 52-96, 1983.
- [45] H. H. Baker, "Depth from Edge and Intensity Based Stereo", in *Proceedings of 7th International Conference of Artificial Intelligent*, 1982, pp. 631-636.
- [46] Y. Ohta and T. Kanade, "Stereo by Intra-and Inter-scanline Search Using Dynamic Programming", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 7(2), pp. 139-154, 1985.
- [47] M. Z. Brown, D. Burschka, and G. D. Hager, "Advanced in Computational Stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25(8), pp. 993-1008, 2003.
- [48] A. Koschan, "What is New in Computational Stereo Since 1989: A Survey on Current Stereo Papers", *Technical Report*, 1993.

- [49] R. Hartley, R. Gupta, and T. Chang, "Stereo from Uncalibrated Cameras", in IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, IL, 1992, pp. 761-764.
- [50] Z. Zhang, "Determining the Epipolar Geometry and Its Uncertainty: A Review", International Journal of Computer Vision, vol. 27(2), pp. 161-195, 1998.
- [51] R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision", vol. 2 Cambridge Univ Press, 2000.
- [52] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms", International Journal of Computer Vision, vol. 47(1), pp. 7-42, 2002.
- [53] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow", Artificial Intelligence, vol. 17(1), pp. 185-203, 1981.
- [54] V. Venkateswar and R. Chellappa, "Hierarchical Stereo and Motion Correspondence Using Feature Grouping", International Journal of Computer Vision, vol. 15(3), pp. 245-269, 1995.
- [55] S. Todorovic and N. Ahuja, "Region-based Hierarchical Image Matching", International Journal of Computer Vision, vol. 78(1), pp. 47-66, 2008.
- [56] C. Harris, Ed., "Geometry from Visual Motion: In Active Vision": MIT Press, 1993, pp. 263-284.
- [57] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, vol. 60(2), pp. 91-110, 2004.
- [58] M. Brown and D. G. Lowe, "Invariant Features from Interest Point Groups", in Proceedings of British Machine Vision Conference, Cardiff, Wales, 2002, pp. 656-665.

- [59] K. Mikolajczyk and C. Schmid, "An Affine Invariant Interest Point Detector", in Proceedings of the 7th European Conference on Computer Vision-Part I, 2002, pp. 128-142.
- [60] M. E. Munich, et al., "SIFT-ing through features with ViPR - Application of visual pattern recognition to robotics and automation", IEEE Robotics & Automation Magazine, vol. 13(3), pp. 72-77, 2006.
- [61] P. Aschwanen and W. Guggenbuhl, "Experimental Results from a Comparative Study on Correlation-Type Registration Algorithms", in Robust Computer Vision, 1992, pp. 268-289.
- [62] J. Banks and P. Corke, "Quantitative Evaluation of Matching Methods and Validity Measures for Stereo Vision", The International Journal of Robotics Research, vol. 20(7), pp. 512-532, 2001.
- [63] T. H. Cormen, et al., "Introduction to algorithms", MIT press, 2001.
- [64] B. M. Maggs, "A Maximum Likelihood Stereo Algorithm", Computer Vision and Image Understanding, vol. 63(3), pp. 542-567, 1996.
- [65] 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, 1994, pp. 674-677.
- [66] S. Birchfield and C. Tomasi, "Depth Discontinuities by Pixel-to-Pixel Stereo", International Journal of Computer Vision, vol. 35(3), pp. 269-293, 1999.
- [67] P. N. Belhumeur, "A Bayesian Approach to Binocular Stereopsis", International Journal of Computer Vision, vol. 19(3), pp. 237-260, 1996.
- [68] S. Roy and I. J. Cox, "A Maximum-Flow Formulation of the N-Camera Stereo Correspondence Problem", in Proceedings of the Sixth International Conference on Computer Vision, Bombay, 1998, pp. 492-499.

- [69] H. Zhao, "Global Optimal Surface from Stereo", in Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, 2000, pp. 101-104.
- [70] I. Thomo, S. Malasiotis, and M. G. Strintzis, "Optimized Block Based Disparity Estimation in Stereo System Using a Maximum-Flow Approach", in Proceedings of the International Symposium on Computer Graphics, Image Processing, and Vision, Rio de Janeiro, 1998, pp. 410-417.
- [71] Y. Boykov, O. Veksler, and R. Zabih, "Fast Approximate Energy Minimization via Graph Cuts", Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 23(11), pp. 1222-1239, 2001.
- [72] V. Kolmogorov and R. Zabih, "Graph Cut Algorithms for Binocular Stereo with Occlusions", Springer, 2006.
- [73] D. Marr, "Vision: A Computational Investigation Into the Human Representation and Processing of Visual Information", MIT Press, 1982.
- [74] X. F. Huang, "Cooperative Optimization for Solving Large Scale Combinatorial Problems", Theory and Algorithms for Cooperative Systems, vol. 4(7), pp. 117-156, 2004.
- [75] X. Huang, "A General Framework for Constructing Cooperative Global Optimization Algorithms", Nonconvex Optimization and Its Applications, vol. 74, pp. 197-222, 2003.
- [76] X. Huang, "A General Global Optimization Algorithm for Energy Minimization from Stereo Matching", in ACCV, Korea, 2004, pp. 480-485.
- [77] C. L. Zitnick and T. Kanade, "A Cooperative Algorithm for Stereo Matching and Occlusion Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22(7), pp. 675-684, 2000.

- [78] Y. Zhang and C. Kambhamettu, "Stereo Matching with Segmentation-Based Cooperation", in Proceedings of the 7th European Conference on Computer Vision-Part II, 2002, pp. 556-571.
- [79] 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, 1993, pp. 34-40.
- [80] J. Sun, N. N. Zheng, and H. Y. Shum, "Stereo Matching using Belief Propagation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 25(7), pp. 787-800, 2003.
- [81] W. Faig, "Calibration of Close Range Photogrammetric Systems: Mathematical Formulation", Photogrammetric Engineering and Remote Sensing, vol. 41(12), pp. 1479-1486, 1975.
- [82] K. W. Wong, "Mathematical Formulation and Digital Analysis in Close-range Photogrammetry", Photogrammetric engineering and remote sensing, vol. 44(11), pp. 1355-1373, 1975.
- [83] O. D. Faugeras and G. Toscani, "Calibration Problem for Stereo", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1986, pp. 15-20.
- [84] S. Ganapathy, "Decomposition of Transformation Matrices for Robot Vision", Pattern Recognition Letters, vol. 2(6), pp. 401-412, 1984.
- [85] J. Weng, P. Cohen, and M. Herniou, "Camera Calibration with Distortion Models and Accuracy Evaluation", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14(10), pp. 965-980, 1992.
- [86] 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, 1990, pp. 246-253.

- [87] H. Bacakoglu and M. S. Kamel, "A Three-Step Camera Calibration Method", IEEE Transactions on Instrumentation and Measurement, vol. 46(5), pp. 1165-1172, 1997.
- [88] H. Gao, et al., "An Improved Two-Stage Camera Calibration Method", in Proceedings of the Sixth World Congress on Intelligent Control and Automation, Dalian, 2006, pp. 9514-9518.
- [89] 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, 1993, pp. 439-446.
- [90] Q. T. Luong and O. D. Faugeras, "Self-Calibration of a Moving Camera from Point Correspondences and Fundamental Matrices", International Journal of Computer Vision, vol. 22(3), pp. 261-289, 1997.
- [91] 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, pp. 238-249.
- [92] L. F. Cheong. 3D Computer Vision [Lecture Note]. Available: <http://couses.nus.edu.sg/course/electf/ee6901>
- [93] E. Boyer and M. O. Berger, "3D Surface Reconstruction Using Occluding Contours", International Journal of Computer Vision, vol. 22(3), pp. 219-233, 1997.

Appendices

Appendix A- the Snell's Law and 3D geometrical analysis

A.1. 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:

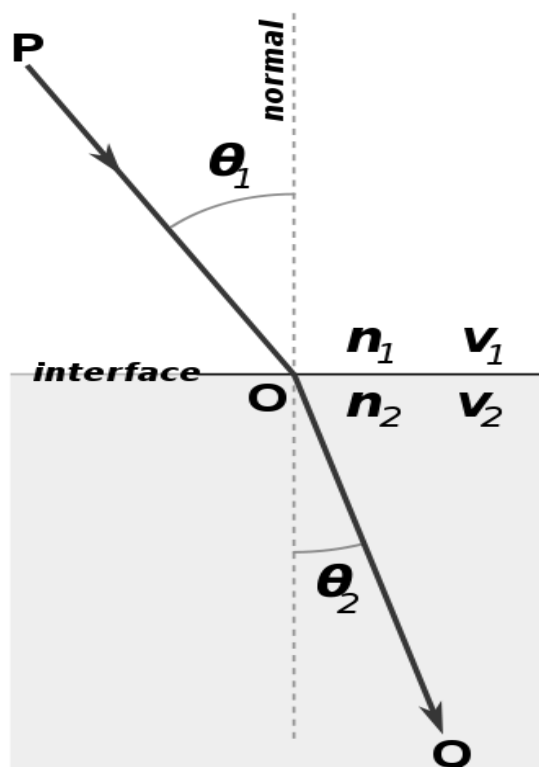


Fig. A.1: Demonstration of the Snell's Law

- (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 unitless) of the respective medium.

A.2 3D geometrical analysis

Any point, line, surface/plane can be expressed in 3D space through the geometrical analysis. It states that:

Straight line

A straight line in 3D space can be expressed as:

$$\frac{X-X_0}{X_1-X_0} = \frac{Y-Y_0}{Y_1-Y_0} = \frac{Z-Z_0}{Z_1-Z_0} , \quad (\text{A.1})$$

Where (X_0, Y_0, Z_0) and (X_1, Y_1, Z_1) are two arbitrary points located on the line.

A straight line in 3D space can also be expressed in the form of:

$$\frac{X-X_0}{A} = \frac{Y-Y_0}{B} = \frac{Z-Z_0}{C} . \quad (\text{A.2})$$

Where (X_0, Y_0, Z_0) is the arbitrary point located on the line and $(A, B, C)^T$ is the direction vector of the line.

The intersection angle α of any two lines in 3D space is defined as:

$$\cos \alpha = \frac{A'A'' + B'B'' + C'C''}{\sqrt{(A'^2 + B'^2 + C'^2)} * \sqrt{(A''^2 + B''^2 + C''^2)}} , \quad (\text{A.3})$$

with $(A', B', C')^T$ and $(A'', B'', C'')^T$ are the direction vectors of the two lines.

Plane/surface

By geometrical analysis, any plane in 3D space can be written as:

$$A_1(X - X_2) + B_1(Y - Y_2) + C_1(Z - Z_2) = 0 . \quad (\text{A.4})$$

Where $(A_1, B_1, C_1)^T$ is the normal vector of the plane and (X_2, Y_2, Z_2) is an arbitrary point on the plane.

The intersection angle β between any of the two planes is equal to:

$$\cos\beta = \frac{A'_1A''_1 + B'_1B''_1 + C'_1C''_1}{\sqrt{(A'^2_1 + B'^2_1 + C'^2_1)} * \sqrt{(A''^2_1 + B''^2_1 + C''^2_1)}} . \quad (\text{A.5})$$

Where $(A'_1, B'_1, C'_1)^T$ and $(A''_1, B''_1, C''_1)^T$ are the two normal vectors of the planes.

Appendix B- experimental results

B.1 Stereo correspondence and depth recovery by camera calibration

B.1.1 Stereo correspondence results at distance of 1400mm by camera calibration

Fig. B.1 is the stereo image taken by our single-lens stereovision system with the distance between camera and calibration board is 1400mm.

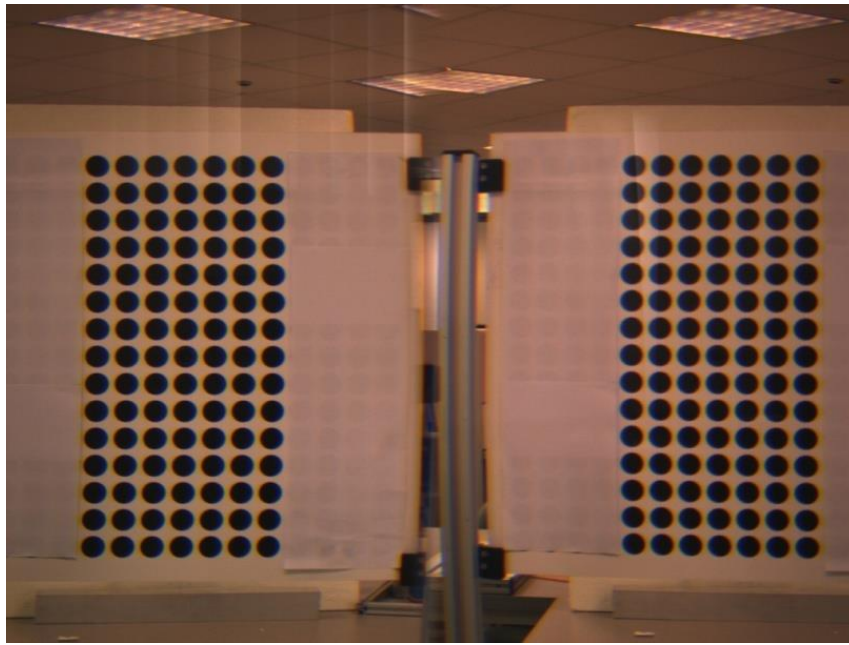


Fig. B.1: Experimental stereo image pair taken at distance 1400mm

We calibrate the left and right virtual cameras by following the discussions in Chapter III, the calibration parameters for left and right cameras are then as follows:

$$\begin{aligned} \begin{bmatrix} X_{lf} k_l \\ Y_{lf} k_l \\ k_l \end{bmatrix} &= \begin{bmatrix} f_l d'_{xl}{}^{-1} & 0 & C_{xl} \\ 0 & f_l d'_{yl}{}^{-1} & C_{yl} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}^L \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \\ \Rightarrow \begin{bmatrix} X_{lf} k_l \\ Y_{lf} k_l \\ k_l \end{bmatrix} &= \begin{bmatrix} 1960 & -10 & 270 & 153690 \\ -50 & 1800 & 400 & 264040 \\ 0 & 0 & 0 & 1370 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}, \end{aligned} \quad (\text{B.1})$$

and

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = \begin{bmatrix} f_r d'_{xr}^{-1} & 0 & C_{xr} \\ 0 & f_r d'_{yr}^{-1} & C_{yr} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{bmatrix} \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (\text{B.2})$$

$$\Rightarrow \begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = \begin{bmatrix} 1990 & -30 & 430 & 305740 \\ 40 & 1770 & 400 & 251820 \\ 0 & 0 & 0 & 1330 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}$$

We randomly select 20 points from the left image in Fig. B.2 and marked them in red circle. The results of the calculated corresponding right image points from Eq. (5.8) are shown in the figure in red square. The numerical values are listed in Table B.1.

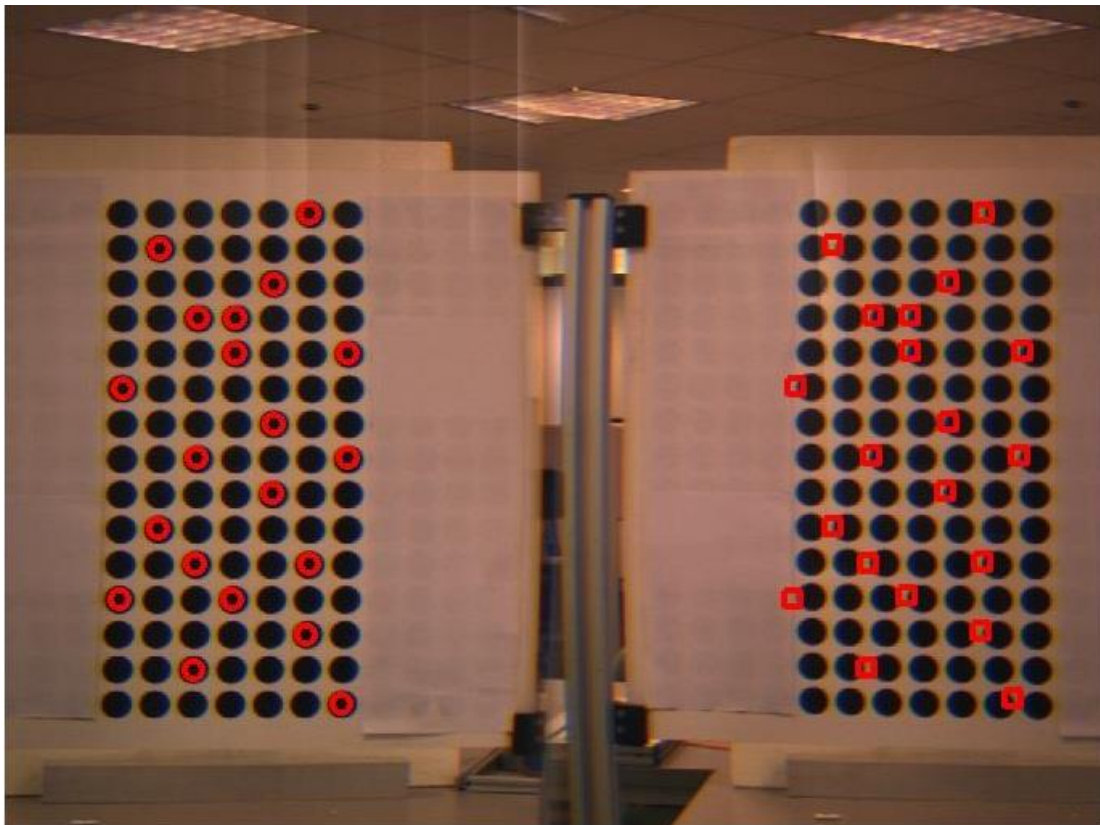


Fig. B.2: Display of the stereo correspondences (distance 1400mm)

Table B.1: Results of stereo correspondence of random 20 points at distance 1400mm

	X_{lf}	Y_{lf}	X_{rf}	X_{rf}^c	$ X_{rf} - X_{rf}^c $	Y_{rf}	Y_{rf}^c	$ Y_{rf} - Y_{rf}^c $
1	285	241	948	935	13	239	240	1
2	182	265	839	828	11	265	263	2
3	264	291	920	909	11	290	290	0
4	209	316	865	855	10	315	314	1
5	237	316	892	883	9	315	314	1
6	237	342	892	883	9	341	340	1
7	318	342	975	962	13	341	341	0
8	154	367	811	800	11	364	365	1
9	264	392	919	909	10	392	390	2
10	209	418	865	855	10	417	416	1
11	317	418	975	961	14	418	416	2
12	263	443	919	908	11	443	441	2
13	181	468	838	827	11	467	466	1
14	208	494	865	854	11	493	492	1
15	289	494	947	934	13	494	491	3
16	152	519	811	798	13	517	518	1
17	235	519	892	881	11	519	517	2
18	288	544	947	933	14	545	541	4
19	206	570	866	852	14	569	568	1
20	314	595	976	958	18	596	591	5

B.1.2 Stereo correspondence results at distance of 1800mm by camera calibration

Similarly, at the case of distance 1800mm, we have the stereo image pair taken in Fig. B.3. And the calibration results for both left and right virtual camera in Eq. (B.3) and Eq. (B.4), respectively.



Fig. B.3: Experimental stereo image pair taken at distance 1800mm

$$\begin{bmatrix} X_{lf}k_l \\ Y_{lf}k_l \\ k_l \end{bmatrix} = \begin{bmatrix} 1980 & -10 & 230 & 277470 \\ -40 & 1800 & 380 & 419310 \\ 0 & 0 & 0 & 1760 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (\text{B.3})$$

$$\begin{bmatrix} X_{rf}k_r \\ Y_{rf}k_r \\ k_r \end{bmatrix} = \begin{bmatrix} 1980 & -20 & 460 & 503630 \\ 30 & 1780 & 380 & 406880 \\ 0 & 0 & 0 & 1720 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \quad (\text{B.4})$$

The calculated stereo corresponding points are marked in Fig. B.4 (c.f. Eq. (5.9)) and the values display in Table B.2.

Table B.2: Results of stereo correspondence of random 20 points at distance 1800mm

	X_{lf}	Y_{lf}	X_{rf}	X_{rf}^c	$ X_{rf} - X_{rf}^c $	Y_{rf}	Y_{rf}^c	$ Y_{rf} - Y_{rf}^c $
1	290	241	948	935	13	239	240	1
2	182	265	839	828	11	265	263	2
3	264	291	920	909	11	290	290	0
4	209	316	865	855	10	315	314	1
5	237	316	892	883	9	315	314	1
6	237	342	892	883	9	341	340	1
7	318	342	975	962	13	341	341	0
8	154	367	811	800	11	364	365	1
9	264	392	919	909	10	392	390	2
10	209	418	865	855	10	417	416	1
11	317	418	975	961	14	418	416	2
12	263	443	919	908	11	443	441	2
13	181	468	838	827	11	467	466	1
14	208	494	865	854	11	493	492	1
15	289	494	947	934	13	494	491	3
16	152	519	811	798	13	517	518	1
17	235	519	892	881	11	519	517	2
18	288	544	947	933	14	545	541	4
19	206	570	866	852	14	569	568	1
20	314	595	976	958	18	596	591	5

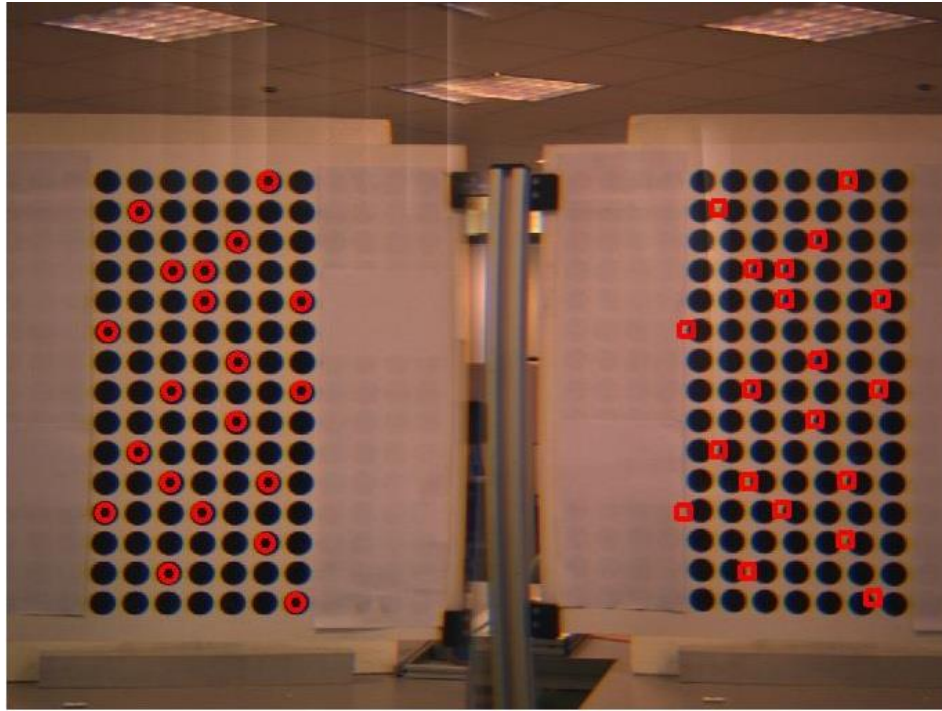


Fig. B.3: Display of the stereo correspondences (distance 1800mm)

B.1.3 Depth recovery by camera calibration at distance of 1400mm by camera calibration

Quantitative results of depth recovery where distance between calibration board and camera optical center is 1400mm are shown in Table B.3.

Table B.3: Results of depth recovery at distance of 1400mm by calibration based approach

	X_{lf}	Y_{lf}	X_{rf}^c	Y_{rf}^c	Z_c	Z_c^c	$ Z_c - Z_c^c $
1	285	195	914	195	1400	1367	33
2	145	227	774	224	1400	1411	11
3	251	260	881	259	1400	1347	53
4	181	292	811	290	1400	1368	32
5	216	292	846	290	1400	1381	19
6	216	325	846	323	1400	1380	20
7	321	325	950	324	1400	1330	70
8	110	358	739	356	1400	1444	44

9	251	391	881	389	1400	1342	58
10	180	423	810	421	1400	1362	38
11	320	423	948	421	1400	1327	73
12	250	456	879	454	1400	1341	59
13	144	489	773	488	1400	1398	2
14	178	522	807	521	1400	1360	40
15	284	522	913	520	1400	1355	45
16	107	554	736	554	1400	1435	35
17	213	554	842	552	1400	1371	29
18	282	587	911	584	1400	1352	48
19	176	620	805	619	1400	1355	45
20	314	652	942	648	1400	1320	80

B.1.4 Depth recovery by camera calibration at distance of 1800mm by camera calibration

Table B.4 below is our depth recovery results obtained at distance of 1800mm.

Table B.4: Results of depth recovery at distance of 1800mm by calibration based approach

	X_{lf}	Y_{lf}	X_{rf}^c	Y_{rf}^c	Z_c	Z_c^c	$ Z_c - Z_c^c $
1	290	241	935	240	1800	1737	63
2	182	265	828	263	1800	1811	11
3	264	291	909	290	1800	1750	50
4	209	316	855	314	1800	1771	29
5	237	316	883	314	1800	1733	67
6	237	342	883	340	1800	1732	68
7	318	342	962	341	1800	1731	69
8	154	367	800	365	1800	1848	48
9	264	392	909	390	1800	1796	4
10	209	418	855	416	1800	1767	33

11	317	418	961	416	1800	1729	71
12	263	443	908	441	1800	1795	5
13	181	468	827	466	1800	1803	3
14	208	494	854	492	1800	1764	36
15	289	494	934	491	1800	1759	41
16	152	519	798	518	1800	1842	58
17	235	519	881	517	1800	1726	74
18	288	544	933	541	1800	1757	43
19	206	570	852	568	1800	1762	38
20	314	595	958	591	1800	1723	77

B.2 Stereo correspondence and depth recovery by ray sketching

B.2.1 Stereo correspondence results at distance of 1400mm by ray sketching

Table B.5 shows the stereo correspondence results by ray sketching based approach with the distance between calibration board and camera optical center is 1400mm. The stereo image pair in Fig. B.1 is used for this experiment.

Table B.5: Result of stereo correspondence by ray sketching based approach at 1400mm

	X_{lf}	Y_{lf}	X_{rf}	X_{rf}^c	$ X_{rf} - X_{rf}^c $	Y_{rf}	Y_{rf}^c	$ Y_{rf} - Y_{rf}^c $	
1	285	195	928	929	1	194	198	4	
2	145	227	789	788	1	226	232	6	
3	251	260	891	892	1	259	267	7	
4	181	292	821	822	1	291	297	6	
5	216	292	856	856	0	291	298	7	
6	216	325	855	856	0	324	329	5	
7	321	325	961	962	1	325	328	3	
8	110	358	752	752	0	356	361	5	

9	251	391	890	891	1	390	391	1	
10	180	423	820	821	1	422	423	1	
11	320	423	961	925	36	423	424	1	x
12	250	456	890	890	0	456	455	1	
13	144	489	786	787	1	487	489	2	
14	178	522	820	821	1	520	520	0	
15	284	522	926	927	1	522	521	1	
16	107	554	752	753	1	552	554	2	
17	213	554	855	856	1	554	555	1	
18	282	587	926	928	2	587	587	0	
19	176	620	822	928	106	618	618	0	x
20	314	652	963	931	32	653	652	1	x

B.2.2 Stereo correspondence results at distance of 1800mm by ray sketching

Table B.6 shows the stereo correspondence results by ray sketching with the distance between calibration board and camera optical center is 1400mm and the image pair for experiment is Fig. B.3.

Table B.6: Result of stereo correspondence by ray sketching based approach at 1800mm

	X_{lf}	Y_{lf}	X_{rf}	X_{rf}^c	$ X_{rf} - X_{rf}^c $	Y_{rf}	Y_{rf}^c	$ Y_{rf} - Y_{rf}^c $	
1	290	241	948	949	1	239	244	5	
2	182	265	839	840	1	265	264	1	
3	264	291	920	921	1	290	294	4	
4	209	316	865	867	2	315	319	4	
5	237	316	892	893	1	315	320	5	
6	237	342	892	893	1	341	346	5	
7	318	342	975	978	3	341	341	0	

8	154	367	811	813	2	364	370	6	
9	264	392	919	920	1	392	392	0	
10	209	418	865	864	1	417	419	2	
11	317	418	975	976	1	418	417	1	
12	263	443	919	919	0	443	443	0	
13	181	468	838	837	1	467	467	0	
14	208	494	865	866	1	493	493	0	
15	289	494	947	947	0	494	494	0	
16	152	519	811	811	0	517	518	1	
17	235	519	892	892	0	519	519	0	
18	288	544	947	948	1	545	544	1	
19	206	570	866	865	1	569	568	1	
20	314	595	976	976	0	596	596	0	

B.2.3 Depth recovery by ray sketching at distance of 1400mm

With the stereo correspondence results gained in Table B.5, we have the depth recovery results by ray sketching as shown below:

Table B.7: The recovered depth of points at distance of 1400mm by ray sketching based approach

	X_{lf}	Y_{lf}	X_{rf}^c	Y_{rf}^c	Z_c	Z_c^c	$ Z^c - Z_c^c $
1	285	195	929	198	1400	1459	59
2	145	227	788	232	1400	1449	49
3	251	260	892	267	1400	1466	66
4	181	292	822	297	1400	1481	81
5	216	292	856	298	1400	1438	38
6	216	325	856	329	1400	1486	86
7	321	325	962	328	1400	1478	78

8	110	358	752	361	1400	1411	11
9	251	391	891	391	1400	1499	99
10	180	423	821	423	1400	1443	43
11	320	423	925	424	1400	1290	110
12	250	456	890	455	1400	1468	68
13	144	489	787	489	1400	1440	40
14	178	522	821	520	1400	1493	93
15	284	522	927	521	1400	1485	85
16	107	554	753	554	1400	1436	36
17	213	554	856	555	1400	1480	80
18	282	587	928	587	1400	1466	66
19	176	620	928	618	1400	1789	389
20	314	652	931	652	1400	1285	115

$$\text{Error Percentage} = \sum |Z_c - Z_c^c| / 20 / Z_c * 100\% = 6.04\%$$

$$\text{For matching points, error percentage} = \sum |Z_c - Z_c^c| / 18 / Z_c * 100\% = 4.53\%$$

B.2.4 Depth recovery by ray sketching at distance of 1800mm

Table B.6 shows the depth recovery results by ray sketching with the stereo correspondence results gained in Table B.8.

Table B.8: The recovered depth of points at distance of 1800mm by ray sketching based approach

	X_{lf}	Y_{lf}	X_{rf}^c	Y_{rf}^c	Z_c	Z_c^c	$ Z_c^c - Z_c^c $
1	290	241	949	244	1800	1888	88
2	182	265	840	264	1800	1862	62
3	264	291	921	294	1800	1907	107
4	209	316	867	319	1800	1930	130

5	237	316	893	320	1800	1898	98
6	237	342	893	346	1800	1841	41
7	318	342	978	341	1800	1853	53
8	154	367	813	370	1800	0	0
9	264	392	920	392	1800	1881	81
10	209	418	864	419	1800	1852	52
11	317	418	976	417	1800	1843	43
12	263	443	919	443	1800	1829	29
13	181	468	837	467	1800	1856	56
14	208	494	866	493	1800	1838	38
15	289	494	947	494	1800	1850	50
16	152	519	811	518	1800	1884	84
17	235	519	892	519	1800	1845	45
18	288	544	948	544	1800	1872	72
19	206	570	865	568	1800	1867	67
20	314	595	976	596	1800	1894	94

All the points are matched, error percentage = $\sum |Z_c - Z_c^c| / 18 / Z_c * 100\% = 3.58\%$

Publications

Journal papers:

1. K.B. Lim, M. J. Zhao, "Stereo Matching of Single-lens Bi-Prism Based Stereovision System", *Journal of Procedia Engineering*, ISSN: 1877-7058, ELSEVIER, 2011
2. M. J. Zhao, K. B. Lim, "Geometrical-Analysis-Based Algorithm for Stereo Matching of Single-lens Binocular and Multi-ocular Stereovision System", *Journal of Electronic Science and Technology*, Vol. 10, No. 2, JUNE, 2012

Conference papers:

1. M. J. Zhao, K. B. Lim, "*Geometrical-Analysis-Based Algorithm for Stereo Matching of Single-lens Binocular and Multi-ocular Stereovision System*", International Conference of Signal, Image Processing and Application, Hong Kong, 2012
2. M. J. Zhao, K. B. Lim, "*Stereo Matching of Single-lens Bi-prism based Stereovision System*", International Conference of ICFIT, Changsha, China, 2010
3. M. J. Zhao, K. B. Lim, "*Stereo Correspondence Problem of Single-lens Bi-Prism Stereovision System*", International Conference of ISDM, Wuhan, China, 2009