

Thesis Submitted for the Degree of  
Doctor of Philosophy

GENERALISING THE IDEAL  
PINHOLE MODEL TO MULTI-PUPIL  
IMAGING FOR DEPTH RECOVERY

Brendan P. Byrne, B.Eng.



Supervisor: Professor Paul F. Whelan

Dublin City University  
School of Electronic Engineering  
September 2012

*I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy is entirely my own work, that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.*

Signed: \_\_\_\_\_  
*Candidate*

ID No.: 52597055

Date: 1<sup>st</sup> September 2012

## Acknowledgement

Firstly, I would like to extend my sincere gratitude to my supervisor Prof. Paul Whelan, who initially afforded me the opportunity to undertake this PhD, and whose continuous support, guidance, and advice have benefited me greatly. I would also like to extend further thanks to Dr. John Mallon for his direction and advice.

I wish to offer a special thanks to my colleague and friend Dr. Aubrey Dunne who, over the course of this study, has provided me with many stimulating discussions on the various topics within this thesis. His meticulous attention to detail, work ethic and exceptional patience are admirable.

I have enjoyed my time working in the Centre for Image Processing and Analysis where my colleagues have always been willing to help and offer advice as well as their friendship. For this I am grateful and would like to thank Seán, Tony, Patricia, Tarik, Laura, Stephen, Ovidiu and Julia. I would also like to thank my friends in the Optics lab Karl and Kai.

My Mam and Dad, Emily and Seán, have always maintained belief in my work while at the same time providing much needed encouragement, kindness and unconditional love and support. For this I am eternally grateful. I would like to thank my sisters Mary, Lisa, brother-in-law John and extended family who have always been supportive and willing to join me in celebrating successes. I also wish to thank Marian, Frank, Karen and Caroline for their support and Frankee who remained joyfully unaware of the pleasant distractions he provided.

Finally, to my girlfriend Louise, whose unconditional support and unfaltering patience has been an inspiration throughout the highs and lows. This work wouldn't have been possible without your love and kindness. I can't express how grateful I am to you.

Acknowledgements are also due to IRCSET, the Irish Research Council for Science, Engineering and Technology, and the Irish State, under the National Development Plan, for funding my PhD studies.

*To Louise, Emily and Seán*

# Contents

Acknowledgement . . . . .	i
Abstract . . . . .	vi
Glossary of Acronyms . . . . .	vii
<b>1 Introduction</b>	<b>1</b>
1.1 Background and Motivation . . . . .	3
1.1.1 Camera Models and Calibration . . . . .	3
1.1.2 Modified Optics - Calibration and Depth Estimation . . . . .	4
1.1.3 Thesis Goals . . . . .	6
1.2 Literature Review . . . . .	7
1.2.1 Camera Calibration . . . . .	7
1.2.2 Modified Optics in Computer Vision . . . . .	14
1.3 Mathematical Notation . . . . .	21
1.4 Contributions . . . . .	22
1.5 Thesis Organisation . . . . .	24
<b>2 Camera Models and Calibration</b>	<b>26</b>
2.1 Camera Models . . . . .	28
2.1.1 Pinhole Camera . . . . .	29
2.1.2 Thin Lens Camera . . . . .	31
2.1.3 Thick Lens Camera . . . . .	35
2.2 Camera Calibration . . . . .	37
2.2.1 Planar Camera Calibration . . . . .	38
Zhang’s Approach . . . . .	39
Alternative Approach . . . . .	43
2.3 Significance of the Pupil Plane . . . . .	46
2.3.1 Location of Pupil Plane . . . . .	49
2.3.2 Location of Pupil on Pupil Plane . . . . .	56
2.4 Discussion . . . . .	62
<b>3 Efficient Camera Calibration</b>	<b>65</b>
3.1 Camera Configurations . . . . .	67
3.1.1 Degenerate Configurations . . . . .	69
3.2 Optimal Image Networks . . . . .	73
3.2.1 Generated Image Networks . . . . .	74
Forming $\hat{H}$ . . . . .	75
Constraints on $\hat{H}$ . . . . .	76
Network orientation and sensitivity . . . . .	79
3.2.2 Selected Image Networks . . . . .	80

Search Algorithm Implementation . . . . .	82
3.3 Experiments . . . . .	83
3.3.1 Selected Image Networks . . . . .	84
Convergence of Image Networks . . . . .	84
Reduced Number of Images . . . . .	85
Distortion Correction . . . . .	86
3.3.2 Generated Image Networks . . . . .	87
Synthetic testing . . . . .	87
Image reproducibility . . . . .	89
Image Network Stability . . . . .	91
3.4 Discussion . . . . .	93
<b>4 Multi-Pupil Imaging</b>	<b>96</b>
4.1 Multi-Pupil Camera Model . . . . .	98
4.1.1 Double Pupil Imaging Model . . . . .	99
4.1.2 Simulated Experiments . . . . .	102
Zemax Verification . . . . .	103
Model Characterisation . . . . .	104
Depth Estimation . . . . .	110
4.2 Shifted Multi-Pupil Camera Model . . . . .	113
4.2.1 Double Pupil Shifted Imaging Model . . . . .	116
4.2.2 Simulated Experiments . . . . .	118
Zemax Verification . . . . .	118
Model Characterisation . . . . .	119
Depth Estimation . . . . .	125
4.3 Calibration of the Pupil Model . . . . .	128
4.3.1 Standard Pinhole Calibration Methods . . . . .	129
4.3.2 DPM Calibration Method . . . . .	130
Simulated Experiments . . . . .	132
4.3.3 DPSM Calibration Method . . . . .	137
Simulated Experiments . . . . .	140
4.4 Real Experiments . . . . .	146
4.4.1 Experimental Setup . . . . .	147
4.4.2 Calibration Procedure . . . . .	150
Standard Calibration Approaches . . . . .	152
Multi-pupil Calibration . . . . .	152
Discussion . . . . .	155
4.4.3 Depth Estimation . . . . .	157
Discussion . . . . .	160
4.5 Discussion . . . . .	163
<b>5 Aberration Modelling, Calibration and Removal</b>	<b>168</b>
5.1 Optical Aberrations . . . . .	169
5.1.1 Aberration Formation . . . . .	169
Spherical Aberration . . . . .	171
Coma . . . . .	173
Astigmatism and Field Curvature . . . . .	174
Distortion . . . . .	175

5.2	Calibration . . . . .	175
5.2.1	Multi-Pupil Formation . . . . .	176
5.2.2	Multi-Pupil Calibration . . . . .	178
5.3	Experiments . . . . .	181
5.3.1	Experimental Setup . . . . .	182
5.3.2	Depth Estimation . . . . .	183
5.4	Discussion . . . . .	186
<b>6</b>	<b>Conclusions and Future Work</b>	<b>190</b>
6.1	Thesis Contributions . . . . .	191
6.1.1	Efficient Camera Calibration . . . . .	192
6.1.2	Multi-Pupil Imaging . . . . .	193
6.1.3	Aberration Modelling, Calibration, and Removal . . . . .	195
6.2	Publications Arising . . . . .	196
6.3	Directions for Future Research . . . . .	196
6.3.1	Camera Calibration . . . . .	197
6.3.2	Multi-Pupil Imaging . . . . .	197
<b>A</b>	<b>Additional OIN Results</b>	<b>A-1</b>
	<b>Bibliography</b>	

# Generalising the Ideal Pinhole Model to Multi-Pupil Imaging for Depth Recovery

Brendan P. Byrne

## Abstract

This thesis investigates the applicability of computer vision camera models in recovering depth information from images, and presents a novel camera model incorporating a modified pupil plane capable of performing this task accurately from a single image. Standard models, such as the ideal pinhole, suffer a loss of depth information when projecting from the world to an image plane. Recovery of this data enables reconstruction of the original scene as well as object and 3D motion reconstruction. The major contributions of this thesis are the complete characterisation of the ideal pinhole model calibration and the development of a new multi-pupil imaging model which enables depth recovery. A comprehensive analysis of the calibration sensitivity of the ideal pinhole model is presented along with a novel method of capturing calibration images which avoid singularities in image space. Experimentation reveals a higher degree of accuracy using the new calibration images. A novel camera model employing multiple pupils is proposed which, in contrast to the ideal pinhole model, recovers scene depth. The accuracy of the multi-pupil model is demonstrated and validated through rigorous experimentation. An integral property of any camera model is the location of its pupil. To this end, the new model is expanded by generalising the location of the multi-pupil plane, thus enabling superior flexibility over traditional camera models which are confined to positioning the pupil plane to negate particular aberrations in the lens. A key step in the development of the multi-pupil model is the treatment of optical aberrations in the imaging system. The unconstrained location and configuration of the pupil plane enables the determination of optical distortions in the multi-pupil imaging model. A calibration algorithm is proposed which corrects for the optical aberrations. This allows the multi-pupil model to be applied to a multitude of imaging systems regardless of the optical quality of the lens. Experimentation validates the multi-pupil model's accuracy in accounting for the aberrations and estimating accurate depth information from a single image. Results for object reconstruction are presented establishing the capabilities of the proposed multi-pupil imaging model.

# Glossary of Acronyms

<b>Acronym</b>	–	<b>Explanation</b>
1D	–	One Dimensional
2D	–	Two Dimensional
3D	–	Three Dimensional
BA	–	Bundle Adjustment
BFL	–	Back Focal Length
CC	–	Centre Circle
CCD	–	Charged Coupled Device
CL	–	Centre Line
DfD	–	Depth from Defocus
DfF	–	Depth from Focus
DLT	–	Direct Linear Transform
DPM	–	Double Pupil Model
DPSM	–	Double Pupil Shifted Model
DSLR	–	Digital Single Lens Reflex
EO	–	Edmund Optics
FFL	–	Front Focal Length
FOV	–	Field of View
GIN	–	Generated Image Network
IAC	–	Image of the Absolute Conic
IN	–	Image Network
LCD	–	Liquid Crystal Display
LCoS	–	Liquid Crystal on Silicon
LM	–	Levenberg-Marquardt
MFT	–	Micro Four Thirds
MLE	–	Maximum Likelihood Estimation
OIN	–	Optimal Image Network
PSF	–	Point Spread Function
RMSE	–	Root Mean Square Error
SD	–	Standard Deviation
SIN	–	Selected Image Network
SVD	–	Singular Value Decomposition
VL	–	Vanishing Line
VP	–	Vanishing Point

# Chapter 1

## Introduction

One of the primary goals of computer vision is to retrieve metric information from images. This has remained a constant goal despite the technological advances in the field of imaging. In fact, as a result of improving technology, the physical size of image sensors has decreased whilst the on board processing power has increased. This has enabled access to a host of new areas to implement imaging system solutions which were not previously possible due to physical and computing constraints. Therefore the field is constantly developing new imaging techniques which take advantage of these technological advances to meet the requirements of new application areas. This thesis sits firmly in this area of developing new imaging models which build on the technological advances of the imaging field. More specifically, a new imaging framework is developed for the purpose of retrieving metric information from images. This is achieved within the realm of modifying optical properties of a conventional imaging system.

Naturally, the problems which are fundamental to the recovery of metric information from images must be addressed. This is primarily, calibration of the camera model. Calibration is the process in which the camera model parameters, which describe the world to image projection, are estimated. Furthermore, an increasing trend amongst manufacturers within the area is to produce ultra-compact imaging solutions. This has a direct impact on the quality of the optical systems within the package and as such, a decrease in quality with miniaturisation is being observed. Therefore in addition to estimating camera model parameters, compensation for optical quality of the imaging system is

required. These are all problems which are addressed in this thesis.

Cameras which are modelled as pinhole describe the projection of 3D world points to 2D sensor points through a single centre of projection. In order to recover the metric information of this process, such as distances between object points, knowledge of the model parameters which describe the projection is required. Naturally, the recovery of camera model parameters is achieved by a calibration process. Typically this process entails imaging a calibration object of known geometry, and subsequently calculating model parameters based on the image to world point relationship. Understandably, a large amount of research has been directed at this problem as camera calibration is a process which is fundamental to all computer vision tasks which require metric information. To this end, many approaches have been investigated, however, it is the planar methods which have taken precedence within the computer vision community due to the relaxation on input requirements coupled with readily available implementations. Consequently, feature points are captured by imaging planar targets in the form of chessboard grids. Although the majority of existing planar calibration methods are based on the same geometric constraints, the focus has been on improving camera parameter estimates from an algorithmic aspect. Thus, the effects on accuracy due to the geometric configuration of the input planar targets has largely been neglected. With the exception of known degenerate configurations, which result in parameter estimation failure, the overall impact of planar target orientation on camera parameter estimation is an area which warrants further investigation. This goal forms the basis for the initial section of this thesis.

The pinhole model is applicable to a multitude of imaging systems, however, within the constraints of estimating object depth from a single image, additional cues are required to obtain depth information. Object depth is lost within the process of central projection, therefore two pinhole cameras in a stereo configuration is the standard method to recover depth information. The natural progression of the pinhole model to lens models is outlined in this thesis. Generally, approaches which estimate depth using a single camera impose lens models which allow image phenomena such as focus and blur to be modelled. These additional cues enable the recovery of object depth information. However, multiple images with varying camera parameter settings is typical to this process. Recently, techniques which modify the imaging process in the optical domain have been promoted as an alternative method to recover depth

from a single image. Typically, these modifications make use of lens properties to encode additional information about the scene in the image. One such approach is modification of the pupil plane within the imaging system. This thesis initially investigates the sensitivity of such modifications firstly in terms of pupil plane location within the optical system and secondly, in terms of the placement of a pupil within the pupil plane. Subsequently, a new multi-pupil model is proposed for estimating depth from a single image. This requires a novel calibration method and the ability to account for optical aberrations present in the imaging lens.

## **1.1 Background and Motivation**

This section gives a brief overview of the principal concepts and ideas that are addressed in this thesis. The aim is to highlight the significance of camera calibration and the evolution of camera models incorporating modified optics for the purpose of depth recovery.

### **1.1.1 Camera Models and Calibration**

A camera captures information from 3D world space and subsequently projects this to a 2D representation in the form of an image. Modelling of the imaging process is a fundamental task of computer vision. Many camera models have been proposed for this purpose as no single model fits all prospective applications. The pinhole camera model is the most basic model for perspective cameras. Additionally, it is the most used geometric camera model in the field of computer vision. A single centre of projection through which all scene rays must pass is assumed. Naturally, the theoretical pinhole represents the ideal case for true perspective projection, and the closer the camera is to this model, the better the performance for certain tasks. However, imperfections within the optical system of imaging lenses leads to deviation from this projection, and as such, models which incorporate optical properties are of paramount importance to accurately represent the imaging process.

Calibration of a camera is the process in which the parameters that form the camera model are recovered. Since the model represents the projection of 3-

space points to image space pixel locations, knowledge of these parameters allows metric measurement and analysis of the information available in the camera image. This is a primary objective for many computer vision applications. There are many approaches to calibrating a camera. Although there are methods which calibrate a camera based on natural scene constraints, known as self-calibration, a scale ambiguity remains in the determination of metric information. Therefore, camera calibration employing targets with known geometry are dealt with in this thesis. The calibration target provides a set of feature points in 3-space which are subsequently imaged by the camera. Camera calibration methods have been proposed for 1D, 2D, and 3D targets. Issues in finding exact 3D locations of feature points has led to techniques using 2D planar targets, of a chessboard nature, being most commonly used in the computer vision community. Generally, these 2D methods are formed under the same geometric constraints. Therefore, the impact of the geometric configuration of planar targets on calibration accuracy has received less attention. This has implications on the input requirements of the practitioner as well as on the accuracy in model parameter determination. This problem is thoroughly investigated in this thesis and results in a novel approach to capturing calibration images with improved geometrical properties.

### 1.1.2 Modified Optics - Calibration and Depth Estimation

With the successful application of a calibration algorithm, accurate model parameters allow tasks such as object depth estimation to be completed. The standard configuration for depth estimation employing a pinhole camera model is that of a stereo vision system. This requires knowledge of the camera offset or baseline coupled with correct alignment of the images captured from each camera. Typical implementations of stereo vision systems are seen with 3D broadcasting of sporting events, and synthetically in the case of generating scenes for 3D animated movies. It is clear that one major shortcoming of the pinhole model is its inability to passively estimate depth from a single image using a single sensor.

In response to these shortcomings, optical properties have been introduced to the camera model in the form of lens models. Additional image artifacts as a

result of deviation from pinhole imaging are retrievable with the inclusion of a lens in the camera model. One such artifact is image blur, which is a product of the aperture diameter and object depth. Depth dependent blur is a well known cue for passive object depth recovery. However the calibration process requires multiple images and camera settings in order to successfully estimate object depth. Additionally, properties which require accurate measurement, such as image edges in the presence of blur, are known to be error prone. Furthermore, there are inherent ambiguities in estimating depth from a single defocused image. Depending on the optical configuration of the imaging system, an object depth which lies a certain distance behind the focal plane will exhibit an equal amount of blur to an object located at a certain distance in front of the focal plane. The second ambiguity relates to measuring blur from edge features. It is difficult to ascertain whether the measured blur is due to a strong edge which is out of focus or a weak edge which is in focus.

Active methods are an alternative to retrieving depth information from a single image. Generally, these methods employ an additional source to aid in the calculation of depth. Examples of such methods are time of flight cameras (Gokturk et al., 2004) which emit infrared pulses and calculate resulting depth information based on the duration between emitting and sensing. The Microsoft Kinect is another example of monocular depth recovery from a single image<sup>1</sup>. An infrared camera is used in conjunction with an infrared projector. Properties of the projected pattern on the scene are used to infer depth information. The main issue with active methods is the requirement of additional sources to fuse with image data, which also complicates the calibration process.

Due to increased processing power and an increased number of imaging pixels, modifying the optical system within a camera is a recent trend which has surfaced in the field of computer vision. The most common modification of these camera models is the intentional alteration of the optical path between scene points and the image sensor. Resulting non-conventional images are thus encoded based on the geometry of these modifications. Therefore additional image cues, which are not available from a single conventional image, can be coded in the image data for subsequent retrieval. Camera models with modified optics have been successfully implemented to complete various computer

---

<sup>1</sup><http://www.xbox.com/en-ie/kinect> accessed November 2011

vision tasks, such as scene depth recovery, single-image variable focus, super-resolution, and extended depth of field. Examples of modified camera models for the purpose of depth recovery are seen with coded apertures (Levin et al., 2007) and with systems implementing diffusers (Cossairt et al., 2010). These methods can also extend the depth of field by post processing the image data. Plenoptic cameras (Adelson and Wang, 1992, Ng et al., 2005, Lumsdaine and Georgiev, 2009), which modify the optical path with a microlens array placed near the sensor, have post-processing applications such as variable focus and super resolution. Recently, the plenoptic work developed by Ng et al. (2005) was commercialised as the Lytro camera<sup>2</sup>. Additionally, a recently formed company<sup>3</sup> offer custom modified optical solutions for various applications such as medical endoscopes, 3D broadcasting and military equipment.

There are clear benefits with the implementation of imaging systems with modified optical properties. Modern processing capabilities allow additional operations to be performed on the captured image. Therefore, new cues for common computer vision tasks, such as depth recovery, can be availed of. However, a common theme amongst these new imaging modalities is difficulty in construction and system calibration. Many of the modifications occur within the system of lens elements or near the image sensor, which inherently requires a high level of accuracy as these areas are most sensitive to optical variation. Since the resulting images are tailored based on optical modification, standard camera calibration procedures are not applicable. These difficulties are addressed in this thesis with the proposition and calibration of a new camera model with modified optical properties.

### 1.1.3 Thesis Goals

The work presented in this thesis is motivated by the fast changing environment in which imaging systems are currently being developed. Smaller image sensors with increased processing power are enabling new approaches to solve the problem of metric information retrieval from images. The primary goal of this work is to develop a new imaging framework, building on a non-conventional approach, which can estimate metric information from a single image. In order to realise this goal, a better understanding of current camera models and their

---

<sup>2</sup>[www.lytro.com](http://www.lytro.com) accessed: October 2011

<sup>3</sup><http://www.isee3d.com> accessed: November 2011

calibration is required as well as the development of novel methods to deal with the new camera model.

Accordingly, the goals of the presented research are

- to evaluate current camera models and their calibration to identify characteristics suitable for depth estimation;
- to develop a new camera model that is applicable to current and emerging imaging modalities and is capable of retrieving metric information from a single image;
- to develop novel and practical calibration methods for the new camera models;
- to demonstrate the suitability and applicability of the new camera model to the modelling, calibration and removal of optical aberrations;

## 1.2 Literature Review

Since the advent of digital cameras, the topic of camera calibration has possibly received more attention than any other aspect of computer vision. With such a large body of work, the following literature survey concentrates on the significant publications which are directly related to the stated goals of the thesis. The following review is conducted under headings which correspond to the main themes within this thesis.

### 1.2.1 Camera Calibration

Camera calibration is a fundamental task of computer vision. Accordingly, it has received much attention and many approaches have been proposed. Generally the process involves imaging a calibration target with known geometry. Within the scope of this thesis, the primary interest is in the configuration of calibration targets and the impact this has on camera parameter estimates. This review presents calibration approaches employing targets of one, two, and three dimensions. The key problems when employing each form of target are outlined and discussed.

There have been numerous methods proposed to calibrate a camera with the earliest recorded work in the area of photogrammetry. Sutherland (1964) developed a system which inferred user drawings into a graphical format. The perspective transformation between the 3D world coordinates and 2D image coordinates from point correspondence was formulated. However, accuracy of this method was not reported as its application was in the graphics domain. This work led to the introduction of linear calibration methods. Abdel-Aziz and Karara (1971) introduced the Direct Linear Transformation (DLT) which provided a linear solution to solving the camera parameters given the world and image coordinate correspondence. The main disadvantage of the linear methods was that lens distortion could not be solved linearly. Additionally, the number of unknowns being estimated was greater than the degrees of freedom within the projection matrix. Therefore, the effect of noise within the imaging system or within the measurement of world point coordinates would have a detrimental effect on accuracy of camera parameter recovery. Subsequent nonlinear methods were proposed which minimised the reprojection error equation. This is the geometric distance between the observed image point and the projection of the world point onto the image (via the projection matrix being estimated). Distortion models were included in the nonlinear minimisation of the reprojection error. The photogrammetric community coined the term “Bundle Adjustment” (BA) for this particular type of calibration (Slama et al., 1980). Other non-linear methods used the linear solution as an initial estimate for the camera parameters which were then refined with BA (Hartley and Zisserman, 2003).

The evolution of camera calibration algorithms resulted in a greater emphasis on the type of calibration object used. Calibration objects of one, two and three dimensions have been proposed. Thus, depending on the type of calibration object employed, various approaches are taken to solve the calibration problem. Classical calibration algorithms in the field of computer vision began to surface in the late 1980s. These algorithms employed 3D calibration targets which were precisely manufactured. Tsai (1987) proposed a calibration approach using a 3D calibration rig consisting of two orthogonal planar facets in which the control point positions were accurately known. Tsai used a radial alignment constraint to decouple the camera intrinsic and extrinsic parameters which enabled a two-stage approach to estimating the camera parameters along with a single parameter radial model for distortion. The initial step in-

volved linearly estimating the camera extrinsic parameters. Subsequently, the focal length (principal point was assumed given by manufacturer) and single radial distortion term were estimated using a non-linear optimisation scheme.

Weng et al. (1992) proposed a calibration technique which considered additional distortion parameters in the form of tangential and thin-prism distortion. Some of the assumptions made in Tsai’s algorithm, such as an assumed principal point, were addressed. Weng’s approach was also completed in two stages. Initially, a closed form solution for the external parameters and some internal parameters was formed based on a distortion free camera model. Image points close to the centre of the image were used for this initial calibration stage. These points are least affected by distortion and thus provide a reasonable approximation of a distortion free camera model. The second step was initialised with the closed form parameter estimates and subsequently refined, via non-linear optimisation, based on a camera model incorporating distortion.

A comprehensive analysis of the early calibration algorithms employing 3D calibration objects is presented in Salvi et al. (2002). It is shown that the performance of both Tsai (1987) and Weng et al. (1992) achieve comparable accuracy. However, for lenses which suffer greater distortion, such as lenses with larger fields of view, the method of Weng et al. (1992) performs better. This is due to its more accurate distortion model compared to Tsai’s, which only models radial distortion.

Alternative camera calibration methods employing similar 3D targets, but without the requirement of exact 3D control point localisation, were proposed using Vanishing Points (VP) and Vanishing Lines (VL). In the seminal work of Caprile and Torre (1990), which was aimed at calibrating a stereo system, the idea of using vanishing points for camera calibration was introduced. The calibration target consisted of an aluminum cube with straight line segments painted on three of the mutually orthogonal facets. Camera intrinsic parameters were calculated based on the property that the three vanishing points of the lines on the planar facets of the cube form the vertices of a triangle whose orthocentre is the intersection of the optical axis with the image plane. This approach is valid provided a zero skew camera is assumed with a known aspect ratio. Recovery of the principal point enabled calculation of the focal length and camera rotation matrix based on the VP constraint. The remaining extrinsic parameters were recovered using properties of the calibration target

in both cameras.

A method based on similar principles to Caprile and Torre (1990) was proposed by Cipolla et al. (1999). The principal point is calculated as the orthocentre of the triangle containing the VPs which allows subsequent calculation of the focal length. The main difference with Cipolla’s method is that the calibration target information is retrieved by the user in the form of marking orthogonal lines in the image. The main application of this method is for reconstruction of architectural scenes. Beardsley and Murray (1992) also proposed a calibration (intrinsic parameters only) method based on the VPs and corresponding VLs. By enforcing orthogonality constraints on the VPs and VLs of multiple planes, the internal camera parameters can be estimated. Guillemaut et al. (2005) presented a method based on line orientation and the constraint that a VP must lie on the imaged line. This allows the camera intrinsic parameters and orientation to be decoupled from the translation component for a two-step calibration. The main benefit of this approach is that it does not require the explicit calculation of VPs, however, it does require accurate knowledge of 3D control points.

Overall, when considering calibrating a camera using a 3D calibration object, the accuracy will be limited by the resolution with which the control points can be located. Even in the case of using VPs and VLs, orthogonality between planar facets of the calibration object is required, such as an accurately manufactured cube object (Caprile and Torre, 1990). In addition to calibration object issues, there are sensitivity issues in the calculation of VPs and VLs which influence the parameter estimation accuracy.

In order to address the limitations of 3D calibration objects, techniques were developed for calibration objects with reduced dimensionality. Zhang (2004) proposed a camera calibration technique employing a 1D object in the form of a stick with three marker points of known separation distances. By anchoring one of the points, the stick target is rotated to new positions (six independent positions required) to form the input images for calibration. Results indicate comparable performance with standard techniques. However, since one point of the stick is fixed, there is a likelihood of capturing orientations which are dependent in the calibration equations. Hammarstedt et al. (2005) address this issue by detailing the configurations of the 1D object which result in degenerate cases. Recently, Miyagawa et al. (2010) presented a method capable

of calibrating certain parameters (focal length and extrinsic parameters) from a single image of a 1D object. Two sticks containing three collinear points, one of which is shared between both sticks, are placed in an orthogonal fashion. Calibration parameters are calculated based on orthogonality constraints of the calibration object. Generally, calibration using 1D objects is implemented in multi-camera networks where the calibration object is required to be present within the field of view of each camera.

Camera calibration employing 2D targets has arguably received the most attention within the computer vision community. The ease with which 2D targets can be constructed has led to this interest. Standard printers can be used to print calibration patterns for attachment to a planar surface. The majority of planar methods are based on the geometric constraints resulting from the Image of the Absolute Conic (IAC). Triggs (1998) introduced the use of the IAC to the calibration domain for the purpose of auto-calibration. It was shown that the IAC encodes the camera intrinsic parameters. However it was the seminal work of Zhang (1998, 2000) which presents its use for the purpose of camera calibration from planar targets. Typically, the calibration patterns employed are either circular or of a checkerboard nature. Zhang (1998) outlined a calibration procedure which required at least two images of a planar checkerboard target to recover the camera parameters. Since the IAC is independent of the position and orientation of the camera, planar targets are captured with varying pose. Based on the target to image planar homography, two constraints, in the form of the two circular points, are found on the IAC. Thus with two images, assuming a zero skew camera, four constraints are sufficient to estimate the IAC which in turn encodes the intrinsic parameters of the camera. Zhang outlines a closed form solution to estimating the extrinsic parameters once the intrinsic parameters have been recovered. Subsequent non-linear optimisation is performed to include a distortion model and to refine the camera parameter estimates. Although this method is universally accepted as the standard planar calibration approach, with the exception of standard degenerate configurations such as translation and rotation about a single axis, it lacks consideration of the overall effect of planar target orientation on the camera parameter estimates.

In the search for a more meaningful representation of the planar camera calibration setup, Gurdjos and Payrissat (2001) and Gurdjos et al. (2002) propose an intuitive geometric framework to solve the planar calibration problem. Al-

though it is based on the same geometric constraints introduced by Zhang (1998), the method exploits the geometric configuration of the planar target to decouple the calibration of the intrinsic parameters. The Centre Line (CL) is introduced, through perspective correspondence of the planar target and its image, and is shown to include the principal point. Thus with two or more images of a planar target, the principal point can be recovered along with the aspect ratio. This allows the calibration of focal length to be decoupled from the other intrinsic parameters. Calibration of a camera with varying focal length is the primary goal of this work. Guillemaut and Illingworth (2008) developed a similar approach to calibrate a zooming camera with the introduction of the normalised IAC. It is shown that the linear approach is identical to that of Gurdjos and Payrissat (2001), however a calibration result can be obtained with a non-zero skew camera. Both methods can recover the intrinsic parameters with three or more images of a planar pattern.

In contrast to the aforementioned planar calibration methods, which all use chessboard planar targets, a number of approaches have employed planar targets with circular patterns. Yang et al. (2000) describe such a method in which a minimum of three concentric circles are printed on a planar surface. Assuming the centre of the conics is located at the origin of the coordinate system, a relationship between the conic target and its image is developed which allows the planar homography to be estimated. This allows the camera parameters to be recovered by estimating the IAC with two or more images of the planar target in varying orientation using the constraints outlined by Zhang (1998). Kim et al. (2005) describes a planar calibration method also employing concentric circles. However, the number of printed circles is reduced (two) along with a relaxation on the constraint of a known centre. A study by Mallon and Whelan (2007b) compares chessboard and circular patterns in order to determine which pattern yields higher accuracy in the presence of distortion. It is shown that chessboard grid features are invariant to both perspective and distortion bias whereas circular patterns are variant to distortion bias. Thus, chessboard patterns are more suitable for calibration algorithms employing 2D targets.

It is clear that there is an abundance of literature in the area of camera calibration. Generally, the approach to solving the problem depends on the type of calibration target employed. Targets of one/two/three dimensions and their respective methods have been outlined, however it is the planar methods (2D

targets) which are predominantly practised within the computer vision community, particularly when used in conjunction with algorithms based on the IAC. This allows 2D targets to be imaged in varying orientation since the IAC is invariant to camera rotation and translation. Thus the constraint of accurate 3D control point knowledge, which is required using targets of one and three dimensions, is relaxed.

Whilst the primary focus of the computer vision community has been on developing new strategies to calibrate cameras from planar targets, the sensitivity of such targets to camera parameter estimation based on their orientation has received much less attention. Sturm and Maybank (1999) describe image configurations which lead to singularities in the planar calibration equations, however, orientations which are near degenerate are not examined. Additionally, there are no reported guidelines on enforcing ideal geometry on the planar target orientations. Wang and Liu (2006) suggest a subset of planar target orientations, which according to them yields good results, but no justification is given. Rupp and Elter (2007) describe a heuristics approach which uses a genetic algorithm to optimise the selection of images from a dataset to yield accurate camera parameter estimates. However, this approach requires the calibration of each possible subset to minimise its selection criteria which is a combination of the re-projection error and a plane fit error. There are two problems with this approach. Firstly, it requires calibration of all possible combinations of images in the data set. This is a slow and rather mechanical process to determine accurate parameter estimates. The author reports the process as taking several hours depending on the data set size. Secondly, the proposed approach does not enlighten practitioners on the significance of the planar target orientation in the overall scheme of planar camera calibration.

Specifying ideal geometry for input planar calibration target images is the main problem that is addressed in Chapter 3. Initially, a more detailed examination of planar calibration (Zhang, 1998, 2000) is given in Chapter 2, along with an insight into the geometric relationship of the planar target orientation (Gurdjos and Payrissat, 2001). Chapter 3 presents a new approach to forming input image data sets, without the need for a pre-calibration stage, for planar camera calibration. Optimal image geometry is imposed, consequently avoiding degenerate and near-degenerate configurations, yielding more accurate calibration results with less input images.

## 1.2.2 Modified Optics in Computer Vision

A primary goal of computer vision is to accurately estimate object depth from images. The pinhole camera model and its calibration has been discussed in detail, however, in the context of depth estimation, object depth information is lost in the central projection of world points to image points. Therefore, depth estimation is not possible from a single image using the pinhole camera model. Traditionally this task has been achieved passively by employing two pinhole cameras in a stereo configuration. Many robotics applications implement stereo vision systems, and increased research in the area has been observed since the 1970s. The geometrical background and many of the early techniques are outlined in Faugeras (1993), while Brown et al. (2003) reviews the more modern advances in terms of point correspondence, handling occlusions and real-time implementation. One of the main drawbacks of stereo configurations is the need for an additional sensor and accurate knowledge of the camera configurations with respect to each other. In terms of generating depth, an additional drawback is that of occlusions i.e. when a feature is detected in one image but is not visible in the second image. A new approach to estimating depth using a single sensor was proposed by Pentland (1987). Pentland moved from the traditional pinhole model to a thin lens model in order to take advantage of the focus/defocus artifacts within an imaging system. It was shown that the blur induced at out of focus step edges could be related to object depth. This process was termed Depth from Defocus (DfD) and stimulated intense research into alternative approaches to Pentland (Chaudhuri and Rajagopalan, 1999, Subbarao and Wei, 1992, Subbarao and Surya, 1994). The main disadvantage of these methods was the requirement of accurate estimation of the blur within the images. Additionally, many of these techniques required well defined edge features for defocus estimation. Lai et al. (1992) proposed an extension to Pentland’s method to capture depth information from a single image. However, the dependency on strong step-like edges remained. Within the scope of this thesis, the primary interest is in retrieving depth information using a single image from a single sensor. This is a task that can not be accurately fulfilled by employing traditional DfD techniques.

Addressing the limitations of DfD led to the emergence of a new area in the form of modifying the optics of a traditional camera for accurate range recovery. Adelson and Wang (1992) introduced the “Plenoptic Camera” which

captures the 4D light-field. In other words, directional information of the light rays striking the sensor is also recorded. This is equivalent to capturing structural information of the world from all possible viewpoints within the camera aperture. Directional information was acquired by modifying the traditional optics of the camera. A lenticular array was placed on front of the image sensor while a field lens was also inserted to ensure that the main lens aperture was placed at optical infinity from the microlenses. Furthermore, a relay lens was introduced to the system to enable a replicated view, using an additional camera, for calibration and alignment purposes. Figure 1.1(a) presents the plenoptic configuration. This optical arrangement provided multiple views from different locations within the system aperture and recorded these onto separate pixels. Thus, correspondence problems were reduced to selecting the correct pixels using a weighted mask. Subsequent depth estimation was performed based on the displacement of image features caused by the lenticular array. An additional application of the acquired light-field was the ability to generate synthetic images from alternative viewpoints. However, the spatial resolution is reduced due to the sampling process. It was also noted that this method was unsuitable for large range depth estimation due to the fixed optical configuration and limitations on the lens diameter.

Ng et al. (2005) proposed a hand-held solution for a plenoptic camera which has been recently<sup>4</sup> made available in the market. Improvements were made to the design of the camera with portability being a key driver. The field and relay lenses were eliminated compared to the original design by Adelson and Wang (1992). The image sensor was located at the focal plane of the lenticular array. This lenticular array was focused at infinity, which due to the diameter of the microlenses, was in fact focused on the aperture or traditional lens of the camera. The optical arrangement for this type of plenoptic camera is presented in Figure 1.1(b). The primary focus of this work was not range estimation but rather post processing captured images for refocusing. Spatial resolution of captured images was reduced due to the sampling of multiple pixels per microlens within the lenticular array. This is the same limitation that applied to the configuration of Adelson and Wang (1992). Additionally, the required placement accuracy of the lenticular array was approximately  $36\mu m$ . This is a consequence of modifying the optical path within the image space of the imaging system, which is highly sensitivity. Thus the physical

---

<sup>4</sup>[www.lytro.com](http://www.lytro.com)

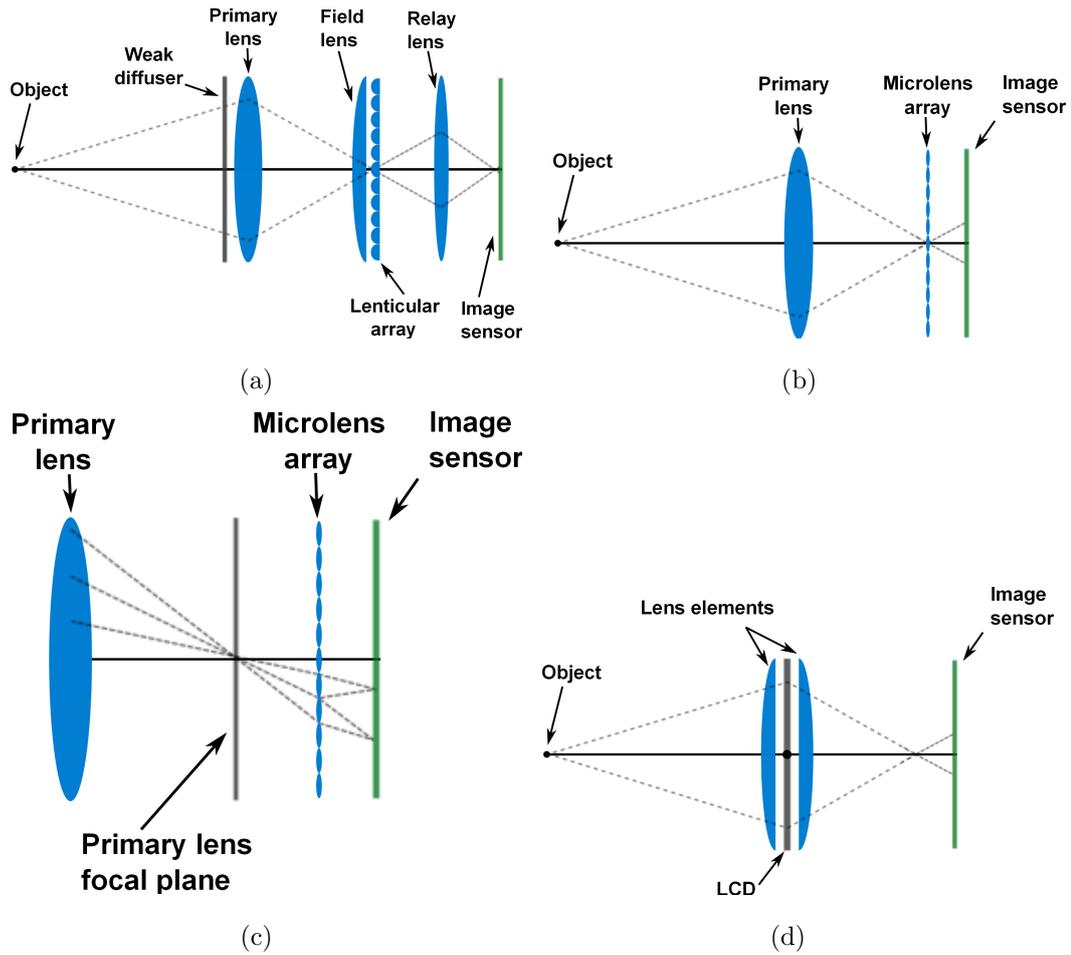


Figure 1.1: Non-conventional imaging approaches (a) Adelson and Wang Plenoptic camera; (b) Ng's plenoptic camera; (c) Georgiev's plenoptic configuration; (d) General modified aperture plane imaging

construction of a plenoptic camera requires precise manufacturing.

Improvement in spatial resolution was the key contribution in the recent plenoptic system proposed by Lumsdaine and Georgiev (2009). The optical configuration (in Figure 1.1(c)) was modified to focus the microlens array at the focal plane of the main lens rather than placing the microlens array at this plane. This increased the spatial resolution of the final images but reduced the angular resolution in sub-sampling the system aperture. Fife (2009) describe a sensor level architecture which is customised to capture the light-field in a similar optical configuration to Lumsdaine and Georgiev (2009). However, the primary focus of that work is in the custom manufacturing of the image sensor (CMOS) with the integrated microlens array, thus applications in depth estimation and image refocusing are not explored. Besides image resolution, the main issue with plenoptic cameras is practicality in their optical config-

uration. Since the optical modifications are made in the image space of the system, sensitivity dictates an extremely fine resolution in the manufacturing process (minimum  $36\mu m$ ). Additionally, metric depth estimation is not a high priority with plenoptic approaches since the main goal is to digitally refocus images or regenerate alternative perspectives from a single image capture.

Alternative methods have been developed by modifying the optical properties either within the system of lens elements or in object space. General modification within this domain are presented in Figure 1.1(d). Dowski and Cathey (1994) describe a method to obtain depth information from a single image and sensor using a “phase plate” located at the lens. This plate is effectively an optical mask which has been designed to be highly sensitive to depth information. Thus, range information can be extracted from images by examining the frequency response of the imaging system. Metric depth estimation is reported for a confined range (approximately 1.3 to 1.5m) with 2% error. However, spatial information is degraded which inhibits image reconstruction. In order to recover image data a new approach was taken which required the construction of an alternative optical mask. Dowski and Cathey (1995) proposed a method to extend the depth of field of standard imaging systems by employing a cubic-phase modulated mask within the lens. The mask was designed in such a way that the system response was to defocus all rays independently of depth by the same magnitude. Therefore a sharp, all in focus image, could be retrieved by a single deconvolution to the image. The main drawback of this system was that metric depth information could not be retrieved.

Further work involving the deployment of optical masks within the imaging system was carried out by Farid and Simoncelli (1996, 1998). An optical system was constructed with a Liquid Crystal Display (LCD) located at its optical centre (see Figure 1.1(d)). The LCD is used to generate two separate optical masks, one of which is the derivative of the other. Therefore, the differential variation in intensities recorded by both masks are related by a scale factor. Thus recovery of the scale factor allows depth estimation with calibration knowledge of the camera (sensor depth and focal length). Depth estimation results are demonstrated for near range planar objects ( $< 170mm$ ). The experiments performed were quite limiting with error in depth estimation reported at 9%. This approach does however allow more flexibility than that of Dowski and Cathey (1994) which requires a complex optical mask. Another limitation of this approach is that two images are required, one for each mask,

therefore depth estimation is not acquired in a single image.

Recently, Zhou et al. (2010) presented an approach to depth estimation using an optical diffuser located in object space. “Depth from diffusion” operates on the same principal as depth from defocus. However, in order to increase the blur baseline in the image, this method is not limited by the physical diameter of the aperture. Increasing the diffusion angle of the optical diffuser is sufficient to increase the sensitivity of the depth estimation at large depth ranges. Metric depth estimation up to  $2m$  is reported with errors less than  $1mm$ . Macro depth estimation is also presented with depth resolution of less than  $1mm$ . The sensitivity of this depth estimation is certainly superior to traditional techniques, however there are limitations with this configuration. Besides requiring two images in configurations with and without the diffuser, a major limitation of this approach is the placement of the optical diffuser. Although it is placed in object space, which is less sensitive to error than image space, it is required to be located near the object being imaged. Therefore, the implementation of a depth from diffusion system using this technique is impractical.

There has been a large shift in the research area of modified optics in recent years, particularly with regard to modifying the aperture of conventional camera lenses. This area of “coded apertures” has been an active area of research in the fields of astronomy and medical imaging for X rays and Gamma rays, but it was the work of Levin et al. (2007) which highlighted the benefits of coded apertures for depth recovery to the computer vision community. The key idea was to construct an aperture which encoded the system response (PSF) to be more distinct in the frequency domain. This enabled accurate recovery of blur scale related to depth. Thus, Levin was able to recover depth information from a single image. Depth estimation results are presented for a scene with objects placed between  $2$  and  $3m$  and exhibit a high degree of accuracy (less than  $2mm$ ). However, accurate depth recovery required an extensive pre-calibration procedure. The camera was set to focus at every depth within the working range at a resolution of  $10cm$ . In order to address the diffractive and distortion elements of the imaging system, calibration images were captured with the target varying across the horizontal plane of the image. Additionally, the developed method requires a deconvolution operation which is known to result in image artifacts. Although this configuration required cumbersome calibration, it reignited interest in implementing modified optical systems for

depth recovery.

Zhou and Nayar (2009) propose a method for designing and evaluating aperture patterns which are subsequently employed in a similar manner to Levin et al. (2007). A criterion based on defocus deblurring is developed, therefore the main application is in the area of post-processing captured images for the creation of all-in-focus images. This approach is the basis for further work (Zhou et al., 2009) in which coded aperture pairs are optimised for DfD. The apertures are designed to complement each other when analysing the frequency response of the imaging system. Thus, a deconvolution is also required to recover depth information. The method also requires two images, one from each coded aperture, thus the capturing process is complicated due to the necessary modification to the internal components of the imaging lens.

To this end, coded aperture techniques have been implemented on LCD displays which are embedded in the imaging lens, such as presented in Figure 1.1(d), (Liang et al., 2008, Gao et al., 2007, Dou and Favaro, 2008) in order to avoid manually changing the coded aperture inside the lens. However, these methods tend to require precise optical arrangement, generally in a laboratory environment, and lack practicality in implementation. A further problem with LCD apertures is the transmission of light through the screen in its “on” state. Nagahara et al. (2010) describe a coded aperture system using a Liquid Crystal on Silicon (LCoS) which addresses some of the issues observed with LCD systems such as poor transmission and large diffraction. The primary issue with coded aperture techniques employing controllable displays at the aperture plane is the constraints imposed on the optical arrangement. Typically, the display is either mounted inside the imaging lens, thus modifying conventional lenses, or optical relay components are used to shift the optical centre to a suitable location for placement of the LCD/LCoS system.

In order to address the complications of aperture plane and code definition, recent work has reverted to more basic codes and alternative methods to extract image information. In an original work, Jones and Lamb (1993) describe a technique using a customised spherical lens which is separated in two halves. An aperture with two slits is inserted at the aperture plane and a composite image is formed. The main contribution of this work was in the area of feature matching across the composite image of both slits. Thus depth estimation results are not outlined, but verification of composite feature match disparity

related to a single depth is reported. Hiura and Matsuyama (1998) proposed a multi pinhole aperture for depth recovery employing a multi-focus camera. The multi-focus camera consisted of three Charge Coupled Device (CCD) sensors which were equipped with custom optics ensuring that each image captured was focused at different scene depths. The aperture was located at the focal plane, thus the optical system was telecentric. Therefore, with unit magnification, the scale of the blur could be obtained from the three images in a DfD style approach. In a more recent approach, Kim and Kanade (2011) formed a telecentric system incorporating a Fresnel lens, however, similarly to Hiura and Matsuyama (1998) the implementation of such systems require specialised equipment and are limited due to physical size and the number of sensors being used.

Bando et al. (2008) describe a system consisting of a conventional lens which has been modified in a similar manner to Levin et al. (2007). The coded aperture contains three separate openings covered with a red, green, and blue filter. By separating the captured red, green and blue image planes, induced colour mis-alignments are used to discriminate between foreground and background information. The main focus of this work is to extract foreground information for matting applications. An indicative experiment was performed to yield a quality metric on foreground segmentation. Therefore no effort was made to retrieve metric depth information. Lee et al. (2010), Sangjin Kim and Paik (2010) propose a similar approach to Bando et al. (2008) employing an “RGB” aperture within the imaging lens. The main goal of these approaches is to recover an all-focused image by registering the three colour plane images. There are two main limitations, other than not estimating metric depth, with the “RGB” aperture approaches outlined. Firstly, the aperture position is limited to the conventional aperture plane of the imaging lens and secondly, the structure of the “RGB” aperture plane may not be modified. Although the approaches of Bando et al. (2008) and Lee et al. (2010), Sangjin Kim and Paik (2010) are optimised for matting and generating all focused images, the “RGB” concept is certainly applicable to metric depth estimation.

Recently Koh et al. (2011) developed a dual aperture with colour filters for the purpose of autofocusing a camera. The aperture is placed at the conventional lens iris diaphragm location. The speed at which the autofocus procedure is completed is of paramount importance, thus the primary objective of this work is to reliably measure the colour dependent disparity generated by each colour

filter using simple techniques. System implementation is completed with two apertures, one for capturing images and the other (coded) for autofocusing. Camera calibration is not completed, rather the imaging configuration data is taken from the image header information which is known to be unreliable. Experiments are presented for auto-focusing on objects at fixed distances between 2 and 11m using the colour disparity properties of the imaging configuration. Metric depth information is not calculated. While this approach certainly merits further work on the possibilities of depth estimation, it is expected that the colour dependent disparity at large object depths (coupled with the lack of accurate calibration information) would not be sufficiently sensitive to accurately estimate multiple objects at these ranges.

Of the reviewed methods based on modified optical configurations, the general applications are to either post-process images to create all-focused images, or to extract foreground information for matting applications. Typically, these methods require the recovery of the modified system Point Spread Function (PSF) for depth map generation, and involve a deconvolution operation to recover the focused images. With the exception of Levin et al. (2007), the calibration of such systems is not addressed which results in a lack of metric depth information employing these systems. Additionally, the location of the modified aperture is generally fixed at the location of the iris diaphragm. Thus the focus of chapters four and five in this thesis is to redress these issues with the proposition of a new modified optical imaging model. The new multi-pupil imaging model defines the location of pupils within the pupil plane and investigates their sensitivities as well as the location of the pupil plane within the imaging system. A practical calibration procedure is developed for a general multi-pupil system and verification of the model is performed via metric depth estimation.

### 1.3 Mathematical Notation

The projective space is represented by  $P^n$ , respectively, where  $n$  is the dimension of the space. Matrices are denoted by upper case letters as  $A$ . The element at row  $i$  and column  $j$  of matrix  $A$  is denoted as  $a_{ij}$ . Matrix  $A_{[i \ j \ \dots \ k]}$  is a new matrix formed from columns  $i, j, \dots, k$  of matrix  $A$ .  $I_{n \times n}$  is the identity matrix of size  $n$ . Transformations are denoted  $T = [\mathbf{t}_1 \ \mathbf{t}_2 \ \dots]$  where  $\mathbf{t}_i$  are the

columns of the transformation with entries  $t_{11}, t_{12}, \dots, t_{33}$ . Rotations,  $R$ , and homographies,  $H$ , are represented similarly.

Vectors and points in  $P^2$  are denoted in bold lower case as  $\mathbf{a} = [a_1 \ a_2 \ a_3 \ \dots \ a_n]^T$  with elements  $a_1, a_2, a_3, \dots, a_n$ . Points in  $P^3$  are represented in bold upper case as  $\mathbf{A} = [A_1 \ A_2 \ A_3 \ \dots \ A_n]^T$  with elements  $A_1, A_2, A_3, \dots, A_n$ . Point  $\mathbf{A}_{[i \ j \ \dots \ k]}$  is a new point formed by selecting elements  $i, j, \dots, k$  from point  $\mathbf{A}$ . Equality up to a non-zero multiple is denoted with the  $\simeq$  symbol. Finally, an image point  $\mathbf{p}$  after aberrations is represented by  $\tilde{\mathbf{p}}$ . An image point which has been corrected for aberration is represented as  $\bar{\mathbf{p}}$ .

The effective focal length of a lens is represented as  $F$ , while the image plane to lens centre distance, the sensor depth, is represented as  $v$ .

## 1.4 Contributions

The major contributions of this thesis correspond to the main themes within this thesis. Contributions of lesser weight, which contribute to and enable the major contributions, are described in the minor contributions section. Together, these contributions encapsulate the most important work of the thesis.

### Major Contributions

- Sensitivity of the pinhole model within the context of planar camera calibration is investigated and outlined. Planar camera calibration target orientations play a significant role in parameter estimation accuracy. Certain orientations cause dependency in the planar calibration equations and result in poor parameter estimates. These planar target configurations are known as degenerate configurations. The manifold of all possible degenerate configurations for a two planar target setup is identified. Additionally, configurations which are near-degenerate, which also contribute to poor parameter estimates are highlighted. A method for recognising these configurations is presented using a real geometrical entity on the image plane in the form of the Centre Line (CL).
- A novel framework is developed for the formation and capturing of input images for planar camera calibration. The input images are called

the Image Network (IN). Characteristics of the image CL are used to specify optimal planar target pose. INs which contain optimal geometry are termed Generated Image Networks (GIN). Planar camera calibration performed using the proposed approach results in more accurate camera parameter estimates as well as higher efficiency due to the avoidance of degenerate and near-degenerate configurations.

- A new imaging model is presented in the form of the multi-pupil imaging model which enables the recovery of metric depth information from a single image. The Double Pupil Model (DPM) is outlined which contains a modified pupil plane located at the conventional aperture location. Extensive model characterisation is completed along with depth estimation experimentation. By relaxing the constraints of pupil plane placement, further enhancement is achieved with the proposed Double Pupil Shifted Model (DPSM). Similar characterisation is performed along with extensive depth estimation experimentation in which it is shown that the multi-pupil imaging models achieve a high degree of accuracy across a large depth range.
- A practical and accurate calibration framework for multi-pupil imaging is presented. The calibration can be performed with a single image given a minimum of two objects for DPM calibration and four objects for DPSM calibration. Only the offset between object depths is required. The multi-pupil calibration approach is benchmarked using standard calibration techniques. Superior accuracy is obtained using the proposed calibration methods as certain lens aberrations are accounted for which are not addressed with the standard techniques.
- A novel approach to modelling, calibrating and removing monochromatic aberrations from imaging systems is developed. This is achieved within a multi-pupil imaging framework. A pupil plane containing three pupils is presented which provides a new cue to estimate aberrations with pupil plane dependencies. This allows the calibration of all monochromatic aberrations within the imaging system using point feature correspondences. The accuracy of the aberration corrected images are validated with depth estimation and object reconstruction experiments.

## Minor Contributions

- An automatic image selection strategy for planar camera calibration which complements the GIN approach is developed. The input requirements are an image data set in which the algorithm selects the most geometrically independent image set. The resulting image network is called Selected Image Network (SIN). Additionally, an augmented application is developed for the realisation of the GIN approach for the replication stage.
- An examination of the sensitivity of placement of the pupil plane within an imaging system is outlined. It is shown that image space placement of the pupil plane is highly sensitive in comparison to object space placement. The relationship between the system PSF and the pupil plane placement is shown to be non-linear when the pupil plane is placed in image space.
- The impact of sub-sampling a conventional pupil plane with an offset pupil is examined in detail. The system response, in the form of the PSF, is shown to sub-sample a single centred ray which passes through the modified pupil plane. This property forms the basis for multi-pupil imaging.

## 1.5 Thesis Organisation

Chapter 2 examines existing camera models used in the field of computer vision. Initially, the pinhole model is introduced which is augmented into the lens models using Gaussian lens laws. The calibration of camera models is discussed, in particular two planar calibration approaches which are used in this thesis are outlined. Finally, an examination of the significance of the pupil plane within an imaging system is presented. Experiments are conducted to analyse the sensitivity of shifting the location of the pupil plane as well as moving the pupil within the pupil plane.

Chapter 3 firstly addresses the issue of degenerate configurations in planar camera calibration. A key link is made between these configurations and the orientation of the image CL. It is shown that the manifold of critical configurations can be avoided with knowledge of the image CL. Thus, the second part

of this chapter develops a new approach to forming an input image network for accurate and degenerate-free planar camera calibration. Experiments are conducted to verify the increased accuracy and efficiency in camera parameter estimates employing the new image networks.

Chapter 4 presents a new imaging model for the recovery of metric depth information from a single image. Initially, the double pupil model is introduced. The sensitivity and accuracy of the model is examined and benchmarked with industry standard optical simulation software. Subsequently, the double pupil shifted model is presented, which relaxes pupil plane constraints of the DPM, along with detailed analysis of its sensitivity and accuracy. A novel calibration procedure is described for both multi-pupil imaging models. Finally, extensive experiments are carried out to verify the accuracy of the calibration procedure as well as the viability of the models for accurate depth estimation from a single image.

Chapter 5 investigates the applicability of multi-pupil imaging for the modelling, calibration and removal of aberrations in images. Initially, the formation and effects of monochromatic aberrations in imaging systems is presented. Subsequently, a pupil plane is introduced containing multiple pupils which are shown to encode additional information on aberrations that have a pupil plane dependence. This contributes to the formation of a calibration procedure which allows all monochromatic aberrations to be estimated from a single image of a control point grid. Evaluation of the accuracy of the multi-pupil calibration and aberration removal approach is demonstrated through object depth and structure estimation experiments.

Chapter 6 summarises the principal contributions of the thesis. Directions for further research are outlined, and a list of publications arising from the work in this thesis is provided.

## Chapter 2

# Camera Models and Calibration

Researchers in the field of computer vision strive towards accurately modelling the imaging process. This consists of capturing the 3D projection of world points onto an imaging sensor. An accurately modelled camera allows metric information about the 3D world to be drawn from the captured 2D image. The extent to which this process can be precisely modelled depends on the camera model imposed, and secondly, on how well this model can be calibrated. Calibration of a camera model is the process of recovering the model parameters which describe the projection from world to image. Typically, a calibration process is undertaken to recover these parameters. Therefore, this chapter is primarily concerned with the camera models used in the area of computer vision, and the calibration of these camera models.

Section 2.1 introduces the pinhole camera model, which is the most basic camera model used in the field. It is based on the central projection of 3D points to a 2D image plane. The camera calibration matrix is defined along with the parameters which describe the location and orientation of the camera with respect to the 3D world coordinate system. The pinhole model gives a precise geometric mapping from 3 space to the image plane, however, it does not capture irradiance or imaging lens phenomena. Consequently, lens models are introduced in the form of the thin and thick lens models. Gaussian optics is outlined and the thin lens law is presented. The introduction of an imaging lens allows optical phenomena such as focusing and blurring artifacts to be captured by the camera model. Since most conventional imaging lenses comprise of multiple lens elements, the thick lens model is presented, which

allows these lenses to be modelled by the cardinal points of the optical system. One of the key aspects of this thesis is the recovery of metric depth information from a single image. Using the pinhole model, this depth information is not captured within the central projection equations. Thus, the introduction of lens models, which capture image focusing properties of an imaging system, provides an additional cue which can be exploited for the purpose of retrieving depth information.

Recovery of metric information from images requires knowledge of the numerical quantities of the camera model parameters. To this end, the calibration of a camera model is presented in Section 2.2. Two calibration methods based on an imaginary projective entity, the IAC, are described. Planar camera calibration is arguably the predominant method practised in the computer vision community when calibrating conventional cameras. The original method of Zhang (1998) is presented. Subsequently, an alternative approach, also based on the IAC is examined, in the form of Gurdjos and Payrissat (2001). This transforms the planar calibration problem into a more intuitive geometric domain and provides a more meaningful cost function for the estimation of the camera parameters. Both planar methods assume a pinhole camera model. The geometric equivalence of the pinhole and lens models is outlined. Certain camera parameters, in particular the sensor depth, are required for the purposes of depth estimation. The focal length parameter retrieved via planar camera calibration is thus applicable to the lens models and laws in the form of the sensor depth.

Section 2.3 examines, in more detail, the significance of the pupil plane within the lens models. In particular, the placement of this plane within a system of lens elements, such as those which form conventional imaging lenses, can have a large effect on the resulting image blur recorded. The proportionality between pupil (aperture) diameter and image blur for object distances which lie outside of the camera focus plane is shown. However, if the pupil plane is shifted outside of the lens, this relationship does not remain constant. Experiments are performed to highlight how this relationship varies for a pupil plane placed in object space as opposed to image space. Object space is defined as the area in front of the apex of the first lens surface (extends to infinity), while image space is the area between the apex of the last lens surface and the image plane. Results indicate that pupil placement in object space leads to a linear relationship between the pupil diameter and the magnitude of image

blur. Conversely, image space pupil plane placement leads to a non-linear relationship. This is shown to be as a result of optical irises in an imaging system. The entrance and exit pupils, defined as irises throughout this thesis, of an optical system are defined and shown to be the limiting factor for light rays entering and exiting the imaging system respectively. Due to the linear relationship of object space pupil planes, an experiment is performed to confirm that a pupil plane located at the lens with a geometrically equivalent pupil diameter can be defined using a pupil plane which is located in object space. Experiments are conducted to examine the properties of the pupil on the pupil plane. Results show that a conventional pupil can be sub-sampled by a pupil of much smaller diameter. Thus, a ray passing through a conventional pupil can be approximated by a sub-sampled pupil centred at that ray's location in the conventional pupil. The intersection of this ray with the image plane is calculated as the centroid of the resulting sub-pupil blur spot on the image plane. Further experiments investigate the impact of the sub-sampling pupil diameter on the accuracy of the ray data recovery. It is shown that smaller pupils lead to better accuracy in centroid estimation. Additionally, diffractive effects of sub-sampling pupils, due to the decrease in pupil diameter, are examined. Results indicate that the diffractive component is increased with a decrease in pupil diameter, however, the centroid estimation accuracy is unaffected.

### 2.1 Camera Models

One of the fundamental tasks in computer vision is examining how the 3D world relates to a captured 2D image. Typically, the process which describes this relationship is governed by the choice of imaging model used. In general, a camera consists of two primary components in the form of an imaging sensor, and an optical system. The optical system, or camera lens, is formed by a variable number of glass surfaces (possibly with different refractive indices) in conjunction with a number of stop planes. The function of the camera lens is to capture multiple rays emanating from object points in the 3D world, and suitably adjust this information for the image sensor. Subsequently, an image sensor captures the object ray information provided by the lens and forms an image. There are a multitude of possible combinations of camera lenses and image sensors. Lenses which increase or decrease the field of view or magnification of the scene being observed all modify the optical properties of

the imaging lens. Image sensors vary in physical size, however optical systems are generally tailored to match the image sensor size.

In spite of the number of variables within the constraints of an imaging system, it is the task of the computer vision practitioner to choose the most appropriate model to describe the imaging process for the required task. This section examines three such models. The initial model examined is the simple pinhole camera model. This model does not account for optical phenomena of the lens. Consequently, the second model under examination considers the optical effects of the imaging lens in the form of the ideal thin lens model. The final model considers additional optical properties of the imaging system in the form of the Gaussian thick lens model.

### 2.1.1 Pinhole Camera

The pinhole model is the most basic and commonly used camera model in computer vision. Historically, the first recording of a pinhole camera was the camera obscura which dates back as far as 470BC. The first images containing perspective effects were recorded during the Renaissance by painters who made use of the pinhole camera. Such pinhole cameras were constructed by inserting a small circular hole in an opaque sheet of material. Naturally, the object scene rays passing through the pinhole were projected onto a canvas sheet, which was parallel to the pinhole plane. This allowed the painters to capture the perspective effects of the scene observed on the canvas sheet.

A general pinhole model is presented in Figure 2.1. A single Euclidean coordinate frame is assumed in which the origin is set as the camera centre,  $\mathbf{C}$ . If the  $Z$ -axis of the Euclidean coordinate frame is chosen as the principal axis of the system, then the image plane can be defined by any plane orthogonal to this axis. Consequently, the image plane is defined as the orthogonal plane at a distance  $f$  from the camera centre. The distance  $f$  is commonly referred to as the “focal length” within a pinhole configuration<sup>1</sup>. The principal point,  $\mathbf{p}$ , is defined as the intersection of the principal axis with the image plane.

An object point in Euclidean space is represented by the vector  $\mathbf{X}$ . In a pinhole

---

<sup>1</sup>This is a common abuse of terms in the field of computer vision and will be clarified in Section 2.1.2

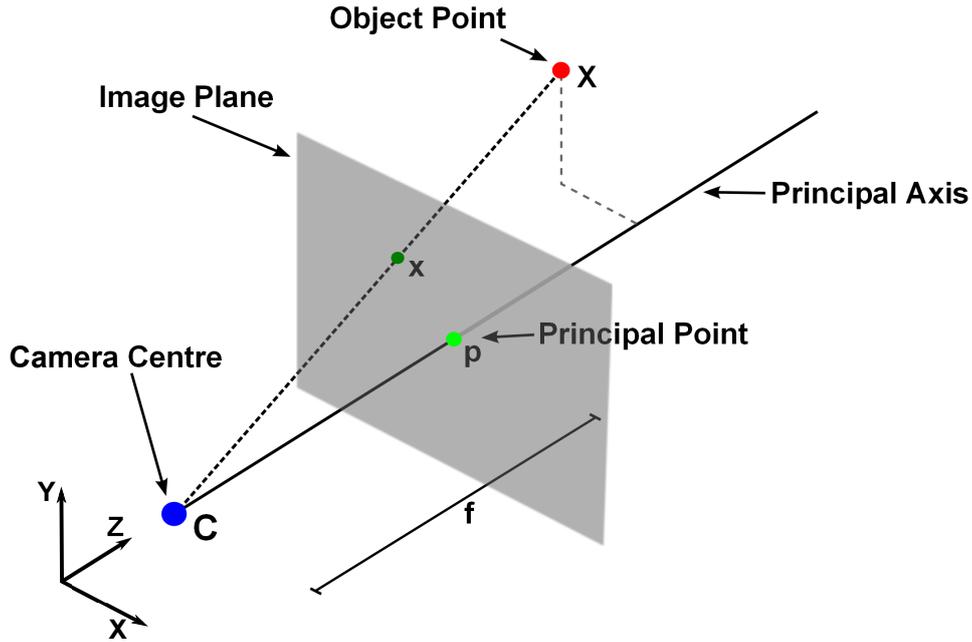


Figure 2.1: Central projection of the pinhole model.

model framework, only a single ray emanating from the object point may pass through the camera centre. This ray intersects the image plane at the point  $\mathbf{x}$ . Assuming the origin of the image coordinate system is the principal point, the projection of a 3D object point to a 2D image point can be described linearly using homogeneous coordinates as,

$$\mathbf{x} \simeq \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \mathbf{X} \quad (2.1)$$

Figure 2.2 presents a generalised pinhole camera model in which all coordinate frames are outlined: the camera coordinate frame (origin at  $\mathbf{C}$ ), the world coordinate frame (origin at  $\mathbf{O}$ ), and the image coordinate frame  $(x_i, y_i)$ . Generally, the origin of a camera coordinate frame is not located at the principal point. Therefore a distinction is made between the camera coordinate system and the image coordinate system. This relationship is described by a principal point offset  $(p_x, p_y)$  and  $f$ . Furthermore, to fully describe the conversion of camera coordinates to image coordinates, non-Euclidean properties of the camera frame must be accounted for. If image pixels are not exactly square, a scale factor is introduced which will modify the camera focal length in each direction. This parameter is realised in the form of the aspect ratio,  $\alpha$ , which is the ratio of pixel height to pixel width. Additionally, a camera skew parameter,  $s$ , is included to account for skewness in the camera sensor

elements. These parameters are independent of the camera orientation and position. They define the camera calibration matrix,  $K$ , which describes the intrinsic parameters of the pinhole camera.

$$K = \begin{bmatrix} \alpha f & s & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.2)$$

The projections described up to this point have assumed that the camera coordinate frame is coincident with the world coordinate frame. However, as depicted in Figure 2.2, this is generally not the case. The world coordinate system is a different Euclidean frame with its origin at  $\mathbf{O}$ . Both coordinate systems are related by a rotation,  $R_c$ , and translation,  $t_c$ , where  $(R_c, t_c)$  represents the camera's orientation and location in the world coordinate frame. The relationship between the object point  $\mathbf{X}$  in camera and world coordinate systems is,

$$\begin{aligned} \mathbf{X}_{cam} &= \begin{bmatrix} R_c & t_c \\ 0 & 1 \end{bmatrix} \mathbf{X}, \\ \text{where } t_c &= -R_c \mathbf{C} \end{aligned} \quad (2.3)$$

These six parameters (3 angles for  $R_c$  and 3 for  $\mathbf{C}$ ) are known as the extrinsic parameters of the camera. Knowledge of the intrinsic and extrinsic camera parameters yields a full description of the 3D to 2D mapping of a pinhole camera model. This defines the pinhole camera projection matrix  $P$ , which consists of Eqns. 2.2 and 2.3,

$$P = K [R_c | t_c] \quad (2.4)$$

Now the projection of an object point in the world coordinate system,  $\mathbf{X}$ , onto the image plane in the image coordinate system can be described as

$$\mathbf{x} \simeq P\mathbf{X} \quad (2.5)$$

### 2.1.2 Thin Lens Camera

In order to model the imaging process of a conventional camera more accurately, an imaging lens is introduced to the camera model. With the pinhole

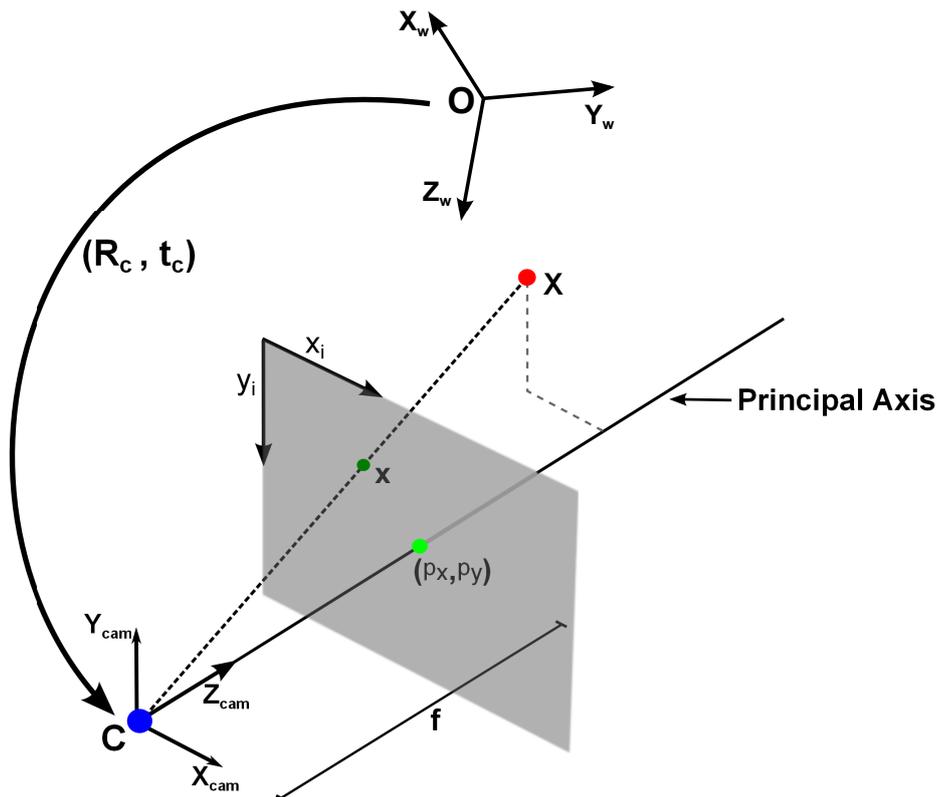


Figure 2.2: Pinhole camera model.

camera model, an assumption was made that only a single ray emanates from each object point. Consequently, image information relating to the camera focus is neglected, thus each imaged object point is always ideally focused. This is not the case with a conventional camera.

The purpose of deploying a lens within an imaging model is to overcome the impracticalities of the pinhole model. Since a pinhole model requires an aperture of infinitesimal size, the amount of light irradiated onto the image sensor is minimal. Thus, image detail is lost - the image is under exposed. If the aperture size was increased in the pinhole model, additional rays would be captured by the imaging sensor. However, these scene rays would not converge to a single image point, and thus there would be a combination of over-exposure within the image coupled with a lack of object detail. The solution to this problem was the introduction of an imaging lens.

Interest in using optical lenses for the formation of images accelerated with the birth of the photographic process in the early part of the 19<sup>th</sup> century (Kingslake, 1989). The primary function of the imaging lens was to capture additional scene radiance and simultaneously focus scene rays at a single im-

age point. This led to increased research in the area of optical lens design and resulted in the proposition of Gaussian or paraxial optics. Gaussian optics assumes that the lens under examination is rotationally symmetric and centred on its optical axis. Scene rays emanating from the object point of interest are assumed to be in a region close to the optical axis. This allows the paraxial approximation to be made, in other words, the slope angles and the incident and refracted angles are assumed equal to their sines and tangents. The introduction of paraxial optics resulted in the proposition of the thin lens model for an imaging system, as presented in Figure 2.3.

The thin lens consists of a single lens element and two refractive surfaces. The thickness of the lens is assumed negligible. In Figure 2.3, the lens is rotationally symmetric about the optical axis. The lens diameter  $D$  is equal to the aperture diameter which, since there is no physical stop, is the limiting factor for scene rays passing through the lens. The optical power of the lens, known as focal length, is defined by the radius of curvature of the lens surfaces and the corresponding refractive index of the lens material. Since there are two surfaces within the thin lens system, which are symmetric, a focal point is defined on both the object and image side of the lens. These points correspond to  $F'$  and  $F$  respectively. Object points are defined at distances  $u_0$  and  $u_1$  on the object side of the lens. The significance of the optical power of the lens and its focal points is highlighted in the formation of the image points. For the case of the object at depth  $u_1$ , three rays are traced from the object point to its formed image point on the sensor. Rays which enter the lens parallel to the optical axis are refracted to pass through the focal point on the image side of the system. Conversely, rays which pass through the focal point on the object side, entering the lens, are refracted and travel parallel to the optical axis on the image side. The object ray which passes through the centre of the lens does not deviate. In this case, the incident angle is equal to the refracted angle. All rays emanating from the object point passing through the lens intersect in a single point and thus form an image at this location.

The image sensor is located at a distance,  $v$ , from the lens. Throughout this thesis, this property is referred to as the “sensor depth”. A common misconception in the field of computer vision is to assume that this property is equal to the focal length of the imaging system. This arises from the use of the term “focal length” within the pinhole configuration to describe the image plane to pinhole distance. The correct use of the term “focal length”

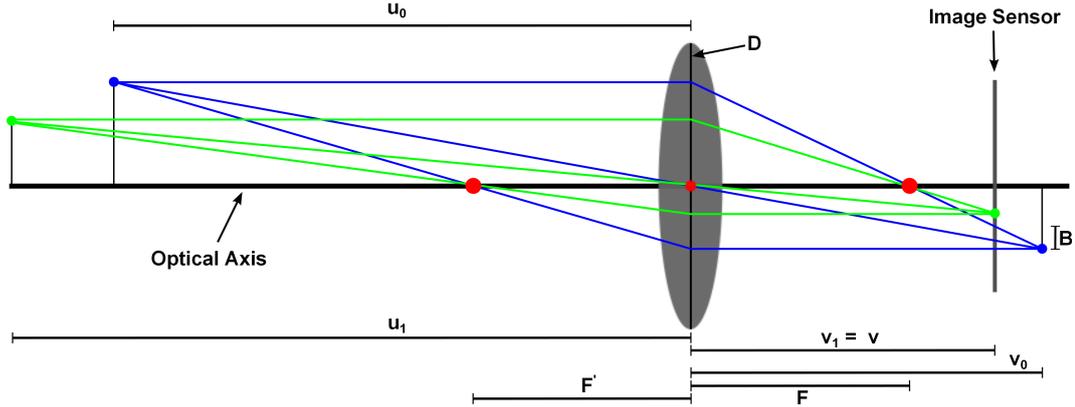


Figure 2.3: Thin lens camera model.

is to describe the optical power of an imaging lens. Therefore, if a pinhole camera model is being used to calibrate a conventional camera, the  $f$  parameter recovered (intrinsic camera parameter from Eqn. 2.2) is in fact the sensor depth parameter.

The main benefit of adding a lens to the imaging model is that optical properties of a conventional camera, such as focus within a scene, are captured. The optical power of the lens is defined by its focal length, therefore it is the sensor depth,  $v$ , which determines the object plane that is in focus. Under paraxial approximation, the thin lens law is derived for an object in focus on the image sensor as

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{F} \quad (2.6)$$

where  $F$  is the lens focal length,  $u$  is object depth and  $v$  is the sensor depth. The importance of the sensor depth parameter is highlighted by this relationship. It defines  $u$  as the object depth which is conjugate to the sensor depth  $v$ . Consequently, an object point which does not lie on the plane at depth  $u$  will not appear in focus at the image plane. Generally, the focus property of a conventional camera is controlled by shifting the lens unit with respect to the fixed image sensor location. Two distinct object depths are shown in Figure 2.3. It is clear that the sensor depth  $v$  is conjugate to the object depth  $u_1$ . The object point at depth  $u_0$  corresponds to a conjugate plane (at distance  $v_0$ ) which is behind the image sensor location in its current configuration. Therefore, the image sensor will capture a blur circle since the object point is out of focus, and the “aperture” shape is circular. If the lens was shifted along the optical axis on the object side by the distance  $v_0 - v$ , then the object point at depth  $u_0$  would become focused on the image sensor. However, the object at

depth  $u_1$  would now form as a blur spot on the sensor. If the image sensor is located at the focal point of the lens ( $v = F$ ), the conjugate object plane in focus is the plane at infinity. This results in the object rays travelling parallel to the optical axis, and as such, object depths of smaller magnitudes will appear out of focus within the image. Typically, conventional imaging lenses allow the sensor depth value to be adjusted to focus object depths within the range of  $200mm$  to infinity.

The aperture diameter,  $D$ , which in this case is the lens diameter, plays a key role in the definition of the amount of image blur observed on the image plane. With reference to Figure 2.3, the object point at depth  $u_0$  is imaged on the sensor out of focus with a blur diameter of  $B$ . The aperture diameter can be seen as the limiting stop for object rays entering the imaging system. Thus, if the aperture diameter was adjustable, the incoming cone of light rays from the object point could be adjusted. Since this cone of rays limits the amount of light reaching the image sensor, it is obvious that the image blur diameter  $B$ , is proportional to the aperture diameter. As the diameter  $D$  is reduced, the imaged object point at  $u_0$  will appear sharper (more focused) on the image sensor. However, this comes at a cost of reduced illumination which was the initial motivating factor for introducing an imaging lens to the camera model.

### 2.1.3 Thick Lens Camera

The thin lens model introduced in the previous section represents the most basic lens model within the Gaussian optics domain. It is primarily applicable when the thickness of the lens can be assumed negligible. When the thickness of a lens can not be assumed negligible, such as the case with modern photographic lenses, an alternative Gaussian model is used called the thick lens model. Figure 2.4 presents the thick lens model. It is defined by the six cardinal points of Gaussian optics (Smith, 2000). These are the first and second focal points, the first and second principal points, and the first and second nodal points.

The focal points of a thick lens system are the points at which incoming rays, from an infinitely distant axial object point, will intersect the optical axis after refraction through the lens system. A focal point is defined on both the image and object sides in the same manner. The distance from the lens vertex to the

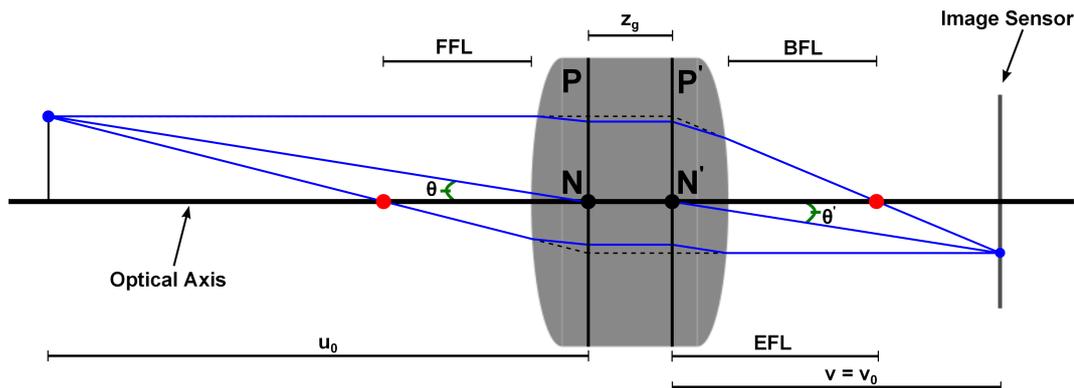


Figure 2.4: Thick lens camera model.

object space focal point is called the Front Focal Length (FFL) for the object side of the lens. Similarly, the distance from the lens vertex on its image side to the focal point in image space is termed the Back Focal Length (BFL). On closer examination of the rays entering and emerging from the optical system in Figure 2.4, it can be seen that a dashed line has been traced which represents the non-refracted path of these rays. The intersection of these rays which pass through both focal points define points on the principal planes of the optical system,  $P$  and  $P'$ . The principal planes are in fact spherical and centred about the focal point, however, under the paraxial approximation, these surfaces become planar. The intersections of each principal plane with the optical axis define the principal points of the lens. The Effective Focal Length (EFL) of the optical system can now be calculated as the distance from either principal point to its corresponding focal point. When dealing with optical systems which are tailored for imaging systems, this is typically the value given for the focal length of the lens. The two remaining cardinal points of a general optical system are the nodal points. A ray which is traced through the optical centre of the lens will have the properties that its incident angle will be equal to its refracted angle. Within a thick lens optical system, the incident and emerging rays will be displaced. If both rays are extended to intersect the optical axis, they define the nodal points of the optical system. Therefore nodal points have the property that rays directed towards the first nodal point emerge from the second nodal point parallel to the original direction. Thus,  $\theta$  is equal to  $\theta'$ , as shown in Figure 2.4. The nodal points will coincide with principal points if the index of refraction is the same in both image and object space. This is generally the case when dealing with imaging systems and lenses as it is air which will occupy the medium on both sides.

Once the cardinal points of the optical system are identified, rays can be traced through the system enforcing refraction to occur at the principal planes. This implies that the rays travel parallel to the optical axis between the principal planes. The distance between both principal planes is defined as the lens “thickness” and is shown in Figure 2.4 as  $z_g$ . If all distances are measured on the object and image side with respect to  $P$  and  $P'$ , then the thin lens law can be enforced. Therefore, from Figure 2.4 it is shown that the object and image distance are related to the effective focal length as

$$\frac{1}{u_0} + \frac{1}{v_0} = \frac{1}{EFL} \quad (2.7)$$

## 2.2 Camera Calibration

One of the most fundamental tasks in computer vision is the recovery of metric 3D information from 2D images. The projection of 3D metric information onto the 2D image is generally dependent on the camera model used. Calibration of a camera is the recovery of the camera parameters which describe this 3D to 2D projection. Therefore, when a camera is deemed calibrated, it is the camera model which is calibrated. This highlights the importance in the choice of camera model as it is these parameters which characterise the accuracy of the imaging process and of subsequent reconstructions.

Generally, the procedure for calibrating a camera involves imaging a structure or scene with known geometry. The relationship between the known geometry and image of this structure/scene is used to describe the projection and subsequently to estimate the camera model parameters. A camera is said to be fully calibrated when both its intrinsic and extrinsic parameters are known. The planar camera calibration methods, proposed independently by Zhang (1998) and Sturm and Maybank (1999), borrowed an approach from the self-calibration domain (Faugeras et al., 1992) in the formulation of their calibration algorithms. A geometric invariant property in the form of the IAC was used to capture properties of the imaging camera. A relationship between the internal camera parameters and the IAC was developed. It was shown that with three images of a planar calibration target, there are sufficient constraints to estimate all camera parameters. In addition to developing an accurate camera calibration algorithm, Zhang (1998) also reduced the input requirements of the camera calibration practitioner. A simple planar target with well de-

fixed feature points was the only requirement. This target was subsequently captured in varying orientations (minimum of three for full calibration). Thus, difficulties with traditional methods such as accurate scene point knowledge and physical setup were no longer an issue for calibrating a camera. These aspects of the planar approach contributed largely to its popularity. Readily available implementations<sup>2</sup> with standard calibration tools (Bouguet, 2008) quickly propelled the planar calibration approach to be the computer vision community's preferred method. Consequently, it is the planar camera calibration method based on the IAC which is dealt with in this thesis. In particular, two approaches are examined in detail.

### 2.2.1 Planar Camera Calibration

Two approaches are examined in this section. Zhang (1998) showed how the IAC was directly related to the internal parameters of the camera. Thus the planar calibration problem was reduced to fitting a conic to the IAC. One of the main issues with Zhang's approach was that it involved the use of an abstract projective geometrical entity in the form of the IAC. As a result, the linear stage of the calibration algorithm was in fact minimising an algebraic distance, which is known to cause instability due to the lack of physical meaning (Hartley and Zisserman, 2003). Consequently the second approach examined in this thesis, even though it is also based on the IAC, puts the calibration problem into a more intuitive geometric domain. Gurdjos and Payrissat (2001) exposed the relationship between the image plane and planar target in order to derive an expression for a real geometric entity on the image plane. This entity, the Centre Line (CL), was subsequently used to calculate the principal point of the image plane. This approach decoupled the intrinsic parameters of the camera for the purpose of calibrating a variable focal length camera. However, in this thesis, the interest is primarily in cameras with constant intrinsic parameters. The most important aspect of Gurdjos's method was the alternative geometric interpretation of the planar calibration problem. This interpretation allows the constraints of plane based calibration to be linearly represented by a geometric cost rather than the algebraic expression developed in Zhang (1998).

Properties of the Absolute Conic (AC) and its image are the basis for both

---

<sup>2</sup><http://opencv.willowgarage.com>

planar calibration methods under examination. The AC, represented mathematically as  $\Omega_\infty$ , is located on the plane at infinity,  $\pi_\infty$ , and is a conic consisting of purely imaginary points. It satisfies the equation

$$0 = \begin{cases} X_1^2 + X_2^2 + X_3^2 \\ X_4 \end{cases} \quad (2.8)$$

and has a radius of  $i = \sqrt{-1}$ , thus it is the conic  $\Omega_\infty = I$  on  $\pi_\infty$ . The invariant properties of the AC are exposed by examining the relationship between  $\pi_\infty$  and the image plane. If points on  $\pi_\infty$  are represented as  $\mathbf{X}_\infty = (\mathbf{d}^T, 0)^T$ , the image captured by a camera (of the form Eqn. 2.3) forms the image,  $\mathbf{x}$  of  $\mathbf{X}_\infty$  as,

$$\mathbf{x} = P\mathbf{X}_\infty = KR_c[I - C] \begin{pmatrix} \mathbf{d} \\ 0 \end{pmatrix} = KR_c\mathbf{d} \quad (2.9)$$

This identifies the relationship between the image plane and  $\pi_\infty$  as a planar homography  $H$  with  $H = KR_c$  and  $\mathbf{x} = H\mathbf{d}$ . Consequently, this mapping only depends on the intrinsic parameters of the camera,  $K$ , and the camera orientation with respect to the world coordinate frame  $R_c$ . Since the AC lies on  $\pi_\infty$ , its image can be determined via the planar homography  $H = KR_c$ ,

$$\begin{aligned} H^{-T}\Omega_\infty H^{-1} &= (KR_c)^{-T}I(KR_c)^{-1} = K^{-T}RR^{-1}K^{-1} = (KK^T)^{-1} \\ \boldsymbol{\omega} &= K^{-T}K^{-1} \end{aligned} \quad (2.10)$$

which defines the image of the absolute conic,  $\boldsymbol{\omega}$ . Similar to  $\Omega_\infty$ , the IAC is an imaginary point conic, however, as can be seen from the expression in Eqn. 2.10, it is only dependent on the intrinsic camera parameters. Therefore, it is an ideal tool for calibrating a camera as it is invariant to the camera's orientation and location. Once  $\boldsymbol{\omega}$  is determined,  $K$  can be calculated via Cholesky factorisation (Hartley and Zisserman, 2003).

### Zhang's Approach

Zhang's (1998) method of planar camera calibration requires imaging a planar target, in the form of a "chessboard" grid, in varying orientations. A pinhole camera model is assumed as described in Section 2.1.1. The planar calibration equations are derived by considering the planar target to be in the  $Z = 0$

plane of the world coordinate system. Thus, points on the planar target are projected onto the image sensor as,

$$\begin{aligned} \mathbf{x} &\simeq K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t}_c \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \\ &= K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}_c \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \end{aligned} \quad (2.11)$$

where  $\mathbf{r}_i$  denotes the  $i$ th column of the rotation matrix  $R_c$ . Therefore, the projection of the planar target points onto the image sensor can be described by a planar homography  $H$  where

$$H \simeq K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}_c \end{bmatrix}. \quad (2.12)$$

A planar homography can be estimated given a minimum of four point correspondences, no three of which can form a line, between the planar target and its image. The relationship in Eqn. 2.12 can be expressed as

$$\begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 & \mathbf{h}_3 \end{bmatrix} = \lambda K \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t}_c \end{bmatrix} \quad (2.13)$$

where  $\mathbf{h}_i$  denotes the  $i$ th column of the planar homography matrix  $H$  and  $\lambda$  is an arbitrary scalar factor. Exploiting the orthonormal properties of the rotation matrix yields two constraints from Eqn. 2.13,

$$\mathbf{h}_1^T K^{-T} K^{-1} \mathbf{h}_2 = 0 \quad (2.14)$$

$$\mathbf{h}_1^T K^{-T} K^{-1} \mathbf{h}_1 = \mathbf{h}_2^T K^{-T} K^{-1} \mathbf{h}_2 \quad (2.15)$$

where  $K^{-T} K^{-1}$  is the IAC. Consequently, each planar homography gives two constraints on the camera intrinsic parameters. Since there are five intrinsic parameters, three images of the planar target in varying orientations are required to uniquely solve for the parameters.

The two constraints obtained in Eqns. 2.14 and 2.15 can be geometrically interpreted as follows. The direction of the planar target in the world coordinate system has coordinates  $\mathbf{l} = [\mathbf{r}_1 \times \mathbf{r}_2]$ . This plane intersects with the plane at infinity in a line which, in turn, intersect the absolute conic in two points, the circular points  $[\mathbf{r}_1 \pm i\mathbf{r}_2]$ . The projection of the circular points into the image are found as  $K[\mathbf{r}_1 \pm i\mathbf{r}_2]$  which is equal to  $[\mathbf{h}_1 \pm i\mathbf{h}_2]$ . Since the circular

points intersect the AC on the plane at infinity, the image of the circular points  $(\mathbf{h}_1 \pm i\mathbf{h}_2)$ , must intersect  $\omega$ , the IAC. Therefore the expressions in Eqns. 2.14 and 2.15 are found by setting both real and imaginary parts of the imaged circular points to zero. This represents fitting a conic to the imaged circular points.

A closed-form linear solution to fitting a conic to the imaged circular points is outlined by Zhang. Firstly, the symmetric conic matrix for the IAC is defined. It consists of the intrinsic camera parameters.

$$\omega = K^{-T}K^{-1} = \begin{bmatrix} \frac{1}{(\alpha f)^2} & -\frac{s}{((\alpha f)^2 f)} & \frac{p_y s - p_x f}{(\alpha f)^2 f} \\ -\frac{s}{(\alpha f)^2 f} & \frac{s^2}{(\alpha f)^2 f^2} + \frac{1}{f^2} & -\frac{s(p_y s - p_x f)}{(\alpha f)^2 f^2} + \frac{p_y}{f^2} \\ \frac{p_y s - p_x f}{(\alpha f)^2 f} & \frac{s(p_y s - p_x f)}{(\alpha f)^2 f^2} - \frac{p_y}{f^2} & \frac{(p_y s - p_x f)^2}{(\alpha f)^2 f^2} + \frac{p_y^2}{f^2} + 1 \end{bmatrix} \quad (2.16)$$

Thus  $\omega$  can be represented as a 6D vector  $\mathbf{b} = [\omega_{11}, \omega_{12}, \omega_{22}, \omega_{13}, \omega_{23}, \omega_{33}]^T$  and a system of equations can be derived, based on the constraints on the IAC, in the form  $\mathbf{A}\mathbf{b} = 0$ . The  $\mathbf{A}$  matrix consists of  $2n$  rows where  $n$  is the number of images captured for the calibration and the two equations are the constraints outlined in Eqns. 2.14 and 2.15. Zhang conveniently denotes these entries as

$$\mathbf{h}_i^T \omega \mathbf{h}_j = \mathbf{a}_{ij}^T \mathbf{b}$$

where,

$$\mathbf{a}_{ij} = [h_{i1}h_{j1}, h_{i1}h_{j2} + h_{i2}h_{j1}, h_{i2}h_{j2}, h_{i3}h_{j1} + h_{i1}h_{j3}, h_{i3}h_{j2} + h_{i2}h_{j3}, h_{i3}h_{j3}]^T$$

thus  $\mathbf{A}\mathbf{b} = 0$  is a  $2n \times 6$  matrix and is expressed as,

$$\begin{bmatrix} \mathbf{a}_{12}^T \\ (\mathbf{a}_{11}, \mathbf{a}_{22})^T \end{bmatrix} \mathbf{b} = 0 \quad (2.17)$$

which can be solved using well known linear techniques, such as Singular Value Decomposition (SVD) (Golub and Van Loan, 1996). With three or more input images, a unique solution is obtained for the IAC, however, if the camera skew parameter,  $s$ , is considered to be zero, two images will suffice to estimate the conic. Once a solution for  $\mathbf{b}$  is found, the camera intrinsic parameters are

calculated as follows,

$$\begin{aligned}
 p_y &= \frac{\omega_{12}\omega_{13} - \omega_{11}\omega_{23}}{\omega_{11}\omega_{22} - \omega_{12}^2} \\
 \lambda &= \frac{\omega_{33} - [\omega_{13}^2 + p_y(\omega_{12}\omega_{13} - \omega_{11}\omega_{23})]}{\omega_{11}} \\
 \alpha f &= \sqrt{\frac{\lambda}{\omega_{11}}} \\
 f &= \sqrt{\frac{\lambda\omega_{11}}{\omega_{11}\omega_{22} - \omega_{12}^2}} \\
 s &= -\frac{\omega_{12}(\alpha f)^2 f}{\lambda} \\
 p_x &= \frac{sp_y}{f} - \frac{\omega_{13}(\alpha f)^2}{\lambda}
 \end{aligned} \tag{2.18}$$

allowing the camera extrinsic parameters to be calculated using Eqn. 2.13.

$$\begin{aligned}
 \mathbf{r}_1 &= \lambda K^{-1} \mathbf{h}_1 \\
 \mathbf{r}_2 &= \lambda K^{-1} \mathbf{h}_2 \\
 \mathbf{r}_3 &= \mathbf{r}_1 \times \mathbf{r}_2 \\
 \mathbf{t}_c &= \lambda K^{-1} \mathbf{h}_3
 \end{aligned} \tag{2.19}$$

This provides a linear solution for the camera intrinsic and extrinsic parameters. One problem with this solution is the fact that the distance being minimised in calculating the camera parameters is algebraic. Algebraic distances have no physical meaning in minimisation problems and are known to cause instability in the estimates, particularly in the presence of noise (Hartley and Zisserman, 2003). Therefore, Zhang proposes a geometric refinement of the camera parameter estimates in the form of a Maximum Likelihood Estimation (MLE). In addition to the MLE, a standard two term radial distortion model ( $c_1, c_2$ ) is imposed on the system (Brown, 1971). Consequently, the cost function minimises the geometric distance between the projected planar target points in the image,  $\hat{\mathbf{x}}$ , under the model in which the parameters are being estimated, and the observed image points,  $\mathbf{x}$ .

$$\sum_{i=1}^n \sum_{j=1}^m \|\mathbf{x}_{ij} - \hat{\mathbf{x}}_{ij}(K, c_1, c_2, R_i, t_i, X_j)\| \tag{2.20}$$

This is a non-linear minimisation problem which is solved using the Levenberg-Marquardt (LM) algorithm. Initial estimates of the parameters in Eqn. 2.20 are taken as the results of the linear estimation stage with  $c_1$  and  $c_2$  set to zero. In summary, Zhang's planar calibration method involves (1) Imaging a

planar target in a minimum of three varying orientations (2) Detecting the planar target feature points (3) Linearly estimating the intrinsic and extrinsic camera parameters using Eqns. 2.17, 2.18 and 2.19 (4) Refining all parameter estimates non-linearly via Eqn. 2.20.

An important aspect of any camera calibration algorithm is awareness of degenerate configurations. These are defined as planar target orientations which do not provide any additional constraints on the intrinsic parameters. Typically this implies that there is an inherent dependency in the system of calibration equations. Since the calibration equations are based on the properties of the rotation matrix, there are a large number of possible orientations in which degenerate cases can arise. This is particularly true in the case of two-plane calibration (assuming zero skew), for which Sturm and Maybank (1999) outline a large family of singularities.

### Alternative Approach

Gurdjos and Payrissat (2001) proposed an alternative approach to the planar calibration problem by introducing the theorem of Poncelet (1862) to the domain. The theorem geometrically proved that two planar figures, for which the first is the central projection of the second, remain in perspective correspondence when one rotates the first plane about its intersection with the second plane. The corresponding centres of projection describe a circle which lies within a plane orthogonal to the intersection of both planes. Figure 2.5 presents a diagram to aid with visualisation of the theorem and its relevance to planar camera calibration. A world plane intersects the image plane in a line, the intersection line. If this line is taken as an axis, about which the world plane may rotate freely, there is an infinite number of world planes which remain in homographic correspondence with the generated image on the image plane. Thus a planar homography  $H$  represents the transformation. Each world plane has a unique centre of projection through which it remains in homographic correspondence with the image. The locus of these centres of projection forms the Centre Circle (CC). This circle lies in a plane orthogonal to the intersection line. The orthogonal projection of the CC onto the image plane forms a line, the CL. Thus, for each projected centre of projection onto the image plane, there is an associated principal point on the CL. When multiple planar targets are captured in varying orientations, each image CL will

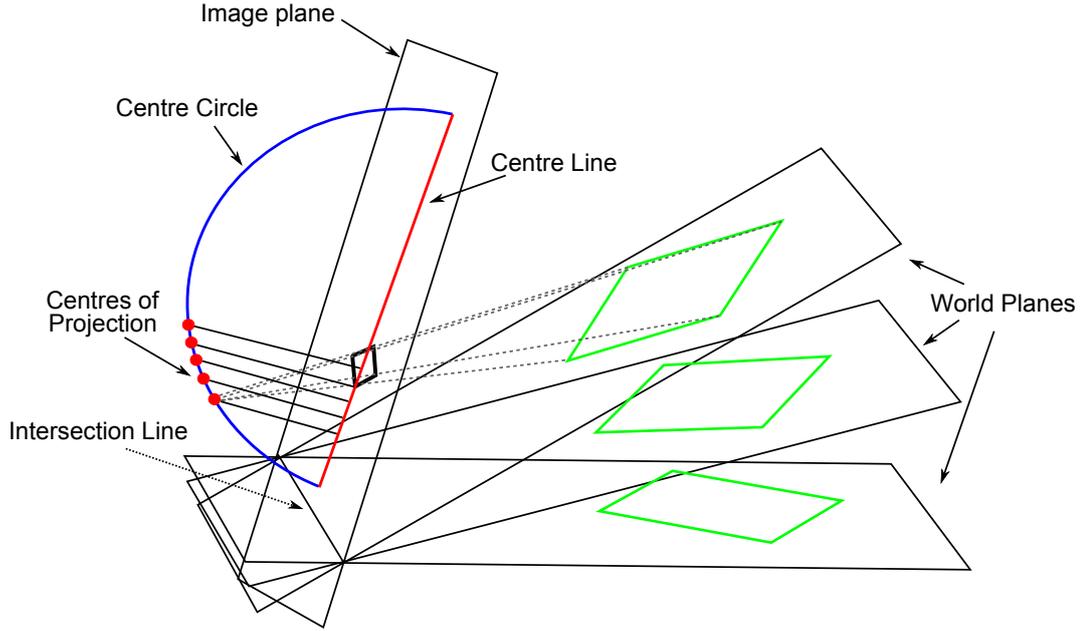


Figure 2.5: Poncelet's theorem and its application to planar camera calibration (Gurdjos and Payrissat, 2001).

intersect at the true camera principal point. This allows a partial calibration of the intrinsic parameters ( $p_x$ ,  $p_y$  and  $\alpha$ ). Gurdjos decouples the intrinsic parameters to calibrate the focal length in a second step based on the estimated principal point and aspect ratio.

An expression for the CL can be derived from properties of the planar homography based on the constraints of Eqns. 2.14 and 2.15. If zero skew is assumed, the IAC is defined as (after scaling by  $(\alpha f)^2$ ),

$$\omega = \begin{bmatrix} 1 & 0 & -p_x \\ 0 & \alpha^2 & \alpha^2 p_y \\ -p_x & -\alpha^2 p_y & \alpha^2 f^2 + p_x^2 + \alpha^2 p_y^2 \end{bmatrix} \quad (2.21)$$

now by algebraically eliminating the  $\omega_{33}$ , of the IAC, a linear equation in  $\omega$  is achieved. With further refinement of the constraints in Eqns. 2.14 and 2.15 an expression is obtained:

$$\varphi^T [\omega_{13}, \omega_{23}, \omega_{22}, \omega_{11}]^T = 0 \quad (2.22)$$

where  $\varphi^T = [\varphi_1, \varphi_2, \varphi_3, \varphi_4]^T$ , and,

$$\begin{aligned}
 \varphi_1 &= (h_{31}^2 + h_{32}^2)(h_{31}h_{12} - h_{11}h_{32}) \\
 \varphi_2 &= (h_{31}^2 + h_{32}^2)(h_{31}h_{22} - h_{21}h_{32}) \\
 \varphi_3 &= (h_{31}h_{12} - h_{11}h_{32})(h_{11}h_{31} + h_{12}h_{32}) \\
 \varphi_4 &= (h_{31}h_{22} - h_{21}h_{32})(h_{21}h_{31} + h_{22}h_{32})
 \end{aligned} \tag{2.23}$$

which yields the equation for a line,  $l$ , containing the principal point.

$$\begin{aligned}
 p_y &= -\frac{\varphi_1\alpha^2}{\varphi_2}p_x + \frac{\varphi_3\alpha^2 + \varphi_4}{\varphi_2} \\
 l &= [-\varphi_1\alpha^2, -\varphi_2, \varphi_3\alpha^2 + \varphi_4]^T
 \end{aligned} \tag{2.24}$$

Confirmation that the line  $l$ , in the image frame, and the CL are equivalent is achieved by illustrating its orthogonality with respect to the vanishing line,  $l_\infty$ , of the planar target. The vanishing line is calculated as the third row of  $adj(H)$  and is equal to,  $l_\infty^T = [h_{21}h_{32} - h_{31}h_{21}, h_{31}h_{12} - h_{11}h_{32}, h_{11}h_{22} - h_{12}h_{21}]$ . Orthogonality between the VL and CL satisfies  $l^T\omega^*l_\infty = 0$ , where  $\omega^*$  is the dual IAC and represented in this case as  $diag(1, 1/\alpha^2, 0)$  (Hartley and Zisserman, 2003).

The function which is minimised when estimating the principal point and aspect ratio is linear and based on the constraint on these parameters given the CL equation (Eqn. 2.24). Thus the solution can be found linearly by minimising Eqn. 2.22 for all input images. The main benefit of Gurdjos' approach to planar calibration is the transformation of the minimisation function into a linear geometric cost function. The previous cost of Eqn. 2.22 is algebraic and therefore has no physical meaning. However, this function can be transformed into a sum of squared Euclidean distances based on the distance between the principal point and the CL. Normalisation of Eqn. 2.22 is performed with the aspect ratio approximated to unity. Thus the function to be minimised is,

$$\arg \min \left\{ \sum_{j=1}^n d^2(m_0, l^{(j)}) \right\} \tag{2.25}$$

where  $d(m_0, l^j)$  is the Euclidean distance from the principal point,  $m_0$ , to the CL  $l^j$ . This can be solved using linear techniques such as the SVD.

A linear expression is developed for subsequent calculation of the camera focal length. A decoupled camera calibration matrix  $K_d$  is formed with the estimated principal point and aspect ratio. For each image of the planar target,

$G = K_d^{-1}H$ . This forms a modified IAC of  $K_d^T \omega K_d \sim \text{diag}(1, 1, f^2)$ . This problem can be seen as solving for  $F = f^2$  the equation  $\Psi[1, F] = 0$  where,

$$\Psi = \begin{bmatrix} G_{11}^2 + G_{21}^2 - G_{12}^2 - G_{22}^2 & G_{31}^2 - G_{32}^2 \\ G_{11}G_{12} + G_{21}G_{22} & G_{31}G_{32} \end{bmatrix} \quad (2.26)$$

in which a linear least-squares solution can be obtained from  $\Psi$ .

Aside from providing a linear method to calculating the camera intrinsic parameters, the CL also provides a geometrical insight into singularities within the system. Degenerate configurations in the recovery of the principal point arise when no clear principal point can be defined. This is the case when CLs of different views are parallel. Such configurations arise when the planar target undergoes translation only, rotation about a parallel axis, and rotation about an axis orthogonal to the plane. In terms of the recovery of focal length, degenerate cases arise when the camera and world coordinate systems are parallel in the X or Y direction as noted by Sturm and Maybank (1999). Both methods of planar camera calibration examined in this section are based on the fundamentals of the IAC and the constraints of fitting a conic to this image. However, it is the geometric domain in which Gurdjos presents the problem, in the form of the CL, which is explored further in Chapter 3.

## 2.3 Significance of the Pupil Plane

The properties of a camera’s aperture, on the pupil plane, are an integral component to describing the true process of imaging 3D world points to the 2D image sensor. Three imaging models have been outlined in the form of the pinhole, thin-lens, and thick-lens models. The pinhole model, which is the simplest of the three models, has been examined in detail. In particular, the calibration of the pinhole model has been outlined. Camera calibration recovers the camera model properties which describe the 3D to 2D projection of world to image points. The model parameters which are recovered are the focal length,  $f$ , the principal point,  $(p_x, p_y)$ , the aspect ratio ( $\alpha$ ) and the camera skew ( $s$ ). One of the primary goals of this thesis is to recover real depth information given a single image of a scene from a single camera. However, this task is not directly possible within the constraints of the pinhole model. A closer examination of the model for central projection (Eqn. 2.1) shows how

object depth information is directly lost as the world point  $(X, Y, Z)$  is projected to the image as  $(fX/Z, fY/Z)$ . Thus direct depth recovery from a single image using the pinhole model is not possible. Generally, when modelling the camera as pinhole, and attempting to recover depth, a stereo configuration is implemented. This consists of two cameras/sensors in which the baseline distance between both camera centres is known. Typically, image pixel disparities can be calculated between both camera images and subsequent triangulation methods allow direct recovery of object depth. The main disadvantage with this approach is that two cameras/sensors are required. The equivalence of  $f$  within the pinhole model and  $v$ , the sensor depth, of the lens models has been highlighted. Within the constraints of a single sensor and single image, knowledge of the sensor depth alone is not adequate for depth recovery. Additional information is required which is not available by modelling the camera as pinhole.

Camera models incorporating lenses were discussed in Section 2.1. An expression for object depth is presented which depends on knowledge of the lens EFL and the conjugate object distance i.e. focused image point distance. The addition of a lens to the imaging model introduces new properties to the camera model which can be exploited for the purpose of depth recovery. The optical phenomenon of focusing and subsequent image blur offers a new image cue which encodes object depth information. Given an image of a natural scene, certain objects within that scene will appear to be more well defined (in better focus) than in other areas of the image. These “in focus” areas generally correspond to the focus plane which can be found with knowledge of the focal length of the lens and the sensor depth value. Objects in front and behind of this focus plane, within the scene, exhibit varying amounts of image blur. The magnitude of this image blur is depth dependent. In addition to being depth dependent, the amount of blur generated at each depth within the image is defined by the pupil within the pupil plane of the imaging system. It was shown in Section 2.1 that the focusing properties of the camera models incorporating a lens depend on properties of the pupil on the pupil plane. Specifically, the physical size of the pupil, which in this case is its diameter, effectively controls the amount of blur within the image. The relationship between pupil diameter and blur radius was identified in Figure 2.3 where  $B$  is proportional to  $D$ . This connection between pupil diameter and blur diameter has been well studied in the area of DfD (Pentland, 1987, Pentland et al., 1989, Subbarao and Wei,

1992, Subbarao and Surya, 1994). However, these methods require multiple images with varying camera parameter settings to estimate object depth.

The pupil and pupil plane are an important aspect of Gaussian optics and of the image formation properties of imaging lenses. Conventional imaging lenses consist of multiple lens elements coupled with multiple stops. The effect of a stop within an optical system is to limit the amount of light entering the system or striking the image sensor. Typically, there will be one stop or diaphragm within the system of lens elements which is adjustable. This is known as the aperture stop or pupil stop. This pupil is the limiting factor for light rays from axial object points entering and exiting the system. Lens systems considered up to this point have assumed a pupil plane located at the centre of the lens. The effect of adjusting the diameter of the pupil on the pupil plane has been observed in the form of image focus and blur. The position of the pupil plane and the diameter of its pupil play an integral role in image formation.

Shifting the pupil plane either side of the effective lens centre gives rise to the entrance and exit irises. The entrance iris is the image of the pupil plane from the object side of the optical system. Conversely, the exit iris is the image of the pupil plane as seen from the image side of the optical system. Thus, the entrance and exit iris are conjugate. In both cases, the diameter of the iris is defined by the diameter of the pupil on the pupil plane and the optical power of the lens elements through which it is being projected. Figures 2.6(a) and 2.6(b) present two optical configurations with the entrance and exit iris shown. In the case of Figure 2.6(a), the pupil plane is situated on the object side of the imaging lens. Therefore the entrance iris is equivalent to the pupil plane position and pupil diameter since there is no optical element between the pupil plane and the object being imaged. The exit iris is defined as the image of the pupil plane and pupil from the image side of the imaging lens. This image corresponds to the exit iris lying closer to the object than the entrance iris. Figure 2.6(b) presents the case with the pupil plane on the image side of the imaging lens. Now the exit iris and pupil plane are coincident while the entrance iris is the image of the pupil plane from the object side of the lens. The most significant property of the entrance and exit iris is that they represent the limiting pupils through which all axial rays must pass. It is seen that the cone of light accepted by the optical system in object space has the entrance iris as its limiting stop. Similarly, the exit iris is the limiting stop of the cone of rays which emanate from the optical system and strike the image

sensor. A representation for a general optical system including the entrance and exit iris is shown in Figure 2.6(c). An off-axis object is imaged where two rays are depicted. The axial ray emanates from the optical axis and has the property that it strikes the limit of the entrance and exit iris. The chief ray emanates from the off-axis object point and passes through the centre of the entrance iris, and, the centre of the exit iris. Additionally, the incident and refracted angle of these rays are equal.

Awareness of the optical properties of imaging lenses is of utmost importance, particularly when considering properties of the pupil plane as a cue for depth recovery. Two different properties of the pupil plane are examined in this section. Firstly, the effects of shifting the pupil plane axially are examined. Secondly, the effects of shifting the pupil on the pupil plane are outlined.

### 2.3.1 Location of Pupil Plane

The importance of the location of the pupil plane has been highlighted. It is the limiting stop for all axial light rays which pass through the optical system. Optical properties, in the form of optical irises, have been introduced and shown to be conjugate images, through the lens, of the pupil plane. The entrance iris is the limiting pupil for rays which enter the system from object space. When the pupil plane is located in object space, it becomes the limiting factor for light rays entering the optical system (as shown in Figure 2.6(a)). Thus it is defined as the entrance iris. As a result, the entrance iris diameter is equal to the pupil plane diameter, and the relationship between the magnitude of observed blur on the image plane is directly proportional to this pupil diameter within the pupil plane. However, placement of the pupil plane within image space leads to a different scenario.

When placing the pupil plane in image space, it must be noted that the limiting factor for rays entering the optical system, the entrance iris, is in fact the conjugate image of the pupil plane through the optical system. Thus, the diameter of the entrance iris may not necessarily be equal to the diameter of the pupil within the pupil plane (see Figure 2.6(b)). This results in a disproportionate relationship between the pupil plane diameter and the observed blur on the image plane. A number of experiments were conducted to identify this relationship and highlight the effects of placing the pupil plane in image space.

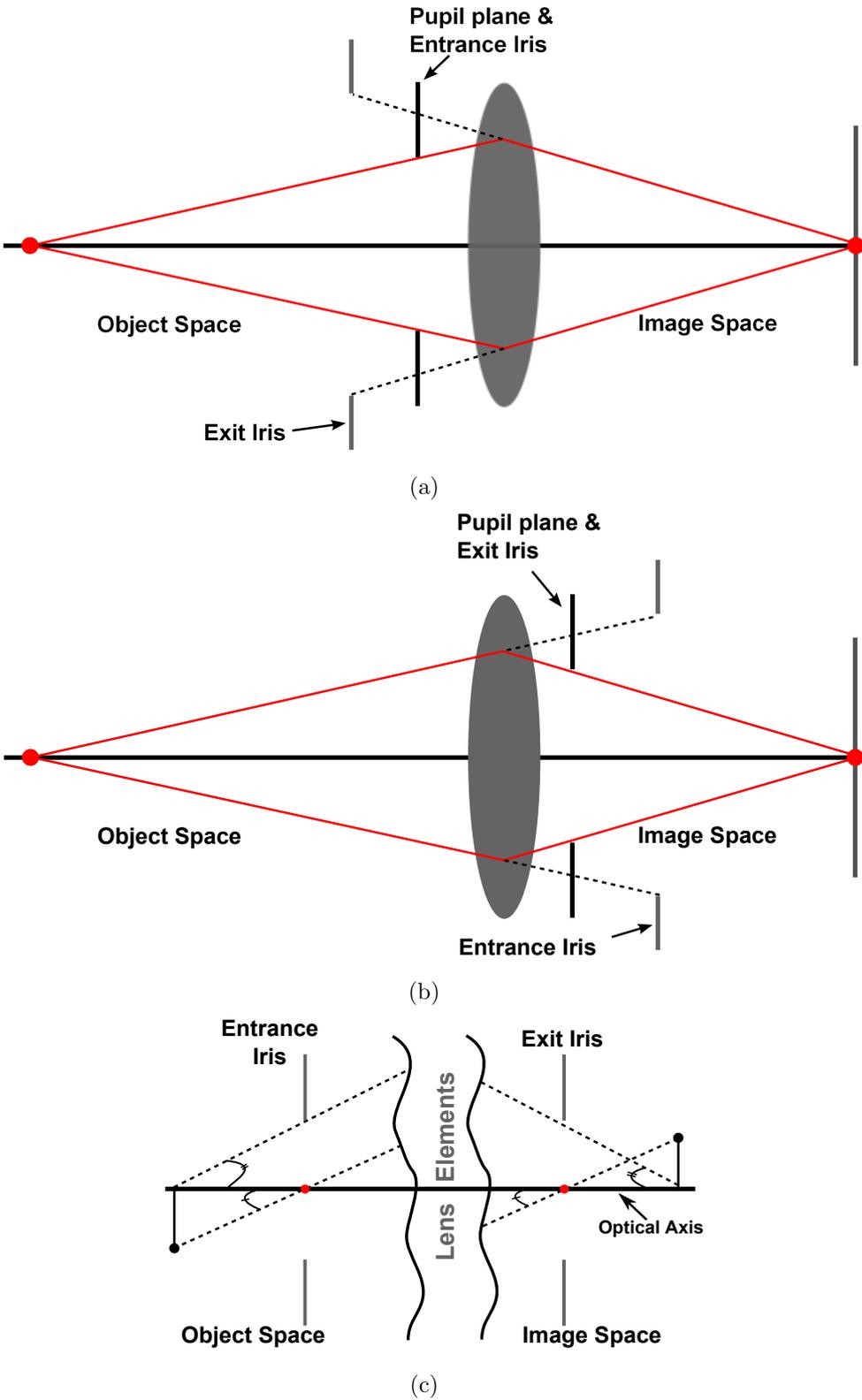


Figure 2.6: Placement of pupil plane and its overall effect on optical irises  
 (a) Pupil plane in object space ; (b) Pupil plane in image space ; (c) General optical configuration depicting both entrance and exit irises

Figures 2.7(a)-2.7(f) present experiments conducted using the optical design software Zemax<sup>3</sup>. Zemax is the industry standard optical design software and contains libraries for numerous lens catalogues. Its ray-tracing algorithms account for all optical properties of the given lenses. Edmund Optics (EO) catalog lenses were chosen for this experiment as they are the lenses which are used in the real experiments in Chapter 4. The aim of this experiment was to highlight the difference in magnitude of the recorded image blur diameter given a system with its pupil plane located in object space, and in image space. An ideal lens (thin lens) was also simulated for comparative purposes. The lens chosen for this simulation was the EO achromat with focal length 75mm and lens diameter of 25mm. The optical system was configured to a near-focus setting (approx. 300mm) and the pupil plane was fixed at 30mm in front of the lens for the object space test. Similarly, the pupil plane was set at 30mm behind the imaging lens for the image space experiment. With the ideal lens, the pupil plane is located at the centre of the lens. For each experiment, six different pupil plane diameters were implemented, with ten axial object depths captured for each diameter. The Point Spread Function (PSF) is measured for each object depth point imaged. The PSF defines the response of an imaging system to an object point or point source. Therefore it describes object point rays which pass through the optical system and strike the image sensor. Zemax provides exact measurements for the width of the PSF.

The main conclusion from the experimental results presented in Figures 2.7(a)-2.7(f) is that the diameter of the PSF, with the pupil plane located in object space, displays the same properties as the ideal imaging system across all object depths. The similarity in performance between the the ideal system and the system with the pupil plane in object space is due to the incoming cone of rays being directly related to the number of rays captured on the image sensor. With the pupil plane located in image space, the limiting stop for incoming rays is defined by the systems' entrance iris pupil diameter, which after projection through the lens, is not equal to that of the pupil diameter within the pupil plane. Similar performance for all axial object depths is achieved with a pupil plane diameter of 25mm as shown in Figure 2.7(a). This is due to the fact that the entrance iris pupil diameter generated with a pupil plane diameter of 25mm in image space is greater than 25mm. However, since the lens diameter is 25mm, the lens becomes the limiting factor for rays entering the optical system

---

<sup>3</sup>[www.zemax.com](http://www.zemax.com)

and the exit iris, in this case the pupil plane, does not affect the relationship between incoming rays and the PSF. The difference in magnitude of the PSF diameter for Figures 2.7(b)-2.7(f) is due to the entrance iris diameter being greater than that of the pupil plane located in image space.

A closer examination of the relationship between the recorded PSF diameter at particular object depths and the pupil plane diameter exposes the effects of pupil plane placement for optical systems. Figure 2.8(a) presents the results for PSF diameter versus pupil plane diameter for each object depth with the pupil plane located in object space. Similarly, Figure 2.8(b) presents the same criteria for the pupil plane located in image space. As expected, the relationship is linear when the pupil plane is in object space. This is due to the pupil plane and entrance iris being equivalent. Therefore when decreasing the pupil plane diameter, a proportional decrease in PSF diameter is observed on the image plane. With placement of the pupil plane in image space, the nonlinear relationship is exposed between PSF diameter and pupil plane diameter. When the pupil plane diameter increases beyond 12.5mm, the PSF diameter does not increase proportionately. This is a result of the actual entrance iris diameter being greater than the lens diameter, causing the lens diameter to become the limiting pupil for incoming rays. Thus the PSF diameter remains constant despite increases in pupil plane diameter. This is an important result in terms of using the relationship between the imaged object point and the pupil plane diameter as a cue for depth information. The linear nature of this relationship with the pupil plane in object space is more desirable for developing a depth dependent model.

The effects of a pupil plane located remotely to the lens in object space is further examined with the experimental results presented in Figure 2.9(a) and 2.9(b). It has been established that when the pupil plane is located in object space, it is equivalent to the entrance iris of the optical system. Due to the linear nature of the relationship of the pupil plane diameter in object space, a remote pupil plane can be modelled by an equivalent pupil plane at the centre of the lens. This equivalent pupil diameter will be slightly greater in magnitude than the physical pupil plane. This is due to the axial rays striking the lens marginally outside the limits of the physical pupil plane (Figure 2.6(a) illustrates this behavior at the lens centre). The results outlined in Figure 2.9(a) and 2.9(b) were captured with a pupil plane located 30mm in front of the lens in object space (Zemax data used with the 75mm focal length EO

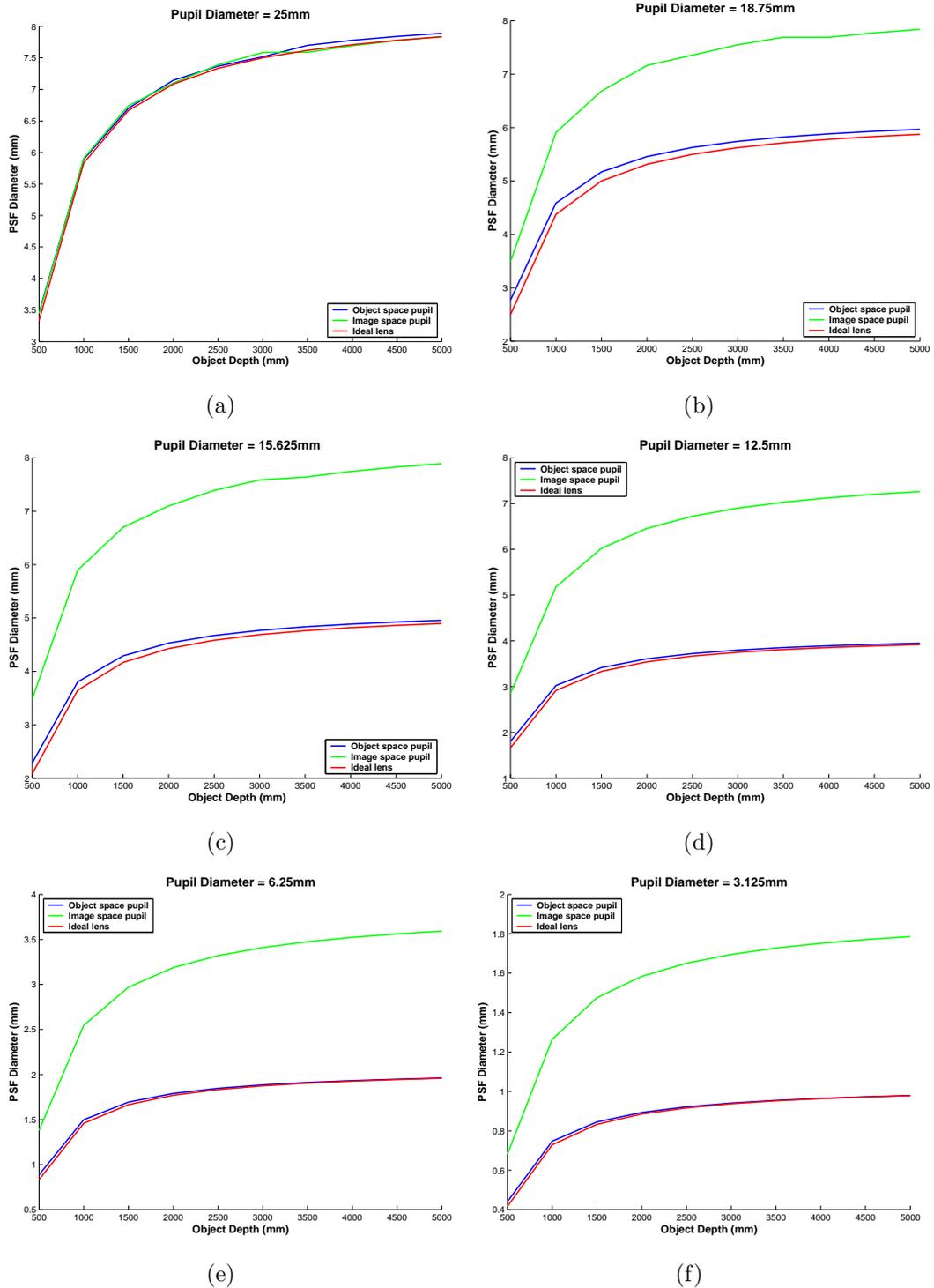


Figure 2.7: Placement of pupil plane and overall effect of optical irises on the PSF (a) Pupil diameter 25mm; (b) Pupil diameter 18.75mm; (c) Pupil diameter 15.625mm; (d) Pupil diameter 12.5mm; (e) Pupil diameter 6.25mm; (f) Pupil diameter 3.125mm

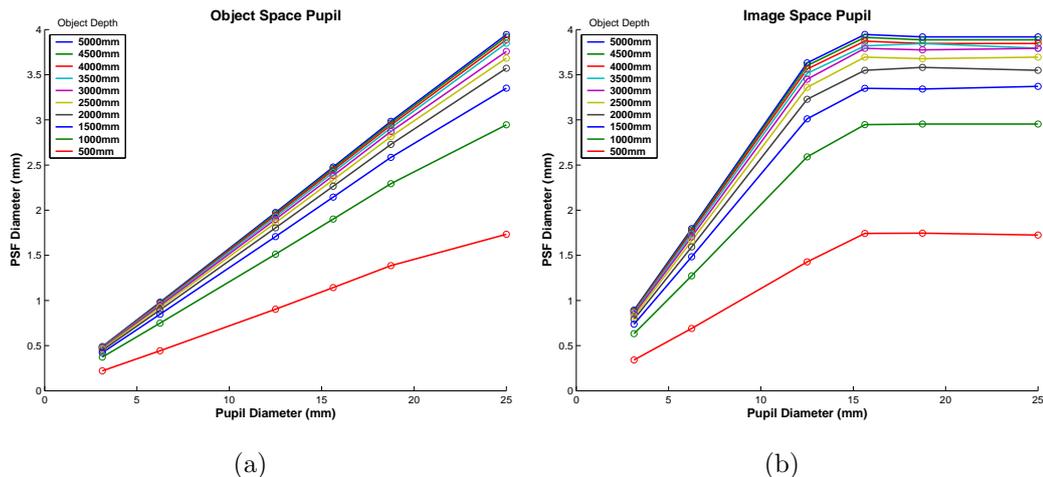


Figure 2.8: Placement of pupil plane and overall effect on optical irises (a) Pupil plane in object space; (b) Pupil plane in image space

lens). Four different pupil diameters within the pupil plane were examined, all of which were less than the lens diameter. Pupil diameters greater than the lens diameter no longer represent the limiting pupil in the system. The solid lines in both figures represents the actual PSF diameter recorded at each object depth at the given remote pupil plane pupil diameter. The circular points represent the data recorded employing the given pupil plane diameter at the imaging lens. Thus, the difference between Figure 2.9(a) and Figure 2.9(b) is that the pupil plane diameter in Figure 2.9(b) has been adjusted to account for the physical remote pupil plane. In contrast, Figure 2.9(a) represents a pupil plane at the lens of the same diameter as the remote pupil plane. It is clear from the results that the adjusted pupil diameter at the lens models the captured image data more accurately than the non-adjusted case.

If knowledge of all camera parameters is available, it is possible to calculate an equivalent pupil at the lens based on the recorded PSF diameter. The practicality of such an equivalent aperture is questionable due to the sensitivity of absolute measurements in image space. The placement of the pupil plane in image space could not be guaranteed with sufficient accuracy to calculate an equivalent aperture. A number of systems have been proposed in the form of “plenoptic cameras” which insert microlens arrays in the image space of the optical system (Adelson and Wang, 1992, Ng et al., 2005, Lumsdaine and Georgiev, 2009). Typically, the placement of the microlens array, which is located at the focal plane in image space of the imaging system (see Figure 1.1(b)), must be highly accurate i.e.  $36\mu\text{m}$  as reported by Ng et al. (2005).

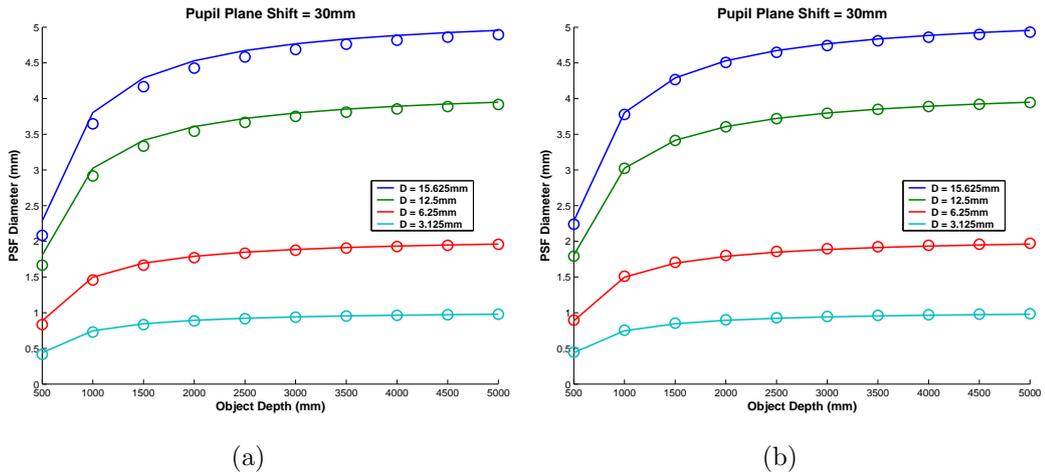


Figure 2.9: Equivalent pupil diameter after axial shift (a) Non-adjusted pupil diameter; (b) Adjusted object space pupil diameter.

These sensitivity issues are avoided with planes in the object space of the imaging system.

Properties of the pupil plane also demonstrate the equivalence between the camera models presented in this chapter. By choosing a suitable location for the pupil plane, it is seen that the lens models are related. Generally, the thick lens model pupil plane diameter is equal to the principal plane diameter, and is located at the first principal plane. Figure 2.10 presents the thick lens, thin lens and pinhole camera models. The thick lens model is defined by its two principal planes which are in unit magnification. The object ray passing through the first principal point of the thick lens emerges from the second principal point with equal angle. If both principal planes are amalgamated and the sensor depth,  $v$ , remains constant, then the thick lens can be equivalently modelled by a thin lens. The thin lens must have its pupil plane (central plane) located at the same position as the first principal plane of the thick lens, and, the sensor depth must be equal to the thick lens sensor depth. This equivalent configuration is depicted in Figure 2.10. Similarly, the thin lens model can be geometrically modelled by the pinhole model. This is achieved by replacing the lens plane with the pinhole pupil and maintaining the image sensor to pinhole distance,  $v$ . Thus the pinhole model is also geometrically equivalent to the thick lens model. The significance of this geometric equivalence between both lens models is that the pupil plane can be remotely located in object space and an equivalent pupil plane and diameter can be defined at the lens principal plane. Consequently, imaging lenses, which generally have multiple

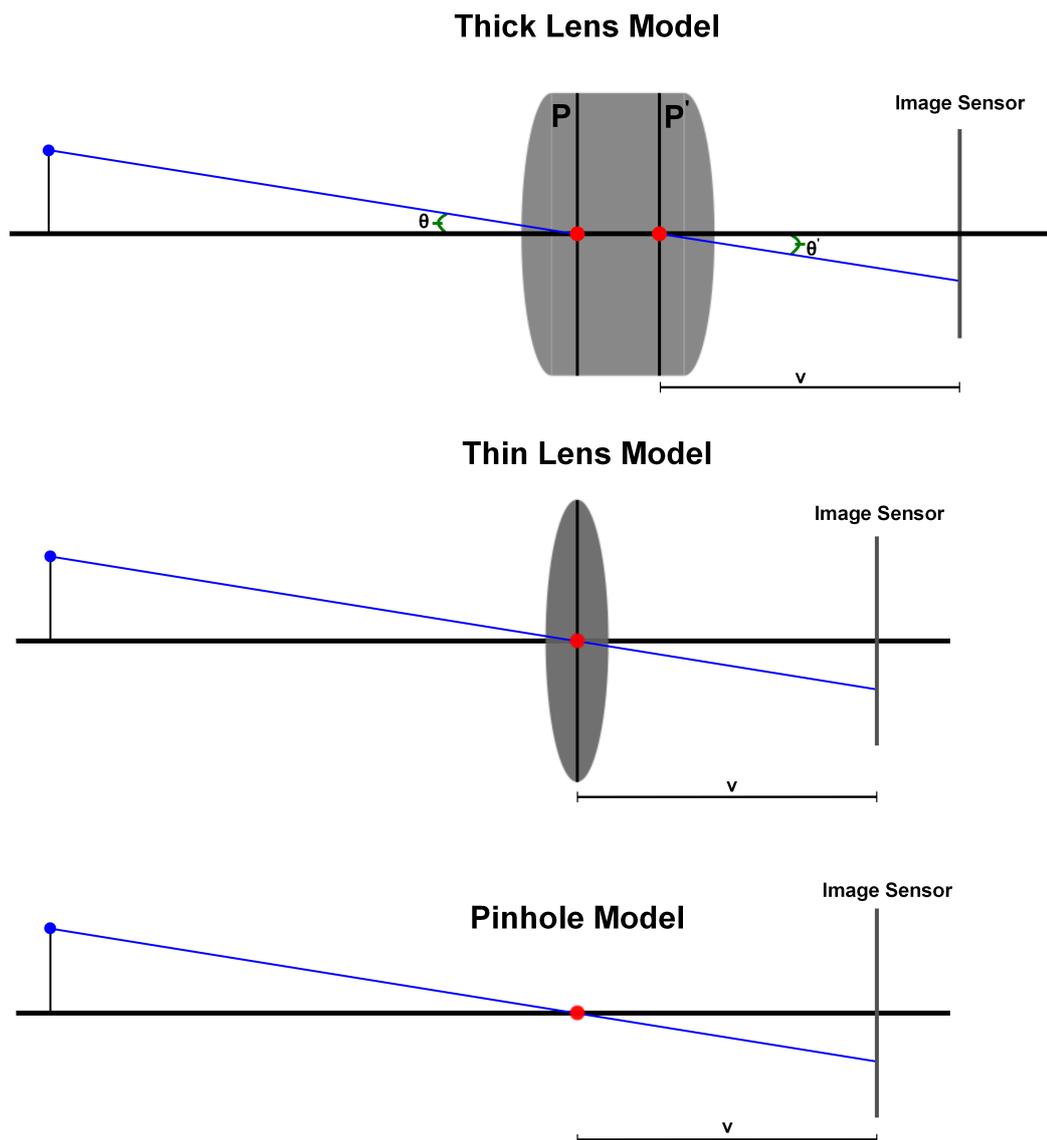


Figure 2.10: Equivalence between camera models.

lens elements, can be modelled by the thin lens camera model provided the principal planes are amalgamated. Furthermore, a remote pupil plane can be defined and geometrically accounted for.

### 2.3.2 Location of Pupil on Pupil Plane

The pupil located on the pupil plane is defined as the circular opening on the opaque pupil plane in which light rays may pass through for image formation. Within conventional imaging systems, such as those modelled by the thin/thick lens models, the pupil is centrally aligned with the optical axis of the system. Since the pupil is circular, it is rotationally symmetric about the optical axis

and is of diameter  $D$  (as in Figure 2.6). The equivalence of camera models examined in this chapter was outlined in Section 2.3.1. It was shown that lens models reduced to pinhole when the image sensor and pupil plane/pinhole configurations were equivalent. That is, the distance of the pupil plane to image sensor and pinhole to image sensor were equivalent. Thus the lens model, which captures a cone of rays from the object point, is geometrically equivalent to the pinhole model which captures a single ray through its pinhole. Since the pupil within the pupil plane is centrally located about the optical axis, the corresponding cone of rays passes through the equivalent pinhole of the lens model. Furthermore, as the pupil is rotationally symmetric about the optical axis, the image formed by the cone of rays will be symmetric on the image plane. Therefore, an object point which lies outside of the focused imaging configuration will be imaged as a “blur circle” on the image plane. The centroid of this blur circle is the cone axial ray i.e. the equivalent image of the object point through the pinhole model. Consequently, the geometric mapping of the cone of rays to the image plane can be approximated by a central ray passing through the pupil and centred on the imaged circle.

The relationship between a single centred ray and the pupil of the pupil plane has been established. An experiment is conducted to verify that this relationship can be extended to sub-sampled pupils within the current pupil on the pupil plane. The experiment was conducted using Zemax and the same lens as in Section 2.3.1. The imaging configuration was set to focus at infinity i.e  $v = F$ , and the pupil plane was located at the lens apex (thus becoming the entrance pupil) and consisted of a conventional pupil of diameter 25mm. The object distance was set to 3m, which in turn ensured that the image would be out of focus and thus would result in a blur circle. Figure 2.11 presents the configuration of the pupil plane for the conducted experiment. The symmetric properties of the pupil have been identified, therefore in order to demonstrate its sub-sampling properties, a number of sub-pupils are implemented at various diameters within the conventional pupil. Besides providing accurate information on each ray which passes through the optical system, Zemax also allows the formation of non-conventional pupil planes within the system. Therefore the experiment consisted of recording the image data for conventional pupils of diameters 3mm, 5mm, 10mm, and 15mm. Subsequently, the experiment was repeated using sub-sampled pupils of diameter 1mm centred on the pupil plane at the aforementioned conventional diameters (red pupils in Figure 2.11).

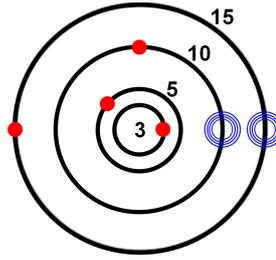


Figure 2.11: Pupil plane examination for conducted experiments. Sub-sampling pupils (red) are 1mm. Pupil diameter variation (blue).

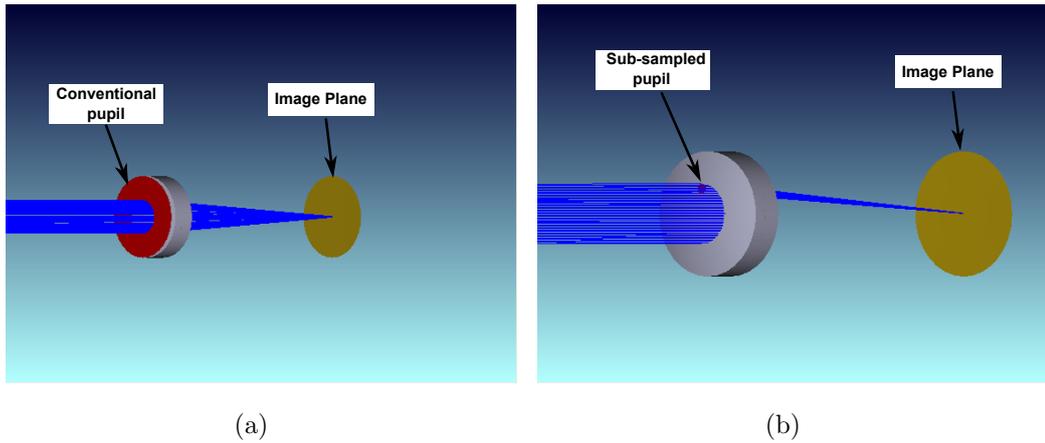


Figure 2.12: Experimental setup for pupil sub-sampling (a) Conventional pupil; (b) Sub-sampled pupil.

Accuracy of the sub-sampling was evaluated by calculating the centroid of the recorded image circle and comparing it with the image data at the corresponding location of the conventional pupil at that diameter. The experimental setup is shown in Figures 2.12(a) and 2.12(b).

Results for the sub-sampling are presented in Table. 2.1 and the accuracy is visualised in Figure 2.13. The accuracy has been defined in terms of the pixel pitch. Pixel pitch is the physical size of a single pixel on the image sensor, which in the case of the Panasonic Lumix G1 (used in chapter 4), is  $4.25\mu\text{m}$ . The results indicate that the centre of the image formed by the sub-sampled pupil is representative of a single ray passing through a conventional pupil with the same diameter as the sub-sampled pupil offset. Of the four offsets examined, the largest error recorded was 0.015 pixels while the minimum was 0.008 pixels. The trend suggests that as the sub-pupil offset is increased within the conventional pupil, an increase in error is observed. This increase, which is extremely small in magnitude ( $1/142$  of a pixel) is due to spherical aberration.

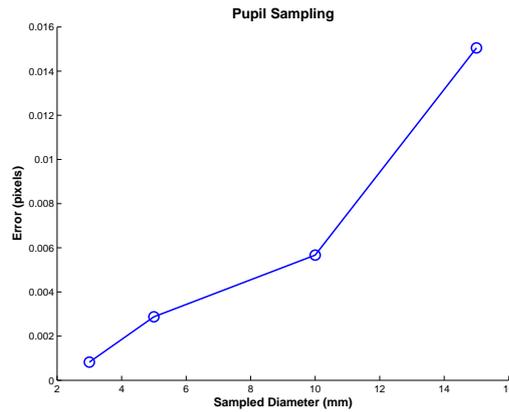


Figure 2.13: Error in pupil plane sampling via sub pupils of diameter 1mm.

Table 2.1: Sub-sampling pupil results.

Conventional Pupil size	3mm	5mm	10mm	15mm
Conventional measured PSF ( $\mu m$ )	25.476	42.179	82.129	120.121
Sub-sampled measured PSF ( $\mu m$ )	25.480	42.167	82.153	120.186
Error ( $\mu m$ )	0.0036	0.0124	0.0245	0.0651

When imaging an axial object point, spherical aberration is the only aberration which affects the imaged point. Rays which pass through the lens further from its centre are refracted slightly more than rays which pass through at the centre. This is caused by the spherical shape of the lens (further examined in Chapter 5). The most significant result of this experiment is the assertion that a conventional pupil can be sub-sampled at any point using a pupil of smaller size. Therefore, movement of the sub-pupil is geometrically equivalent to sampling rays from the conventional pupil.

The second experiment conducted examined the effect of increasing the sub-sampled pupil diameter. The results outlined in Figure 2.13 were acquired with a 1mm diameter sub-sampled pupil. Thus, an examination range between 1mm and 5mm was chosen for the sub-pupil diameter within two conventional pupil diameters of 10mm and 15mm. Formation of the pupil plane for this experiment is represented by the blue pupils in Figure 2.11. Results are shown in Table. 2.2 while the error is presented in Figure 2.14(a) for a conventional diameter of 10mm and in Figure 2.14(b) for 15mm. In both cases, it can be seen that as the sub-pupil diameter is increased, the error in sub-sampling the conventional pupil also increases. Similarly to the initial experiment on sub-sampling pupil location within the pupil plane, spherical aberration increases the error. Initially, the error is increased with the conventional pupil diameter

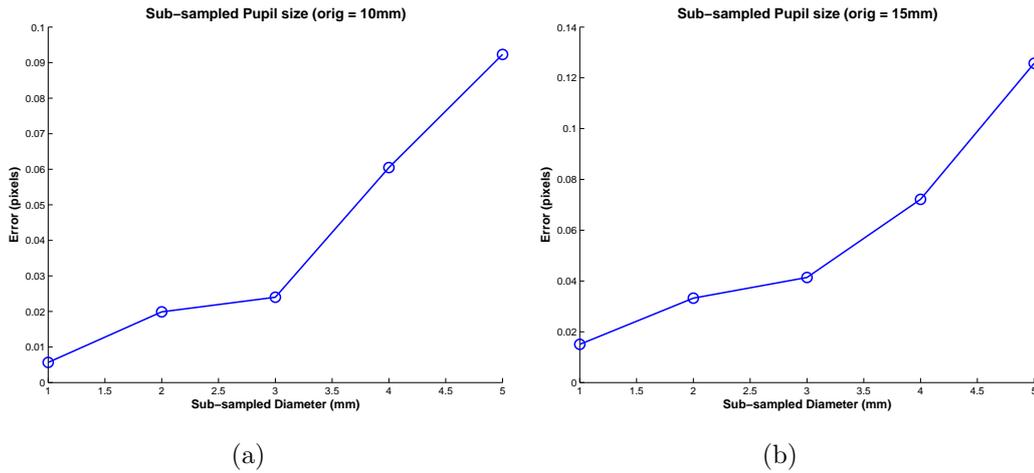


Figure 2.14: Error in sub-sampled pupil due to diameter variation (1mm-5mm); (a) Sub-sampling a pupil diameter of 10mm; (b) Sub-sampling a pupil diameter of 15mm.

Table 2.2: Sub-sampled diameter variation.

	Conventional Diameter ( $\mu\text{m}$ )	Sub-sampled Pupil Diameter ( $\mu\text{m}$ )				
		1mm	2mm	3mm	4mm	5mm
<b>10mm</b>	82.13	82.15	82.04	82.03	81.87	81.73
<b>15mm</b>	120.12	120.19	120.27	120.30	120.43	120.67

increase (from 10mm to 15mm). Subsequently, the increase in sub-sampled pupil diameter (from 1mm to 5mm) contributes to a larger magnitude of error in the final centroid estimate. This is due to the slight skew of the blur circle introduced by the spherical aberration present within the system. The scale of the induced error is approximately 1/8 pixel for a 5mm diameter sub-sampled pupil at 15mm and 1/11 pixel for the same sub-sample diameter at 10mm. Although the error remains relatively small in magnitude, it is desirable to have the minimum error achievable for the sub-sampled ray to better represent the effect of the full pupil. Therefore a sub-sample pupil diameter of 1mm is best suited for this task.

A closer examination of the physical size of the sub-sampling pupil leads to an investigation of the effect of diffraction within the imaging system. Diffraction occurs within an imaging system due to the wave nature of light passing through it. Interference between light waves occurs when it passes through the pupil. Naturally, the diffraction increases as the pupil size is decreased. The pattern observed on the image plane is a combination of constructive and

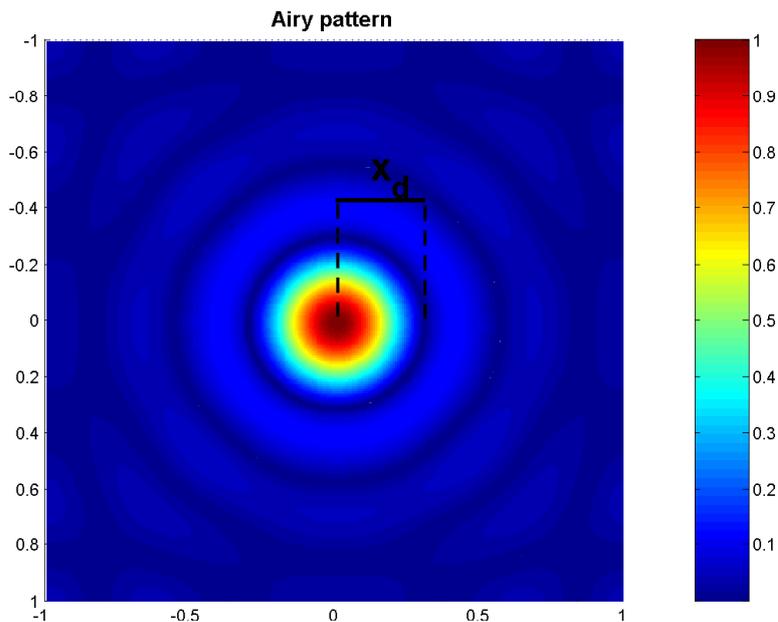


Figure 2.15: Airy pattern.

destructive interference between the waves of light which passed through the pupil. This pattern takes the form of a centrally illuminated circle which is enclosed by light and dark rings. The rings correspond to the interference between the propagated light through the pupil. This pattern is known as an “Airy” pattern (presented in Figure 2.15) in which the distance from the pattern centre to the first dark ring,  $x_d$ , is approximated by,

$$x_d \approx \frac{1.22\lambda f}{d} \quad (2.27)$$

where  $\lambda$  is the wavelength of the incident light,  $f$  is the sensor depth, and  $d$  is the pupil diameter. Therefore, an approximation can be made for the increase in diameter of the imaged point due to diffraction. Within the constraints of the Zemax experiments which were conducted, a single wavelength at  $465nm$  was used with the sensor depth at  $75mm$ . Given a sub-sampled pupil diameter of  $1mm$ , the diffractive contribution to the imaged point will be approximately  $42.5\mu m$  which equates to 9.8 pixels. However, since it is the centroid of the imaged circle which is being estimated, and since the pattern is symmetric, the diffractive effects can be ignored for the purpose of sub-sampling a conventional pupil.

## 2.4 Discussion

This chapter is primarily concerned with modelling the imaging process. Three camera models, which are commonly used in the field of computer vision, are examined in detail and the calibration of these models is outlined. Section 2.1 introduces the traditional pinhole camera model and details the parameters which describe its projection properties. One of the main limitations of the pinhole model is that it does not capture scene radiance. Therefore, lens models are introduced which capture more light rays emanating from the object point and focus these onto the image sensor. One of the key goals of this thesis is to capture depth information from a single image using a single image sensor. Therefore additional cues are required to regain the depth information which is lost with the pinhole model. Depth estimation techniques incorporating pinhole models are briefly discussed but typically require the use of multiple sensors. Consequently, a closer examination of lens models is conducted. The simplest lens model, the thin lens model, places an ideal lens at a distance  $v$  from the the sensor where an aperture or pupil is located at the centre of the lens with a diameter  $D$ . The diameter of the pupil determines the quantity of light rays captured by the imaging system. With knowledge of the lens focal length and sensor depth, Gaussian optical laws facilitate the determination of object depth. The true nature of imaging lenses are neglected by the thin lens model since most photographic lenses contain multiple lens elements and stops. Therefore, the Gaussian thick lens model is introduced. The cardinal points of the thick lens define how object rays are refracted through a system of lens elements. With knowledge of the cardinal points, the lens law can be used to calculate object depth. By incorporating a lens into the imaging system, only a single object depth can be focused to a single point on the image sensor. This object depth is known as the focus depth. A point at any object depth which does not lie on the plane of the focus depth will be captured on the image plane “out of focus” as a blur circle. The quantity of blur captured on the image plane is related to the pupil diameter and, more importantly, the object depth. Thus, the image blur is in fact a depth cue captured by the camera models incorporating lenses.

Calculating object depth requires knowledge of certain camera parameters when using the lens law. Generally, information on the lens focal length is provided by the manufacturers. Therefore, it is the sensor depth which needs

to be recovered to calibrate the optical configuration. The equivalence of the sensor depth of the lens models and the “focal length” parameter of the pinhole model was established. Section 2.2 presents the standard method for recovering camera parameters based on a pinhole model. Planar camera calibration requires three images of a planar target in varying orientation. Subsequent fitting of the IAC based on the object point to image homography is performed to recover the camera calibration matrix. The sensor depth parameter is recovered from this matrix. An alternative method for planar camera calibration is explored in the form of Gurdjos and Payrissat (2001). The standard method of Zhang (1998) is based purely on fitting an imaginary entity in the form of the IAC, and involves minimising an algebraic error to calculate the calibration matrix. Gurdjos and Payrissat (2001) propose a more intuitive geometric framework which, although based on the relationship of the IAC, involves real entities on the image plane and thus minimises geometric error in the determination of the calibration matrix.

Section 2.3 examines the impact of particular camera models on recovering object depth. It was established that lens models introduced blur to the imaging process which could be harnessed as a cue for depth recovery. Since the pupil diameter and magnitude of image blur are proportional, this relationship is further examined. In particular, the impact of the pupil plane location and location of the pupil within the pupil plane are outlined. Conventional imaging lenses consist of multiple lenses and stops. The limiting stop within the system is known as the pupil within the pupil plane. Location of this plane within the lens system plays an important role in the formation of the image blur for an object depth which is out of focus. Subsequently, the concept of optical irises is introduced. The entrance iris and exit iris are defined as the images of the pupil from object space and image space respectively. Therefore optical irises define the axial cones of rays which enter and exit the optical system. When using image blur as a cue for depth, its relationship with the pupil plane must be known. However, it is the entrance iris which defines the object rays entering the system, and the eventual magnitude of the image blur on the image sensor. Consequently, the location of the pupil plane and the diameter of its pupil must be known. It is shown through experimentation that when the pupil plane is physically located outside the system of lens elements in object space, it becomes equivalent to the entrance iris. However, when the pupil plane is located in image space, it is the exit iris and the entrance

iris becomes its image in object space. Experimentation highlights the linear relationship between pupil diameter and image blur for a system in which the pupil plane is the entrance iris. Similarly, an experiment is conducted to verify that this relationship becomes non-linear when the pupil plane is located in image space.

The geometric equivalence of the camera models presented in this chapter is outlined. It is shown that a thick lens can be modeled as a thin lens which is centred at the first principal plane of the thick model. In turn, the thin lens can be decomposed to a pinhole by replacing the lens with a pinhole. With the additional knowledge of the pupil plane, properties of the image blur formed on the image plane can be exploited. Thus, subsequent experimentation examines the impact of remotely locating the pupil plane in object space and of sub-sampling the pupil of the pupil plane. The experiments verify that a pupil plane may be remotely located in object space by modelling it with an equivalent pupil diameter at the imaging lens. Further experimentation examines how the pupil plane can be sub-sampled with a pupil of a smaller diameter. Results show that sub-sampled pupils form an image blur circle on the image plane which, when its centroid is found, accurately models the ray which is passing through an equivalent (conventional) pupil at that diameter. Further testing was carried out which examined the accuracy of this sub-sampling based on increasing the diameter of the sub-sampled pupil. The smallest pupil diameter (1mm) was found to yield the most accurate sampling of the conventional aperture. An investigation into the diffractive properties of the imaging system was subsequently performed. Smaller diameter pupils will produce more diffraction on the image plane. However, since it is the centroid of the image circle that is calculated, the diffraction will not effect the sub-sampling results. Additionally, the object depths which are examined lie in areas which are outside the object focus distance. Therefore, there is considerable image blur present on the image plane which, in the majority of cases, is larger in magnitude than the resulting diffraction of the sub-pupils.

Planar camera calibration is further investigated in Chapter 3. In particular, the method in which calibration images are acquired and the effect this has on camera parameter estimates are examined in detail. Properties of the pupil plane outlined in this chapter are the foundations for the multi-pupil imaging model in this thesis. The results presented in this chapter lead to the formation of two multi-pupil imaging models which are presented in Chapter 4.

## Chapter 3

# Efficient Camera Calibration

The calibration of cameras is an area which has received much attention in the computer vision community over the last 25 years. The drive towards increased resolution with smaller sensor sizes has seen a continued growth in the digital camera market and contributed to the need for flexibility when calibrating a camera. In recent years, the mobile phone camera market has become the largest market for digital sensor suppliers<sup>1</sup>. The emergence of the smartphone has led to increased onboard processing power and high resolution displays, which, coupled with a high resolution camera, is allowing these devices to become more accessible to computer vision researchers. Camera calibration is emerging as a key factor for computer vision researchers in this area for a number of reasons. The lenses used in camera modules manufactured for mobile phones are of a lesser quality (and cost) than a conventional point and shoot digital camera. As a result, lens distortion and aberrations are increased in the images. The increase in demand for mobile phone applications that make use of augmented reality and metric depth information have also contributed to the renewed interest in camera calibration as it is a fundamental step to solving these problems.

Planar calibration targets, or calibration grids, are the de facto standard method of calibrating cameras for computer vision tasks. They are easily manufactured, for example by printing the grid pattern on a desktop printer and then mounting the pattern on a planar surface. The planar calibration meth-

---

<sup>1</sup><http://www.strategyanalytics.com/default.aspx?mod=reportabstractviewer&a0=6216>  
accessed: September 2011

ods of Zhang (1998) and Gurdjos and Payrissat (2001) introduced in Chapter 2 both employ checkerboard planar patterns for the calibration procedure. In this thesis, the input images which form the image data set for planar camera calibration are called the Image Network (IN). The implications for calibration of the planar target orientation in world space with respect to the camera is an area of planar camera calibration which has received little attention. With the exception of the work by Sturm and Maybank (1999), in which degenerate configurations are outlined, there is no indication of what constitutes a good image network geometry for the purpose of planar camera calibration. Thus, the primary focus of this chapter is to address this issue by specifying optimal planar pose for calibration targets in INs.

Section 3.1 examines, in detail, the geometric relationship between the planar calibration methods of Zhang (1998) and Gurdjos and Payrissat (2001). Although both methods solve the camera parameters based on the Image of the Absolute Conic (IAC), the geometric domain in which Gurdjos' method is presented gives further insight into the configuration of the planar target and constraints generated on the IAC. The Centre Line (CL), which was derived in Chapter 2, encodes the geometric configuration of the planar target orientation with respect to the camera. The CL is shown to intersect the IAC at its centre, while also being orthogonal to the vanishing line of the plane. Simulated experimentation is conducted to examine the impact of planar target orientations on the calibration results. Planar targets which are degenerate and near degenerate are outlined. Target orientations which conform to degenerate configurations lead to singularities in solving the planar calibration equation, and thus, result in poor parameter estimates. It is shown that consideration of the image CL orientation can avoid such degenerate configurations. Additionally, near degenerate configurations, which also result in poor camera parameter estimates, can be identified and thus avoided when calibrating a camera.

The CL offers an attractive method to avoid singularities as it is calculated from the image to plane homography. Therefore, an IN strategy is proposed based on enforcing ideal geometry in the IN. By enforcing geometric independence in the planar target orientations, independence in the circular points on the IAC is achieved. This enables more accuracy in the estimation of the IAC and subsequent camera parameters. The proposed IN strategy, Optimal Image Network (OIN), is presented in Section 3.2. Optimal Image Networks (OIN) are formed based on the properties of the CLs within the IN. Two forms

of OIN are proposed in Generated Image Networks (GIN) and Selected Image Networks (SIN). In the case of SINS, it is assumed that the user has acquired a large data set of planar calibration images. Subsequently, image filtering, in the form of an image search, is performed to select images with maximum geometric independence to form the IN. However, in the case of GINs, the input requirements are that of a single image of the planar calibration target. Based on the geometric properties of this image CL, synthetic images are generated which correspond to planar target poses which enforce maximum independence in the IN when formed with the initial image. Section 3.2 outlines the necessary steps in forming the synthetic planar poses, and subsequently replicating these images.

Experiments conducted with OINs are presented in Section 3.3. An application of SINS is outlined by examining a data set captured by a webcam through a video sequence. Calibration results using SINS are compared with these for a random selection of images. Further accuracy assessment of camera parameters is quantified by performing a distortion correction with both image network strategies. Experimentation with GINs is undertaken with simulated and real data. The increase in camera calibration accuracy employing GINs is demonstrated with simulated data across a large range of INs. Additional experiments are performed to validate the image replication process. Efficiency of GINs is highlighted by examining the stability of the camera parameter estimates with real data. The implications of OINs in terms of the increased accuracy and stability in camera parameter estimates are clearly shown and discussed.

### 3.1 Camera Configurations

The planar calibration approach of Gurdjos and Payrissat (2001) was presented in Chapter 2. It outlined a more intuitive geometric approach to solving the problem in comparison to the traditional planar methods. However the exact relationship between both methods was not discussed. Figure 3.1 depicts the Poncelet configuration of Gurdjos' method in traditional planar calibration terms i.e. in the form of the IAC. The assumed camera has zero skew, thus two images of a planar target are the minimum requirement to calibrate the camera. Examining Figure 3.1 from the point of view of the traditional method

(Zhang, 1998), the VL of each planar target  $VL_i$  will intersect the IAC ( $\omega$ ) in the two circular points  $c_{i1}$  and  $c_{i2}$ . Thus, there are four constraints on the IAC to recover the four camera parameters  $(f, \alpha f, p_x, p_y)$ .

Analysing Figure 3.1 from a Poncelet perspective is best served by initially examining the configuration for a single planar target. Poncelet's theorem outlined in Figure 2.5 showed that the image of a planar target remained in perspective correspondence with the planar target rotating about its intersection with the image plane. Each rotated plane has an associated camera centre and Principal Point (PP), however only one plane corresponds to the true camera centre and PP. The projection of the camera centre locus onto the image plane results in the CL, which is a real line on the image plane and consists of all plausible PPs. Taking a single plane, for example the plane which corresponds to  $VL_1$  in Figure 3.1, according to Poncelet there are an infinite number of plausible camera centres and resulting PPs in this geometric configuration. It was shown in Chapter two that the CL is perpendicular to the VL of the planar target. Thus  $CL_1$  is shown in Figure 3.1 and it intersects the IAC at its centre. The implication of Poncelet's geometric configuration is that the circular points  $(c_{11}, c_{12})$  also intersect the infinite number of plausible IACs which correspond to different camera centres and PPs. These IACs are shown in Figure 3.1 as the red dashed conics. Since the CL represents all plausible PPs in the image frame, it also intersects the plausible IACs at their centres. Subsequently, when a second planar target is taken into consideration ( $VL_2, CL_2$  and the green IACs) to complete the calibration, it is clear that the CL of this planar configuration will only intersect  $CL_1$  in a single point. That single point is the true PP of the camera and thus corresponds to the correct IAC.

It is clear from Figures 2.5 and 3.1 that the CL encodes geometric information about the camera and planar target configuration. The focus of Gurdjos and Payrissat (2001) was to decouple the camera intrinsic parameters for the purpose of calibrating a camera using images at varying focal lengths. Thus the CL was employed to initially calibrate the aspect ratio and principal point. Consequently, the broader geometric significance of the CL within the planar calibration framework was neglected. The work carried out in this chapter focuses on using properties of the CL to increase the accuracy and stability of planar camera calibration.

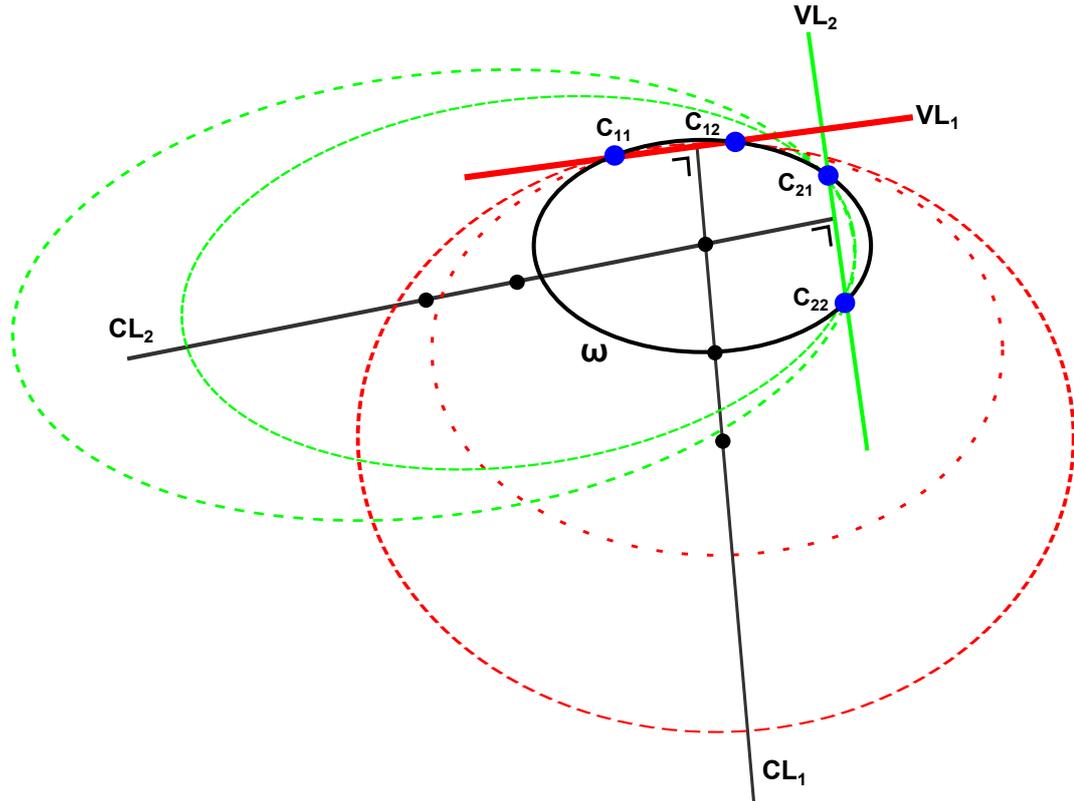


Figure 3.1: Poncelet geometric interpretation in planar calibration domain.

### 3.1.1 Degenerate Configurations

A degenerate configuration results in a singularity arising in the planar calibration equations. Geometrically, this is defined as a planar target which is in an orientation such that it does not contribute independently to the current system of planar calibration equations. Awareness of degenerate configurations is of utmost importance when calibrating a camera, particularly in the minimum case of using two planar targets. If both images do not contribute independently i.e. define a unique IAC, recovery of the camera parameters is not possible.

Degenerate configurations of planar camera calibration have been studied and identified in the computer vision community. The work of Sturm and Maybank (1999) gives the most comprehensive analysis of existing singularities. However, the impact of configurations which are close to degenerate on parameter estimation accuracy has largely been neglected. Additionally, there has been no straightforward method reported to avoid degenerate and near-degenerate configurations when calibrating a camera. This issue is addressed in this chapter by exposing the nature of the relationship between the geometry of the

calibration images and their CLs.

The geometric significance of the CL has been demonstrated. It encapsulates the geometric configuration of the planar target with respect to the camera. Therefore it can be used as a guide to identifying favourable geometry when considering planar target orientations for planar camera calibration. Conversely, it can aid in the detection of degenerate configurations. Extensive simulated experiments were conducted in which the minimum planar calibration requirements were examined for a range of planar target orientations. A zero skew camera was assumed, and a dense set of planar target orientations was tested. Figure 3.2 presents the results for a range of two plane calibrations. An initial planar target orientation was taken with a pitch angle of 20 degrees, a yaw angle of 20 degrees, and a roll angle of 7.5 degrees. Planar calibrations were performed using this orientation as a seed input image. The seed image was calibrated with every combination (every 1 degree change) of planar target with orientations of pitch ranging between -50 and 50 degrees, yaw of -50 to 50, and a random roll angle between -7.5 and 7.5. These orientations approximately represent the limits at which planar orientations are captured for planar camera calibration due to feature extraction constraints.

In order to examine the stability of the calibration results, the condition number of the matrix containing the system of calibration equations (given the two input images) was calculated for each image pair. It is represented by the colourbar in Figure 3.2. The condition number is a quantity which represents how well a system of equations is solved. It is calculated by taking the largest singular value of the equation matrix and dividing this by the smallest singular value. Large condition numbers indicate dependence in the system of equations. Given that the seed image had a (pitch, yaw) orientation of (20, 20), it is evident that an image with identical pitch and yaw orientation is a degenerate configuration. This is represented by the spot area in Figure 3.2 at (20, 20). Variation of the roll angle equates to an in plane variation of the planar target orientation. Thus, roll angle variation does not effect degenerate configurations. On closer examination of Figure 3.2, the degenerate configurations outlined by Sturm and Maybank (1999) can be identified. The planar target orientations at (0, 0), (20, -20), and (-20, 20) are such configurations. Due to the symmetric nature of the IAC (as a result of the zero skew assumption), there are a number of degenerate configurations which arise as a result of dependence between the circular points in certain orientations. Sturm

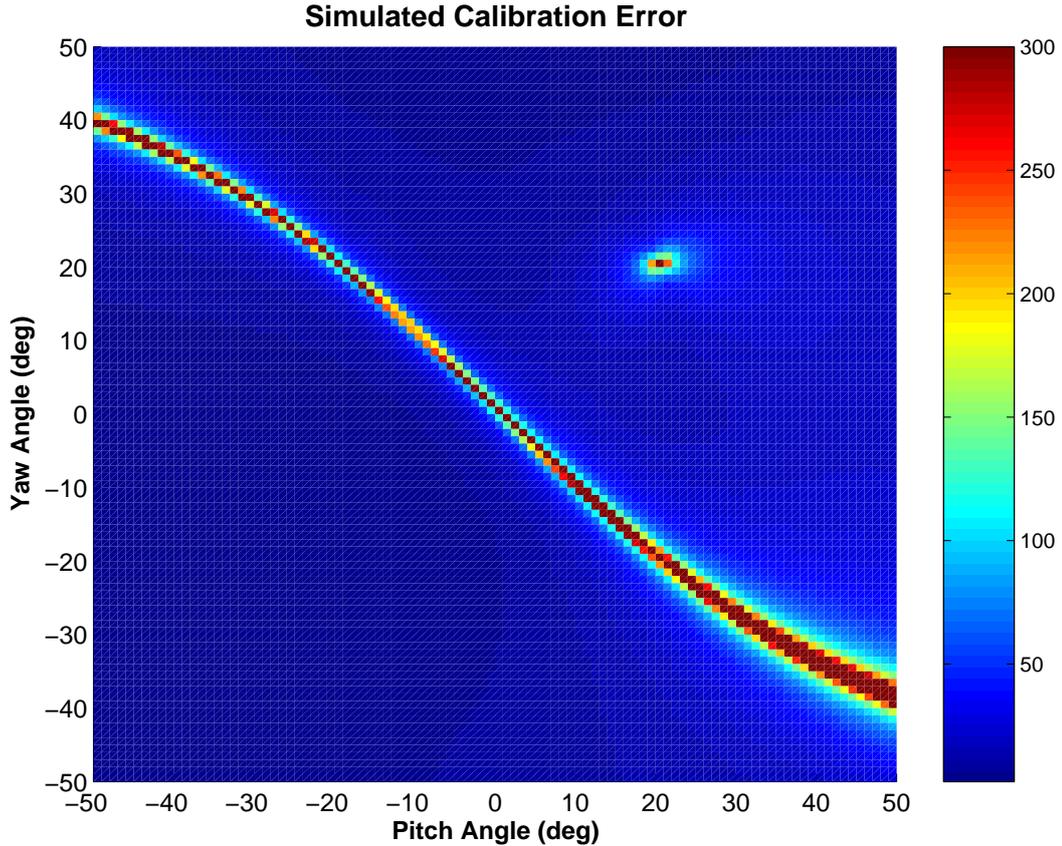


Figure 3.2: Degenerate configurations - simulations.

and Maybank (1999) called this the “reflection constraint”. However this only offered a brief explanation to the root cause of these degeneracies.

The reflection constraint arises when the vanishing lines of both planar targets are reflections of each other by both a horizontal and vertical line in the image. This configuration is outlined in Figure 3.3. The vanishing lines  $VL_1$  and  $VL_2$  intersect  $\omega$  at the circular points  $(c_{11}, c_{12})$  and  $(c_{21}, c_{22})$ . The IAC is symmetric about the axes which intersect at the centre of the conic. It is evident that  $VL_1$  and  $VL_2$  are reflections of each other about the same horizontal axis and a vertical line drawn in at the intersection of both VLS. Therefore, a dependency arises in the circular points, and thus it is not possible to fit a unique conic and solve the calibration equations. One of the major contributions of this chapter is to link this particular family of singularities to the CL. Since the CL is orthogonal to the VLS, it also captures the geometric dependence of the circular points on the IAC. Thus, equal angle with a horizontal and vertical line in the image identifies degenerate configurations equivalent to the “reflection constraint”. This is an important result as the simple formation of the CL

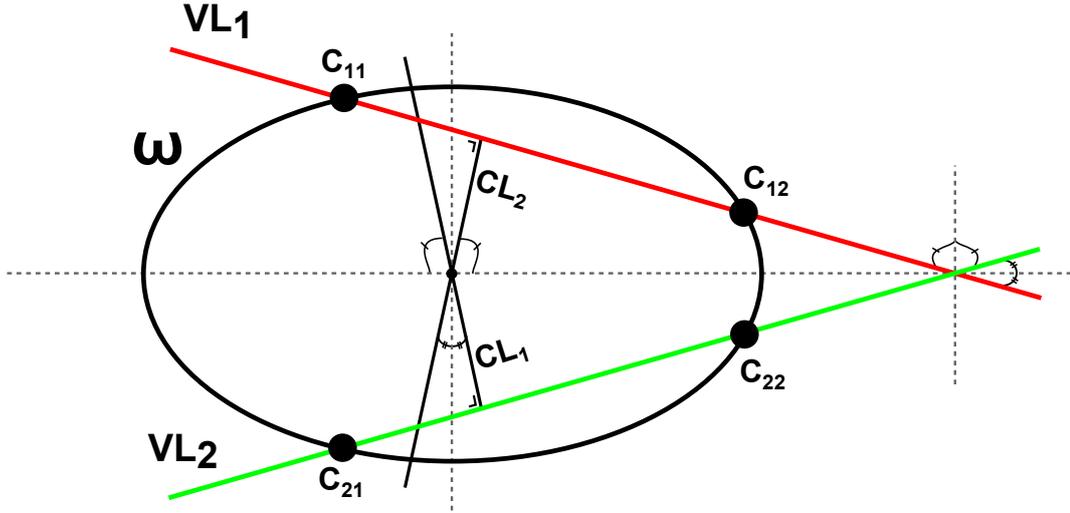


Figure 3.3: Degenerate configurations - reflection.

in comparison to the VLs makes it an attractive proposition for identifying degenerate configurations. Since the CL always intersects the centre of the IAC, it is sufficient to calculate its orientation with either the horizontal or vertical direction to confirm its reflective properties.

Singularities which correspond to the reflection constraint are present in Figure 3.2 in areas between the already identified degeneracies. Thus, there is a manifold of degenerate configurations for the initial seed image at (20, 20) which corresponds to the “hot areas” in Figure 3.2. This manifold also includes near degenerate configurations, (the light blue/green areas) which are not categorised in the literature but generally result in poor parameter estimates. Examination of the image CLs for all the simulated data in Figure 3.2 confirms the link between the CL and image configurations. The seed image has a CL with orientation  $-53.84^\circ$  with respect to the horizontal direction. The total number of simulated second image orientations was 10,201 of which 134 resulted in degenerate (red areas of Figure 3.2) and 201 in near degenerate configurations (light blue/green). In addition to the 335 critical configurations, a further ten images correspond to orientations which are close to the actual seed image. These are the bright areas in Figure 3.2 at (20,20). The average CL angle recorded for these configurations was  $-54.37^\circ$  with a Standard Deviation (SD)  $1.45^\circ$ . Whilst in the case of the 335 degenerate and near degenerate configurations, the resulting CL angle was  $54.13^\circ$  with SD  $2.43^\circ$ . Therefore all degenerate and near degenerate configurations are characterised by examining the relationship between image CLs. Orientations of CL which are similar,

reflected, and near both cases of similar and reflected are either degenerate or near degenerate configurations.

The results presented in Figure 3.2 represent a single seed image orientation. The same simulation was completed for seed images with pitch and yaw orientations across the entire range outlined in Figure 3.2. The same constraints on the CL were exhibited for the degenerate manifold in all cases. If additional input images are used in the planar calibration framework, the number of constraints on the IAC is increased (two for each image). Cases of three or more input images will generally produce a solution for the IAC. However if there is dependency between input images, less stable parameter estimates are found. Therefore image CLs are an ideal tool for examining the geometric configuration of input calibration images regardless of the number of input images. The focus of the remainder of this chapter is the implementation of calibration image networks which contain ideal image geometry for the purpose of accurate and efficient calibration parameter estimates.

## 3.2 Optimal Image Networks

The minimum number of images required in an IN for planar camera calibration is two. Properties of the CL have been introduced and shown to encode geometric information of the planar target configurations. Particularly in the cases of degenerate configurations, the CL can be used as a guide to determine whether or not images within the IN will result in singularities in the parameter estimation stage. Additionally, INs close to degenerate which also result in poor camera parameter estimates can be identified via the CL. Consequently, an optimal image network is defined as an IN which contains ideal image geometry (through optimally independent calibration equations) for the purpose of planar camera calibration.

Two forms of optimal image networks are introduced in this section. Both methods make use of the image CL to determine the geometric configuration of the INs. The first method assumes that a large number of input calibration images have been supplied. Subsequently the CL is used as a guide to filter the ideal images from the degenerate cases. This form of optimal image network is termed Selected Image Network (SIN). The second proposed method of optimal

image networks requires an initial input image from the camera. Based on the CL of this initial image, additional images are generated synthetically. These synthetic images contain ideal IN geometry and are subsequently acquired by the user to form a GIN. SINs and GINs evaluate image geometry based on the orientation of the CL. Independence in the IN is achieved by ensuring that the CLs of images are not in degenerate configurations.

Both methods require an initial seed image, therefore it is imperative that the orientation of the seed image is not in a degenerate configuration or possible degenerate configuration when additional images are added to the IN. CLs with orientations of  $0^\circ$  and  $90^\circ$  correspond to fronto parallel planar target orientations (pitch and yaw of orientation of  $(0^\circ, 0^\circ)$ ) and a rotation of the target about either the pitch or yaw axis respectively. These are known to be degenerate and must be disregarded as seed images for optimal image networks. Depending on the number of images in the IN, which is defined by the user for both SINs and GINs, distinctiveness between CLs must be enforced to ensure a unique determination of the IAC. This is achieved by choosing an adequate angle between image CLs based on the IN size, whilst at the same time ensuring that the reflection constraint is avoided. An example of such avoidance is ensuring the angle between two CLs in a two IN is  $90^\circ$  and that the seed image is not of the orientations  $0^\circ$  and  $90^\circ$  (degenerate) or  $45^\circ$  (reflective degenerate).

Since both forms of optimal image networks contain favourable geometry, an increase in parameter estimation accuracy and stability is achieved. Guidelines on planar target poses for camera calibration have not been adequately addressed in the literature. Therefore both SINs and GINs have implications on the overall scheme of planar camera calibration, particularly in the case of non-expert practitioners.

### 3.2.1 Generated Image Networks

The first step in forming GINs is to define the IN size and capture an initial image of a planar calibration target in a non-degenerate configuration (as discussed in section 3.2). Subsequently, synthetic planar poses are generated based on the properties of the initial CL in the IN. The key contribution of GINs lies in the generation of synthetic images of planes in poses based on

the image CL. In order to generate synthetic images for the GIN, properties of the camera used to capture the initial image must be approximated. This allows the formation of a synthetic planar homography, which in turn defines the synthetic calibration image to be replicated by the user. Once the synthetic poses have been captured, the GIN is formed. The replication process is performed in realtime using the camera’s “live-view” screen<sup>2</sup>. The synthetically generated image of the planar grid with ideal pose is augmented (in a semi-transparent manner) to the current live view of the camera. Therefore, the task of replication is reduced to visually aligning the current live-view of the planar grid with the synthetically augmented grid. An example of this process is shown in Figure 3.11.

Approximating the camera parameters to generate synthetic images requires the construction of a planar homography matrix which encodes the internal parameters,  $K$ , of the camera and the new pose of the target. Estimating the principal point  $(p_x, p_y)$  at the centre of the image allows an approximate calculation for the camera sensor depth, or in pinhole model terms, the focal length  $f$ . This yields an approximated camera matrix  $\hat{K}$  which reduces the problem to the construction of the synthetic planar target pose  $\hat{R}$ . The synthetic homography  $\hat{H}$  is formed by  $\hat{K}$  and the pose  $\hat{R}$ . The rotation matrix,  $\hat{R}$ , is decomposed into the pitch ( $r_\alpha$ ), yaw ( $r_\beta$ ) and roll ( $r_\gamma$ ) angles, which are used to estimate the new pose of the planar target. Constraints on the CL are imposed, reducing the solution space and ensuring that a valid pose is determined.

### Forming $\hat{H}$

The relationship between the IAC and a point homography was outlined in Chapter 2. Coupled with the knowledge of the principal point and use of  $H$  from the initial image captured by the user,  $\hat{H}$  can be partially formed. For a complete formation, the target pose being generated ( $\hat{R}$ ) must be considered. A general planar homography matrix  $H$  can be decomposed as,

$$H = K[R_{12} | t], \tag{3.1}$$

---

<sup>2</sup>A webcam, Logitech<sup>®</sup> QC3000 is used for the experimentation with GINs (at VGA resolution)

where  $R_{12}$  represents the first two columns of the rotation matrix. Image generation of the planar target is obtained by applying  $\hat{H}$  to an image of the calibration target in its canonical position. A valid  $\hat{H}$  is constructed similarly to Eqn. 3.1. The first step of the formation takes advantage of the assumption of  $p_x$  and  $p_y$  (at the image centre) by eliminating these terms in Eqn. 3.1. This is accomplished by pre-multiplying  $H$  by a matrix  $U$  where  $\hat{H} = UH$  with  $U = [I \mid -c]$  and  $c = (p_x, p_y, 1)^T$ . This leaves an expression containing the focal length,  $f$ , and the scaling factor,  $\mu$ , of the  $H$  matrix,

$$\hat{H} = \begin{bmatrix} \frac{fr_{11}}{\mu} & \frac{fr_{12}}{\mu} & \dots \\ \frac{fr_{21}}{\mu} & \frac{fr_{22}}{\mu} & \dots \\ \frac{r_{31}}{\mu} & \frac{r_{32}}{\mu} & \dots \end{bmatrix} \quad (3.2)$$

The third column, which contains the translation, does not affect the forming of  $f$ . By manipulation of the planar calibration equations (Eqns 2.14 and 2.15), the parameter  $f$  can be estimated from  $\hat{H}$  as,

$$f = \sqrt{\frac{-\hat{h}_{11}\hat{h}_{12} - \hat{h}_{21}\hat{h}_{22}}{\hat{h}_{31}\hat{h}_{32}}} \quad (3.3)$$

which can now be used to find the scaling factor in Eqn. 3.2. Calculating the scaling factor is straightforward, since it is well known that the columns of a rotation matrix form an orthonormal basis (Hartley and Zisserman, 2003).

With the ability to decompose the planar homography matrix from the first image ( $H$ ) into a calibration matrix  $\hat{K}$  and scaling factor  $\mu$ , the planar homography matrix for the synthetically generated image ( $\hat{H}$ ) can be constructed via these parameters coupled with the optimisation parameters  $\hat{R}_\alpha$ ,  $\hat{R}_\beta$  and  $\hat{R}_\gamma$  which form the estimated rotation matrix  $\hat{R}$ .

$$\hat{H} = \mu\hat{K}\hat{R} \quad (3.4)$$

### Constraints on $\hat{H}$

The criteria for generating a valid  $\hat{H}$  have been identified. The assumption of the principal point at the image centre recovers an approximated  $\hat{K}$  which, in turn, recovers the scale from the original image homography  $H$ . Therefore, the key to determining the GIN is the estimation of the pose ( $\hat{R}$ ). Forming a  $\hat{H}$  which yields a correct solution requires constraints to be imposed on  $\hat{H}$

while  $\hat{R}$  is estimated. These constraints are based on geometric properties of the image, primarily the CL.

An expression for the CL was derived in Eqn. 2.24. This can be written in linear form as,

$$p_y = \Gamma p_x + \Lambda \quad (3.5)$$

where  $\Gamma$  is the slope of the CL given by,

$$\Gamma = \frac{-\varphi_1 \alpha^2}{\varphi_2} \quad (3.6)$$

and  $\Lambda$ , the y-intercept given by,

$$\Lambda = \frac{\varphi_3 \alpha^2 + \varphi_4}{\varphi_2} \quad (3.7)$$

The first constraint applied to  $\hat{H}$  is on the slope of its CL. As discussed in Section 3.2, once the GIN size (including seed image),  $k$ , has been chosen, a non-degenerate seed image is defined. Therefore equal angle between image CLs enforces independence in the geometry of the IN. The angle between CLs is defined by its slope and the slope of the seed image,  $m_1$ . Therefore, depending on the number of images to be synthetically generated,  $n$ , the optimum CL slope  $\Gamma_i$  is defined as,

$$\Gamma_{i|i=1:n} = \tan \left( \tan^{-1}(m_1) + \frac{ia\pi}{180} \right) \quad (3.8)$$

where  $a$  is the angle between CLs, found as  $a = 180/k$ . Thus, generation of synthetic images require their CLs to have a slope of  $\Gamma_i$ . As the slope is nonlinear in the parameters of  $\hat{H}$ , a cost function is developed to estimate the optimal slope value. The cost function  $\mathfrak{C}_i(\mathbf{1}, \hat{R})$  for this constraint (where “1” indicates the first constraint) is expressed as,

$$\mathfrak{C}_i(\mathbf{1}, \hat{R}) = (\hat{\Gamma}_i - \Gamma_i)^2 \quad (3.9)$$

where  $\hat{\Gamma}$  represents the slope of the CL for the homography being estimated ( $\hat{H}$ ). The implications of  $\mathfrak{C}_i(\mathbf{1}, \hat{R})$  lead to multiple solutions of equal angle between CLs which are determined by the y-intercept,  $\hat{\Lambda}_i$ . Thus a second constraint is imposed on  $\hat{H}$  with the aim of generating the correct CL. This is enforced by ensuring the generated CL passes through the assumed principal point. The optimal intercept value ( $\Lambda_i$ ) is calculated as,

$$\Lambda_i = p_y - \Gamma_i p_x \quad (3.10)$$

The cost function  $\mathfrak{C}_i(2, \hat{R})$  is formed similarly to  $\mathfrak{C}_i(1, \hat{R})$ .

$$\mathfrak{C}_i(2, \hat{R}) = (\hat{\Lambda}_i - \Lambda_i)^2 \quad (3.11)$$

The minimisation of  $\mathfrak{C}_i(1, \hat{R})$  and  $\mathfrak{C}_i(2, \hat{R})$  yields a manifold of possible solutions. In order to select poses that are realisable, a third constraint is applied on the solution space which examines the extent of compression/expansion in the generated synthetic images. This prevents the generation of GIN images that would be practically unattainable and difficult to perform feature extraction on.

When a transformation  $\hat{H}$ , is applied to an image, it can result in the expansion and compression of pixels in the image. This compression and expansion can be measured locally in the image by examining the singular values of the Jacobian (Mallon and Whelan, 2005). The Jacobian of a single point  $p = (x, y)$  in the image is defined as:

$$\mathfrak{G}_p \Rightarrow \mathfrak{J}(\hat{H}, p) = \begin{bmatrix} \frac{\delta \hat{x}}{\delta x} & \frac{\delta \hat{x}}{\delta y} \\ \frac{\delta \hat{y}}{\delta x} & \frac{\delta \hat{y}}{\delta y} \end{bmatrix} \quad (3.12)$$

Each point,  $p$ , in the image has two corresponding singular values  $\sigma_1(\mathfrak{G})$  and  $\sigma_2(\mathfrak{G})$ . For a transformation  $\hat{H}$ , if  $\sigma > 1$  there is an expansion of pixels and if  $\sigma < 1$ , the overall effect is compression. It is desirable to ensure that the singular values of the Jacobian at each point in the image are as close as possible to 1. This limits perspective distortion and ensures realisable images and feature extraction. The Jacobian can be calculated at each point in the image, or alternatively over a grid of points in the image. This constraint is expressed as,

$$\mathfrak{C}_i(3, \hat{R}) = \sum_{i=1}^l [(\sigma_1(\mathfrak{G}_i) - 1)^2 + (\sigma_2(\mathfrak{G}_i) - 1)^2] \quad (3.13)$$

where  $l$  is the number of points in the grid used. The total cost function ( $\mathfrak{T}$ ) which incorporates all the constraints is,

$$\mathfrak{T}(\hat{R}) = \sum_{i=1}^n [\mathfrak{C}_i(1, \hat{R}) + \mathfrak{C}_i(2, \hat{R}) + \mathfrak{C}_i(3, \hat{R})] \quad (3.14)$$

where  $n$  is the number of images to be generated in the image network. Initial estimates of  $r_\alpha$ ,  $r_\beta$  and  $r_\gamma$  for the optimisation are taken as the inverse orientation of the initial calibration image provided by the user (approximating the orientation by decomposing  $H$ ). The minimisation of  $\mathfrak{T}(\hat{R})$  will generate the

synthetic target poses for the GIN through the formation of the optimal  $\hat{H}$  matrices. The Levenberg-Marquardt (LM) algorithm is used for this purpose. The algorithm generally converges within 20 iterations.

### Network orientation and sensitivity

Given that the rotational parameters are being estimated for the GIN optimisation, the sensitivity of  $\hat{R}$  is examined. Properties of the CL were investigated in Section 3.1 and it was noted that the roll angle of the planar grid  $r_\gamma$  has no effect on the CL orientation as it represents an in-plane rotation. Therefore  $r_\gamma$  does not influence  $\mathfrak{C}_i(\mathbf{1}, \hat{R})$  or  $\mathfrak{C}_i(\mathbf{2}, \hat{R})$ . However, it does influence the image grid constraint  $\mathfrak{C}_i(\mathbf{3}, \hat{R})$ . If the roll angle is large in a perspective distorted image, the feature extraction process can be complicated, particularly with regards to the ordering of the control points. Additionally, the guided image acquisition can become more convoluted due to large roll angles which would, for example, require the user to capture the images of a calibration grid rotated by  $90^\circ$ . Thus, by including the  $r_\gamma$  in the optimisation, it prevents large roll angles and yields images which are more practical for the replication process.

The sensitivity of the  $\hat{R}$  parameters to random perturbations is tested. These perturbations are examined as they represent situations where the practitioner may be unable to reproduce the synthetic pose exactly. They are modelled as random variations ( $r_{err}$ ) in the range of  $\pm 10^\circ$  in  $r_\alpha$  and  $r_\beta$ . Typically the user error is within  $\pm 2^\circ$  of the optimum angle as outlined in the experiments in Section 3.3.2.

The variations of  $\hat{R}$  are used to form the perturbed planar homography matrix  $H_{per}$  which gives a perturbed CL.  $H_{per}$  is formed similarly to  $\hat{H}$  in Section 3.2.1 using Eqn. 3.1.

$$\begin{aligned}\hat{R} &= (r_\alpha + r_{err1}, r_\beta + r_{err2}, r_\gamma) \\ H_{per} &= \hat{K}\hat{R}\end{aligned}\tag{3.15}$$

Simulated testing was performed for 1000 instances of random perturbations in the  $\hat{R}$  of a homography describing a CL with known angle relative to another CL. Figure 3.4 shows the relationship between the perturbed error on  $\hat{R}$  and the corresponding induced angle error between the CLs. It can be seen that for small perturbations of  $\hat{R}$ , the induced angle error between CLs remains

Sensitivity of R to Random Perturbations

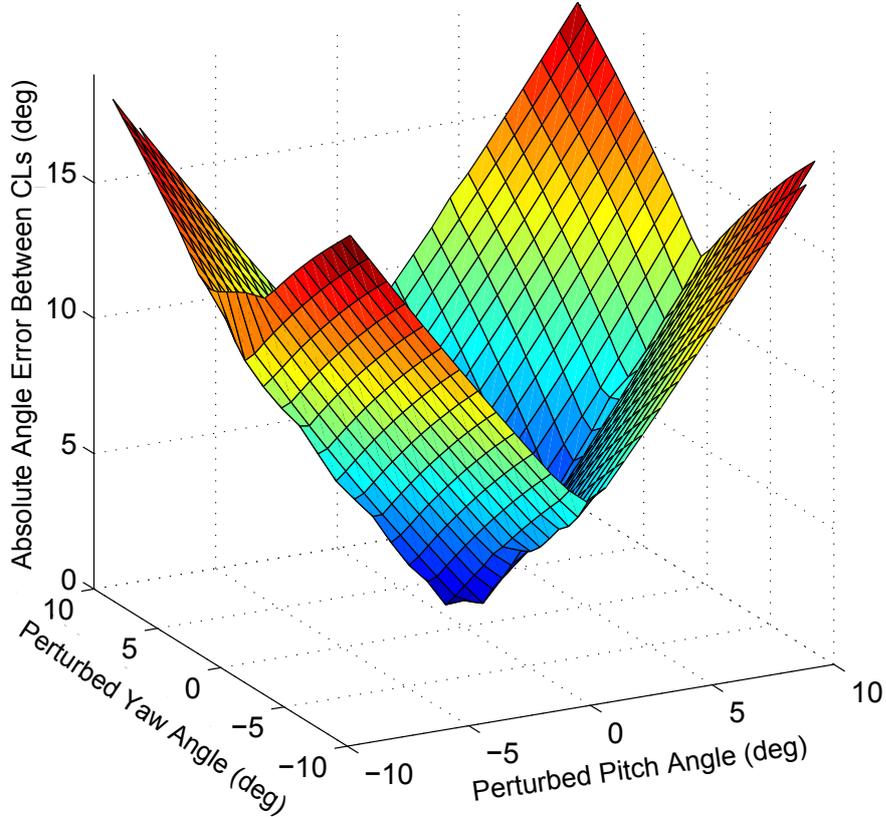


Figure 3.4: Effects of perturbations when estimating the ideal image via the planar homography  $H$  and the rotational parameters  $R$ .

bounded. This is an important result as it ensures that small error in the  $\hat{R}$  estimation will not dramatically affect the calibration results.

### 3.2.2 Selected Image Networks

Selected image networks are the second method employed to obtain INs with favourable geometry for planar camera calibration. SINS are formed given a large data set of input calibration images. The user input requirements are similar to that of the GINs in that an IN size must be specified before the SIN is formed. Therefore, once a  $k$  sized IN is chosen, images from the given data set can be selected to form a SIN based on the orientations of their CLs. Additionally, the same constraints apply to selecting the seed image in terms of degeneracy avoidance. The ideal angle between CLs is calculated in the same manner as the GINs i.e.  $a = 180^\circ/N$ . The only constraint on selecting calibration images is the slope of the CL. It can be assumed that all image

CLs pass through the principal point since the data set provided for the SIN formation is captured with the same camera. Automatic image selection is implemented in two stages. The initial step requires each image CL to be determined, with the angle of each CL relative to the x-axis calculated via the slope (see Eqn. 3.6). The second step involves an algorithmic search through the data set of images to select images with the desired geometry, or CL angles. In practice, a tolerance of  $\pm 1^\circ$  is enforced on  $a$  for the selection process.

The underlying search method of the proposed image selection strategy is a binary search approach (Knuth, 1998). Figure 3.5 presents the proposed strategy. Each node represents an image number while the number adjacent to each node is its CL orientation. In a real situation all nodes are connected to each other where the connecting lines represent the angle between image CLs. To aid explanation all possible connecting lines are not shown, instead the valid search paths are shown i.e each line is in fact equal to the binary search key which is  $|a| \pm 1$

In this example,  $k$ , the number of images required in the SIN, is set to four therefore  $a$  is  $45^\circ$ . The search begins with the seed node 1. When a route corresponding to the search key is identified (node 3) the search continues with node 3 as the seed. In this case there are two possible routes, node 12 and 17. Since a binary search exploits only one route at a time, the proposed strategy implements the search in a cascaded fashion. Therefore the search will return to this point and follow other routes in the next cycle. If node 12 is chosen there are three candidates to complete the SIN: nodes 22, 31, and 64. When one possibility is chosen the SIN is formed and the node is deleted. This will allow the binary search to find the next valid IN with subsequent searches. When all possible SINs along the (1, 3, 12) route have been formed, all image nodes are reintroduced but node 12 is deleted. This enables the algorithm to return to the point (1, 3) and follow the route through node 17. The search continues in this fashion until all valid routes have been identified. In cases of multiple SINs being formed, further refinement can be implemented by increasing the resolution of the search key.

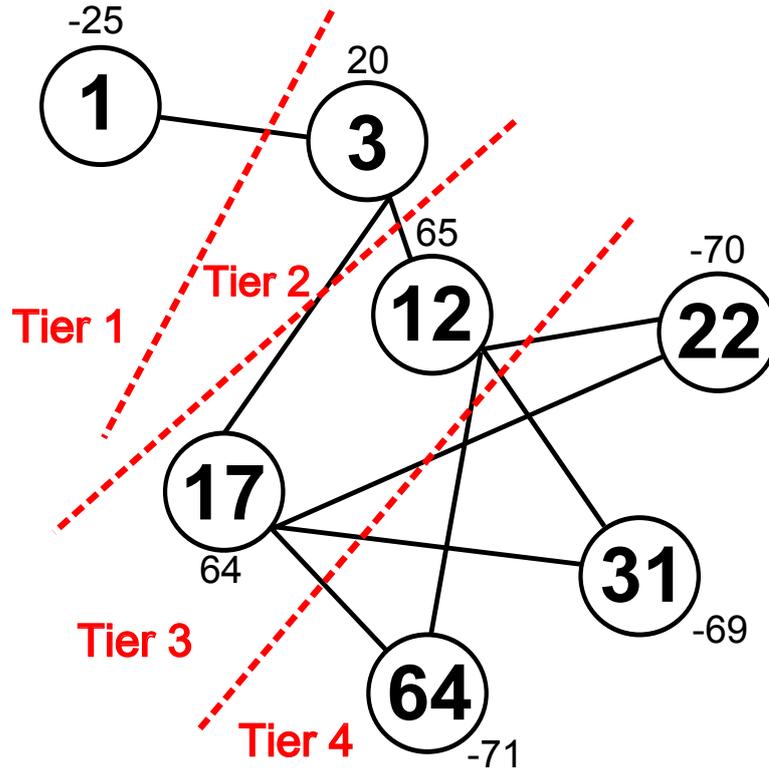


Figure 3.5: SIN image search space.

### Search Algorithm Implementation

The implementation of the search algorithm is presented in Figure 3.6. The algorithm is initialised with the seed node which is stored in the  $IN$  matrix. The binary search is implemented and, if successful, the current  $IN$  size,  $N_T$ , is incremented and the image found,  $imt$ , is stored in  $IN$ . A check is performed to see if  $N_T$  is equal to  $k$  the desired  $IN$  size. If true, the current  $imt$  is deleted from the search space and the algorithm loops back to run the binary search again. This is a similar situation as in Figure 3.5 when tier 4 was reached, the node was deleted so other nodes could be found in subsequent searches. In the case where  $N_T \neq k$ , the algorithm simply loops back to find the next image in the network via the binary search.

The most significant branch point in the algorithm is the binary search junction. When a search fails it indicates that there are no more routes in a particular tier of the search space. A check is performed on  $N_T$  to see if the search has failed in tier 2 (i.e. if  $N_T = 1$ ). If it has failed, the overall search is finished, and all valid  $IN$ s are stored in the  $IN$  matrix. On the other hand, if  $N_T \neq 1$ , all nodes are reintroduced to the search space and the  $N^{th}$  node

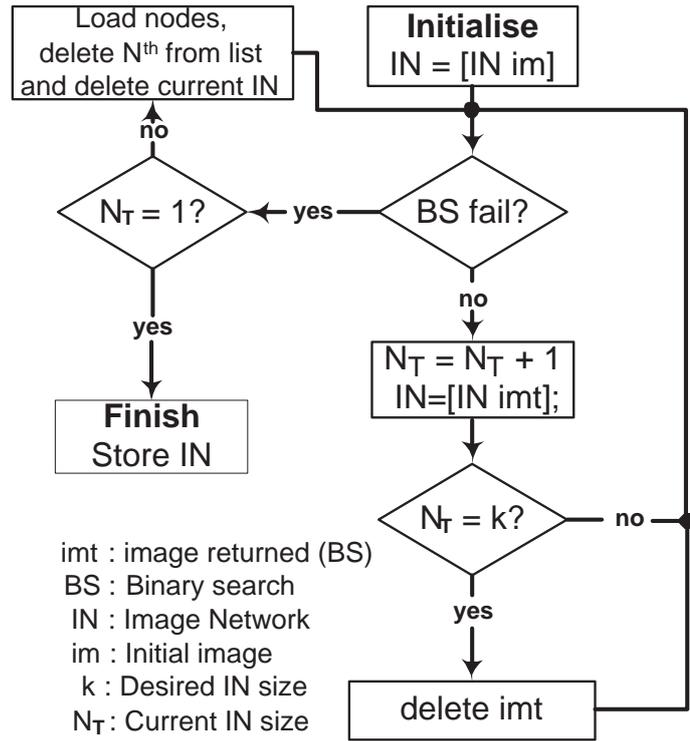


Figure 3.6: SIN search algorithm flowchart.

of the current IN is deleted from the list. This is a similar situation to the example given in Figure 3.5 where the  $IN$  was (1, 3, 12). When all tier 4 nodes were found and subsequently deleted, node 12 was removed which allowed the binary search to follow the (1, 3, 17) route. The search continues until the binary search fails, and  $N_T = 1$  which indicates that all possible routes have been explored. The algorithm is implemented in this fashion for each node in the search space. Further refinement can be applied to the algorithm if duplicate INs are stored.

### 3.3 Experiments

Experimentation is performed to evaluate both approaches to obtaining INs with ideal geometry for planar camera calibration. In all cases, the image network consensus employed for comparison with GINs and SINs is that of a random selection. In the case of simulated data (used in some experiments for GINs), planar target orientations are randomly drawn from an interval of -50 to 50 degrees on pitch and yaw angles, whilst the roll angle was randomly

drawn from a range between -7.5 and 7.5 degrees. The random consensus, Random Image Network is termed RIN. The planar calibration target used for the real tests is a standard checkerboard pattern. The planar calibration method of Zhang (1998) is employed to calculate the camera parameters.

### 3.3.1 Selected Image Networks

The data set used for SIN experimentation consisted of 300 frames captured by a webcam of a planar target in varying orientations. Three experiments are conducted to illustrate the calibration accuracy and overall benefit of using SINS. The first experiment examines the accuracy of the calibration result as the number of views (images) increases. This is followed by a rigorous comparison of the IN configurations with fewer images. The final experiment validates the calibration accuracy of SINS by analysing the lens distortion in the images. As there is no ground truth a calibration result is calculated using all 300 frames and subsequently taken as an indicative optimal result for comparison.

#### Convergence of Image Networks

This experiment examines the accuracy of the calibration result as the number of images in the IN increases. In order to quantify the accuracy, a ground truth calibration result is calculated using all 300 images within the data set. The ground truth intrinsics are summed and subsequently used as an indicative optimal calibration result. Therefore, results are presented as a percentage error of the optimal calibration result. For each image network instance (2 to 25 INs) ten SINS are compared with ten RINs from the data set. It is clear from Figure 3.7 that the SINS produce more accurate and reliable calibration results than the RINs, particularly in the case of smaller INs. As the number of views increases the stability of the SIN results still outperform the RINs, however, once this increases beyond 7 images, comparable results are observed. This is due to the number of constraints on the IAC. Each image provides 2 constraints on fitting the IAC (in the form of the circular points). Assuming zero skew, four parameters are being estimated, and with 8 images there are 16 constraints on the IAC. Therefore, in both cases of RINs and SINS, the systems are over-constrained, and therefore, the difference in performance is

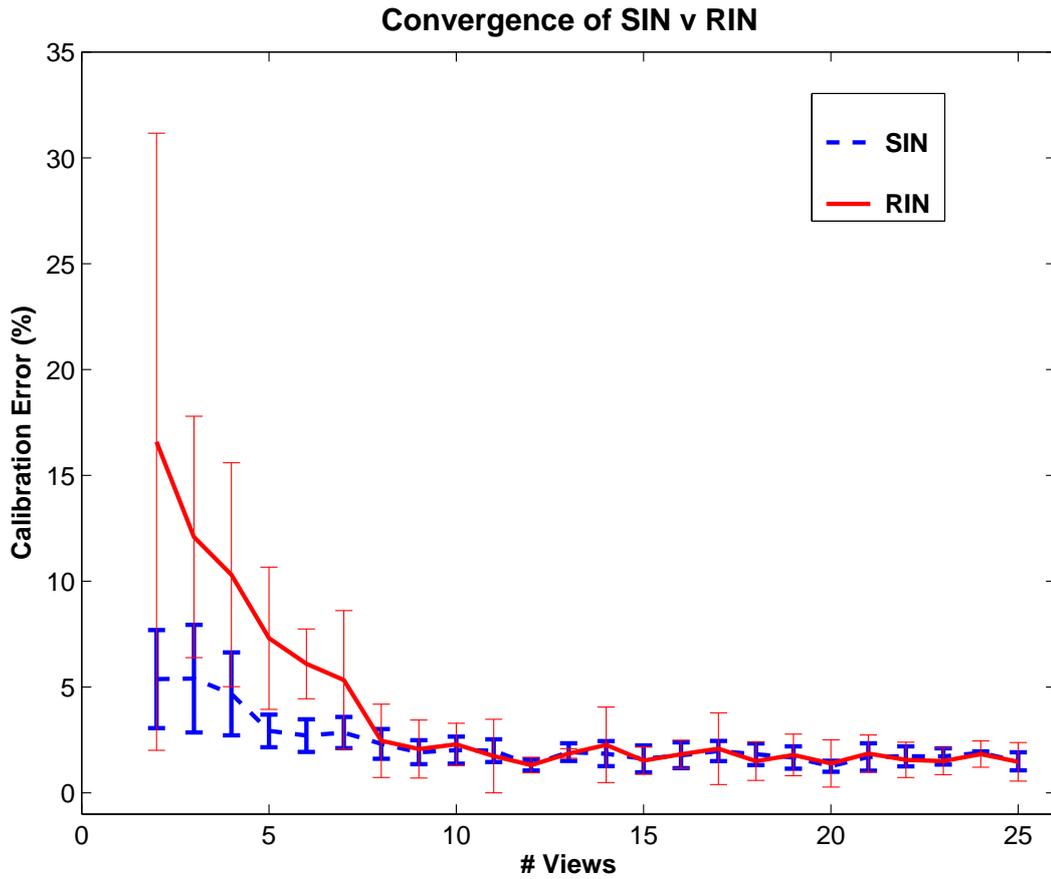


Figure 3.7: Calibration accuracy as the number of views increases from 2 to 25 INs. Standard Deviation (SD) is represented by bars at each IN instance.

negligible. The most important result is that in cases of INs with few images, the SINs significantly outperform the RINs.

### Reduced Number of Images

The number of images for each IN configuration is reduced for this set of experiments to validate the previous result which identified that SINs can achieve more accurate results than the RINs with fewer images in the IN. All results in Table 3.1 are given in terms of percentage error with respect to the indicative optimal calibration result for each parameter. There are 50 two INs, 30 three INs, and 25 four INs in the SIN cases of Table 3.1. RINs were randomly drawn from the data set of 300 images. Similar to the SINs, 50 two INs, 30 three and 25 four INs were examined.

Results presented in Table 3.1 indicate that the SINs significantly outperform

Image Network	No.	fx	fy	u	v
Consensus	Images	$\bar{x}$ ( $\sigma$ )	$\bar{y}$ ( $\sigma$ )	$\bar{u}$ ( $\sigma$ )	$\bar{v}$ ( $\sigma$ )
Random	2	23.3 (21.7)	20.7 (21.2)	23.7 (24.3)	23.5 (27.4)
SIN	2	11.5 (11.7)	11.1 (11.7)	13.3 (16.6)	24.1 (26.5)
Random	3	12.2 (10.2)	12.5 (10.8)	14.2 (11.1)	15.1 (11.3)
SIN	3	3.8 (3.9)	3.6 (4.3)	5.3 (7.5)	9.9 (7.9)
Random	4	8.9 (5.2)	8.9 (5.1)	10.8 (6.7)	11.1 (6.6)
SIN	4	1.59 (1.55)	1.49 (1.69)	2.74 (3.31)	3.88 (3.32)

Table 3.1: Percentage error results for image network configuration comparison.

the RINs for INs of smaller sizes. This trend is similar to that which was observed in Figure 3.7. These results indicate that INs which comprise of images with geometry based on properties of the CL, enforce more independence in planar calibration equations, and thus solve the camera parameters more accurately.

### Distortion Correction

A radial distortion correction experiment is undertaken to further quantify the accuracy of the calibration results employing SINs as opposed to RINs. The benefit of un-distorting the images is that the canonical calibration plane can be used as a ground truth. This allows the distortion correction residuals to be estimated and thus quantify the accuracy of the distortion correction parameters. The resulting calibration of a two image RIN and SIN are examined as this represents the minimum number of images required to solve the planar calibration equations. The plots in Figure 3.8 show the residual error vectors after un-distortion (scaled by a factor of 20). The SINs have removed the radial distortion more accurately from the image (mean residual 0.2 pixels)

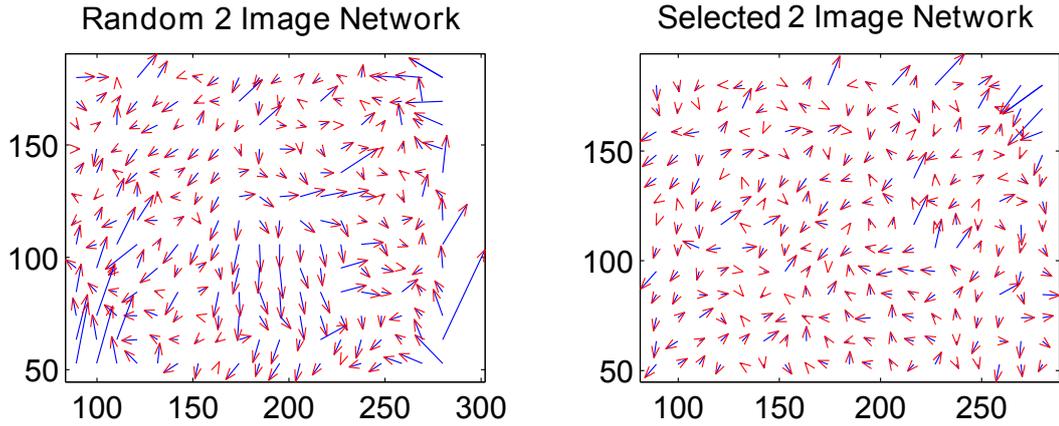


Figure 3.8: Residuals after un-distortion stage (in pixels).

compared to the random network, which still has significant distortion present (mean residual 0.39 pixels).

### 3.3.2 Generated Image Networks

Experiments are performed with real and synthetic data to illustrate the advantages of GINs over RINs. Initially, simulated data is used to verify the increase in calibration accuracy achieved when employing GINs. Subsequently, an examination on the reproducibility of the synthetic planar poses is undertaken. The process by which this is achieved and the accuracy obtained is outlined. Experimentation with real images is conducted to verify the increased efficiency of the calibration process using GINs. A benefit of GINs is highlighted in Figures 3.9(a) and 3.9(b) which presents an imaging configuration where the planar targets appear to be distinct in pitch and yaw orientation. However, on examination of the CLs of both images, it is found that the angle between the CLs is approximately  $5^\circ$ . Therefore the configuration is near-degenerate and results in poor camera parameter estimates. GINs provide a means to prevent such geometric configuration arising.

#### Synthetic testing

A synthetic camera was formed with  $f = 500$ ,  $\alpha = 0.99$ , and  $(p_x, p_y) = (300, 200)$ . Synthetic grids (of control points) were simulated in  $P^3$  and subsequently projected onto the image plane in  $P^2$ . Each grid formed an image

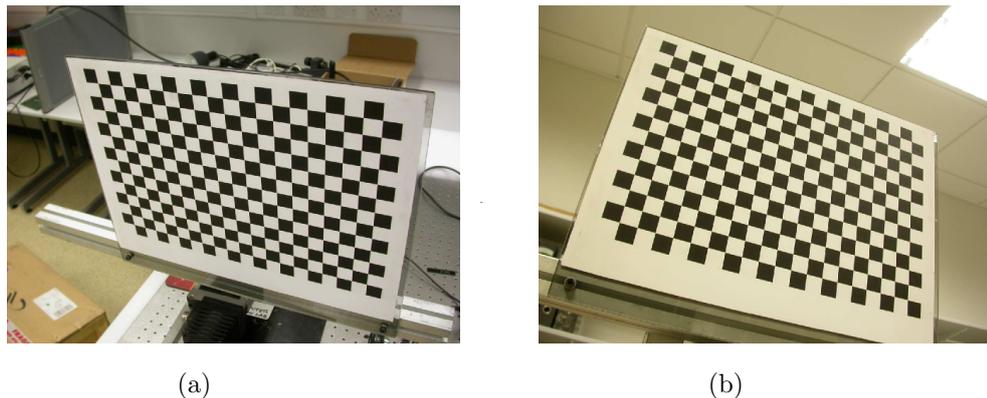


Figure 3.9: Image network with an angle of  $5^\circ$  between the CLs.

of  $600 \times 400$  pixels. A two parameter lens distortion model was applied to the images, and random noise with standard deviation 0.5 pixels was added to the grid point locations to simulate image noise. The test set comprised of 50 trials of each image network instance from 2 to 25 images. GINs were formed by taking one image from each RIN and subsequently generating the optimal image networks. The metric used to compare image network configurations is the mean of the absolute error between the ground truth parameters and the RIN/GIN parameter estimates for each network trial.

Figure 3.10 presents results for the internal parameters of the camera which were estimated using GINs and RINs. The error is presented in terms of the mean of the absolute error in the estimated parameters. It is clear from the results for each parameter that the geometric configuration of the GINs contribute to a better estimation of the camera parameters. The requirement for solving the planar camera calibration equations is that a minimum of two images are used thus giving four constraints on fitting the IAC and recovering the camera parameters. As the number of input images is increased, so too are the number of constraints on the IAC. Thus, it is evident in Figure 3.10 that beyond 6 input images, the additional constraints do not significantly aid in estimating the IAC more accurately. The increased performance of the GINs for INs with five or less images is a result of the planar target orientations being sufficiently independent to estimate the IAC. Degenerate configurations and near-degenerate configurations are avoided with GINs. However, with RINs, there is a greater likelihood of the planar target orientations being in a configuration which exhibits dependency, such as near-degenerate configurations (similar to those outlined in Section 3.1.1). Therefore, it is clear that consideration of image network geometry, in cases of less than five images, is

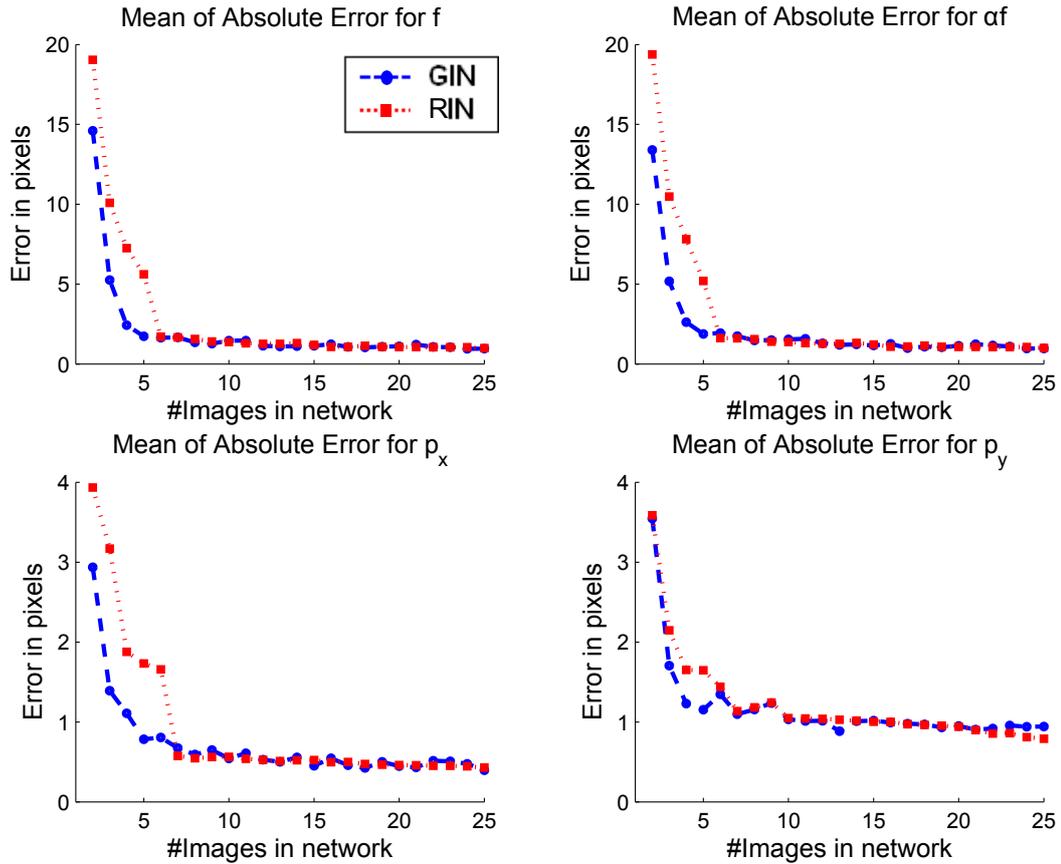


Figure 3.10: Intrinsic parameter estimation results for RINs and GINs for 2 to 25 image networks.

beneficial as an increase in calibration accuracy is achieved.

### Image reproducibility

The required accuracy for reproducing the synthetically generated images, such that the desired image can be captured, was outlined in the sensitivity analysis in Section 3.2.1. It was shown that the induced error on the angle between image CLs due to error in the orientation estimation remains bounded. Therefore, if general users capture the synthetic orientation within  $\pm 5^\circ$  on the grid pitch and yaw angles, the desired CL orientation is obtained.

In order to examine the accuracy to which users replicate the synthetic images, an indicative experiment was conducted. Four non-expert practitioners were provided with 25 optimal planar poses for which to undertake the replication process. The replication process is presented in Figure 3.11. The synthetic planar pose is augmented in a semi-transparent manner to the current camera



Figure 3.11: Augmented replication process for capturing GINs

Table 3.2: Non-expert practitioner results for optimal image reproducibility (Note: optimal angle is  $90^\circ$ )

Practitioner	Average Angle Error (deg) (SD (deg))
1	1.55 (0.51)
2	1.78 (0.88)
3	2.13 (1.06)
4	0.51 (0.47)

live-view. The border of the synthetic grid is highlighted for alignment purposes. A successful replication is achieved when the synthetic planar pose is visually aligned with the current live-view of the physical planar calibration target. The results presented in Table 3.2 represent a single image capture per optimal planar pose. Results indicate that each user was able to replicate the synthetic planar poses with sufficient accuracy to ensure good conditioning on the CL constraint. The largest average error, recorded by user three, was  $2.13^\circ$  with a Standard Deviation (SD) of  $1.06^\circ$ . This level of accuracy is sufficient to obtain accurate GINs.

### Image Network Stability

This experiment was conducted to verify the stability of calibration results obtained using GINs as opposed to RINs. The simulated experiments conducted have shown GINs to outperform RINs, particularly in cases where few input images are captured for the IN. Therefore, this experiment deals with three and four image networks. Since there is no ground truth calibration data available, the measurement used to quantify the performance of the GINs and RINs is the estimated parameter uncertainty. This can be calculated from an estimate of the Fisher information matrix ( $\mathfrak{F}$ ) which is formed upon convergence of the non-linear estimation process (Walter and Pronzato, 1997). The Fischer information matrix is defined as,

$$\mathfrak{F}(\hat{R}) = \frac{1}{\sigma_i^2} \sum_{i=1}^n \frac{\delta e^T(c_i, \hat{R}_k)}{\delta \hat{R}} \frac{\delta e(c_i, \hat{R}_k)}{\delta \hat{R}^T} \quad (3.16)$$

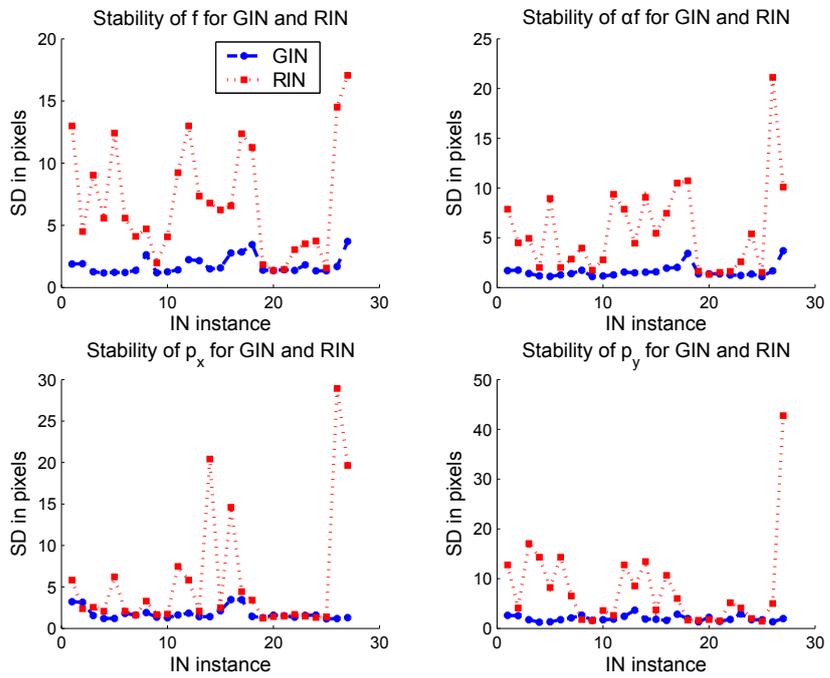
where  $(c_i, \hat{R})$  represents the cost function being minimised and  $\sigma^2$ , the unknown noise variance can be approximated as

$$\sigma^2 = \frac{1}{n_t - n_p} \sum_{i=1}^{n_t} e_i^2 \quad (3.17)$$

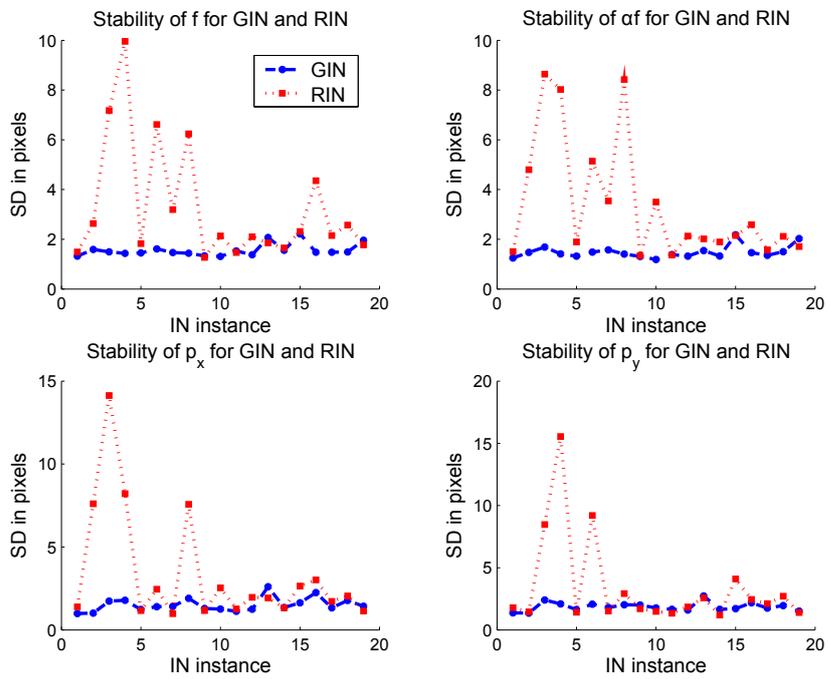
with  $n_t$  being the number of measurement points and  $n_p$  the number of parameters being estimated. The norm of the residuals from the estimation algorithm is defined as  $e_i^2$ . With the approximation of  $\mathfrak{F}$ , the uncertainty in the parameters is found by taking an estimate of the standard deviation as the square root of each diagonal element in  $\mathfrak{F}^{-1}(\hat{R})$ ,

$$SD(\hat{R}_i) = \sqrt{\text{diag}_i(\mathfrak{F}^{-1}(\hat{R}))} \quad (3.18)$$

Figure 3.12(a) presents the results of the uncertainty in the parameter estimation for 27 RINs and GINs in a three IN configuration. Similarly, Figure 3.12(b) depicts the same parameters for 19 RINs and GINs with four input images. Calibration accuracy is given by the SD associated with the uncertainty of the parameters estimated (in pixels). The estimated camera parameters employing GINs display superior stability to those of the RINs. This is a result of the desirable image network geometry of GINs. On closer examination of four image network instances, there is a noticeable increase in the stability of RIN estimates, however, GIN estimates exhibit greater stability across all 4 IN instances. Tabulated results of the average of the SD across all image network instances is given in Table 3.3.



(a)



(b)

Figure 3.12: Stability of intrinsic parameters for three (a) and four (b) image RINs and GINs

Table 3.3: Parameter uncertainty for three and four image networks. Note: Results are given in terms of SD (in pixels)

	$f$	$\alpha f$	$p_x$	$p_y$	$k_1$	$k_2$
3 RIN	6.88	5.67	5.50	7.74	0.008	0.050
3 GIN	1.79	1.59	1.76	1.99	0.005	0.040
4 RIN	3.09	3.31	3.31	3.39	0.007	0.037
4 GIN	1.55	1.48	1.51	1.85	0.005	0.036

Experiments undertaken with optimal image networks have shown GINs to improve the efficiency of planar camera calibration by taking the geometric configuration of the planar targets into consideration. By choosing suitable planar poses, independence in the planar calibration equations is achieved, thus fitting a unique IAC and avoiding configurations which lead to singularities. This is the key contribution of optimal image networks.

### 3.4 Discussion

This chapter addresses the issue of defining suitable images for planar camera calibration. The geometric configuration of input images used for calibrating a camera is an area which has largely been neglected. Degenerate configurations have been identified in the literature, however, near degenerate configurations, which also contribute to poor camera parameter estimates, are ever present when choosing planar target orientations. The proposed IN strategy, optimal image networks, addresses this issue by using the CL as a guide to identifying images which contribute independently to solving the camera parameters. As a result, planar calibration methods which employ OINs as an IN strategy can obtain greater accuracy and stability when solving the camera parameters.

Section 3.1 provides an analysis, in terms of planar target orientation, of the planar calibration problem. The criteria for independent image network geometry is outlined. Additionally, image network configurations which are degenerate are illustrated. The relationship of the CL and planar target orientation is detailed, which in turn highlights the applicability of the CL orientation (in the image frame) with regard to the overall image network configuration. Simulated experiments in Section 3.1.1 show that all degenerate and near de-

generate configurations are identified by analysing the relative orientation of image CLs. CLs which are equal by a reflection about a horizontal or vertical axis in the image provide no additional constraints when estimating the IAC. Thus, CLs which display these characteristics, will decrease the accuracy of the camera parameter estimates.

Optimal image networks are presented in Section 3.2. The ideal geometry of OINs is outlined and two subsequent strategies based on ideal geometry are proposed. The first method employs the CL approach to select planar calibration images from a large dataset of supplied images. A binary search approach is taken by analysing the CLs of the images in the data set. SINs are defined by forming an IN consisting of the selected images. The second approach proposed employs the CL constraints to generate synthetic images with ideal planar target orientations. This method is called generated image networks, GINs. The main advantage of implementing GINs is the reduction in input requirements in comparison to SINs, which require a large data set. Furthermore, there is a possibility that the data set provided may not contain images with ideal geometry. Thus, GINs ensure that the formed INs will contain the desired geometric configuration to obtain accurate camera parameter estimates.

Evaluation of OINs is completed in Section 3.3. As there is no standard method of selecting planar target poses, a random selection strategy is implemented, called random image networks. Pitch and yaw orientations are randomly drawn over an interval of typical values i.e. values which do not hinder the feature extraction process. The first OIN method, in the form of SINs, is evaluated given a data set consisting of 300 frames captured by a webcam of a planar calibration target. An image selection algorithm filters the data set to form SINs. These INs are evaluated against RINs chosen from the data set. Results demonstrate that the SINs clearly outperform the RINs for INs which contain few images. Further experimentation, in the form of distortion correction, is undertaken using the calibration results in the minimal case for both IN configurations. The mean residual error in the case of SINs is less than that of RINs, and it is clear that a greater quantity of the distortion has been removed using the SIN results as opposed to the RIN results.

Section 3.3.2 examines the accuracy of GIN parameter estimation using real and synthetic data. Simulated experiments show that GINs yield more accu-

rate camera parameter estimates for each individual parameter, particularly for INs of less than 6 images. Similar performance is achieved with RINs for IN containing more than 6 images. An integral component of GINs is the ability of the user to replicate the synthetically generated images of the ideal planar target orientation. Consequently, an indicative experiment was undertaken in which four non-expert practitioners completed the replication process for 25 images. Results demonstrate that the replication process is straightforward, with each user capturing the synthetic pose with sufficient accuracy to form GINs. Further investigation of INs with three and four images is conducted with real data to demonstrate the stability of the camera parameter estimates of GINs compared to those of RINs. Results indicate that GINs produce more stable and efficient results across a large range of three and four INs.

The improved performance observed with OINs confirms that, by considering image network geometry when selecting calibration images, more accurate calibration results can be achieved. This improved performance is primarily observed for image networks with 6 or less images. As discussed in Section 3.1, each input image to the IN provides two constraints in the form of the circular points to estimate the IAC. In the minimal case, two images provides four constraints which is adequate to recover camera parameters. However, if there is dependence between input images, a unique IAC may not be defined. Thus, by enforcing maximum independence (geometrically) between input images, configurations which will lead to poor camera parameter estimates are avoided. OINs provide such independence by manipulating geometric aspects of the CL in the form of the angle between image CLs. As the number of images in the IN increases, so too does the number of constraints on the IAC. Therefore beyond 6 images, which is equivalent to 12 constraints on the IAC, the difference in the IAC being estimated is minimal between OINs and RINs. Therefore similar accuracy is achieved with both IN configurations. However, the increase in accuracy of OINs for INs of less than 6 images greatly reduces the input requirements to the planar calibration problem and guarantees that degenerate and near-degenerate configurations are avoided.

# Chapter 4

## Multi-Pupil Imaging

Standard camera models used in the computer vision community are presented in Chapter 2. The pinhole camera model, and its calibration, was shown to lose depth information during the projection of 3D world points to 2D image points. Traditionally, methods for obtaining depth information using the pinhole camera model involved employing stereo vision techniques. Early examples of this are reported in the robotics area, and are summarised by Faugeras (1993). The main problem with stereo techniques from a practical sense was that either two sensors, or two images with known rotation and translation between both views, were required to recover metric information. Coupled with the computational overhead the computer vision community sought alternative methods to infer metric depth from images. Two such approaches to depth recovery have been proposed in the forms of Depth from Focus (DfF) (Grossmann, 1987, Ens and Lawrence, 1993) and DfD (Pentland, 1987). Both methods make use of camera models which introduce optical elements to the model such as focusing and aperture effects. Typically, multiple images are required along with multiple camera configuration settings to recover metric depth information. Additionally, the process of measuring image sharpness, and image blur, has long been identified as an error prone process in the computer vision community. Consequently the main goal of this chapter is to develop a new camera model which is capable of retrieving metric depth information from a single image, and avoids the tedious task of measuring optical depth of field artifacts.

The significance of the aperture or pupil plane was highlighted in Chapter 2. In

a thin lens capacity, it is the limiting factor for the rays entering and exiting the imaging lens. It also dictates the amount of image blur and sharpness present in the formed image. Experimentation in Chapter 2 revealed that a pupil plane could be subsampled by a smaller circular pupil. By calculating the centre of the bundle of rays which pass through the smaller pupil, it was shown that this was equivalent to selecting a single ray centred at that pupil's location in the overall scheme of the pupil plane. It is this basic principle upon which the multi-pupil imaging models are developed. Coupled with knowledge of the imaging lens focal length, image sensor size and metric makeup of the pupil plane, the proposed imaging models can retrieve metric depth information. The Double Pupil Model (DPM) is proposed in Section 4.1. It consists of a modified pupil plane containing two pupils separated by a known distance. A complete characterisation of the proposed model is undertaken to identify sensitivities of the depth estimation to model parameters. Industry standard optical design software is also used for the purpose of validating the DPM. Simulated results aim to demonstrate the accuracy of the proposed camera model for the purpose of depth estimation.

Generally, the optical arrangement of imaging systems consists of predefined glass surfaces and a number of stop/plane surfaces. The pupil plane of the imaging system plays a key role in defining the quality of the formed image. The location of the pupil plane within the system of lens elements is generally fixed (for a fixed focal length lens). Furthermore, lens designers fix the pupil plane location in such a way as to minimise certain aberrations present within the imaging lens. Therefore, in order to apply multi-pupil imaging techniques to standard imaging systems, the constraint on the placement of the pupil plane within the imaging system must be relaxed. A shifted multi-pupil camera model in the form of the Double Pupil Shifted Model (DPSM) is proposed in Section 4.2. This model builds on the presented multi-pupil theory and allows flexibility in the placement of the pupil plane. Axial shifts of the pupil plane are permitted along the optical axis. Industry standard optical design software is used to verify the DPSM. Subsequently, a complete characterisation of the model is also presented along with simulated results for depth estimation experimentation.

Camera calibration and its importance in recovering metric information from images was introduced in Chapter 2 and further developed in Chapter 3. An optimal scheme for selecting planar calibration targets was outlined for cameras

which could be modelled with the pinhole camera model. Section 4.3 examines the applicability of standard pinhole calibration methods to the multi-pupil imaging framework. The main parameter to be recovered for the multi-pupil imaging systems is the sensor depth. It is this parameter which defines the focus configuration and subsequently the image disparity profile of the DPM and DPSM systems. Calibration from both planar targets and spherical targets is implemented. These results are directly compared with a new proposed calibration method for the DPM camera model. However, with the DPSM, an additional parameter in the form of the pupil plane axial offset is required. Consequently, a novel calibration algorithm is proposed for the DPSM imaging system which recovers both sensor depth and pupil plane offset. Therefore, the comparison with standard pinhole approaches can be completed with knowledge of the pupil plane offset. Simulated experiments are carried out to identify favourable conditions to conduct the proposed calibration procedures. Subsequent depth estimation simulations are conducted to quantify the accuracy of the calibrations.

Experimentation with real data is undertaken in Section 4.4 to validate both proposed multi-pupil imaging models and their calibration. Initially, the proposed calibration algorithms are examined and their accuracy determined by comparison with ground truth data for a general imaging system configuration. Standard calibration approaches are also implemented and compared with the proposed approaches as well as with the ground truth. Section 4.4.3 presents depth estimation results across an object depth range of approximately  $4,300\text{mm}$ . Results for the DPM and DPSM imaging models employing calibration estimations of the standard approaches and the proposed multi-pupil approaches are presented and discussed.

## 4.1 Multi-Pupil Camera Model

The main property of the multi-pupil camera model is the pupil plane of the imaging system. A conventional pupil plane is generally located within the system of lens elements. The pupil, located on the pupil plane, is the limiting factor for light rays passing through the optical system. In Chapter 2, the imaging geometry associated with a conventional and unrestricted pupil is examined. It is shown that a conventional pupil, or aperture, can be sub-

sampled by a pupil of lesser size located within the aperture region without affecting the imaging geometry. Thus, for geometric image calculations, it can be assumed that the sub-sampled aperture allows a single centred ray to pass through the pupil plane at the location of the sub-sampling pupil.

The camera model developed in this chapter differs from the conventional camera models in that it contains multiple pupils in the pupil plane. The effect of each pupil is that of a single ray passing through an effective aperture at its location. This is similar to selecting rays from certain parts of a conventional aperture. Thus, the theoretical foundation on which the model is developed is based on geometrical ray tracing.

Paraxial optics are assumed for development of the camera model. The pupil plane, situated at the centre of the ideal system, contains two pupils separated in the transverse axial direction. For any separation of this nature, the two pupils can be seen as sampling the extreme rays in the pupil plane of an equivalent aperture with a diameter of the given transverse separation distance. Assuming on-axis objects, an optical system focused on a world point will form an ideal focused point at the centre of the image plane. Conversely, an object point not in focus will form a double image of the point with each pupil of the pupil plane forming its own image of the object point. The relationship between the transverse pupil separation and the image point separation, or disparity, along with system optical properties, provides the basis for calculating true object depth information.

### 4.1.1 Double Pupil Imaging Model

The double pupil imaging model is developed using the optical system presented in Figure 4.1. This system contains an ideal thin lens in which the pupil plane is centrally located. Pupils of the optical system are separated by the distance  $P_g$  and are assumed to be of a diameter that allows a single ray to pass through the pupil plane. An axial object point is located at a distance  $u_0$  from the pupil plane. The configuration of this optical system is such that the image conjugate distance,  $v_0$ , is less than the pupil plane to image plane distance  $v$  (the sensor depth). Such optical configurations are known to be “out of focus”, therefore the image plane will contain an image of the object point emanating from each pupil. The distance between both image points is

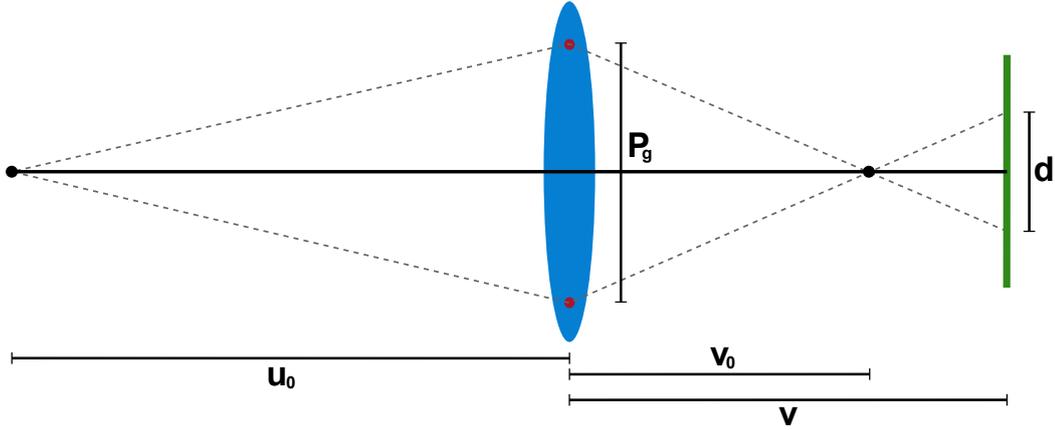


Figure 4.1: Double pupil imaging model.

known as image disparity and is represented by the  $d$  term.

The disparity,  $d$ , can be related to the object distance,  $u_0$ , through manipulation of the ideal lens model.

$$\frac{1}{u_0} + \frac{1}{v_0} = \frac{1}{F} \quad (4.1)$$

where  $F$  is the focal length of the lens. This equation can be rearranged to yield a result for the object distance  $u_0$  in terms of the focal length and the conjugate image distance.

$$u_0 = \frac{v_0 F}{v_0 - F} \quad (4.2)$$

By comparing similar triangles, the triangle with the disparity  $d$  as its base and the conjugate image point as its apex is similar to the triangle sharing the same apex point and with the pupil separation distance  $P_g$  as its base. An expression for the image conjugate distance  $v_0$  can be derived from this similarity and is given by

$$v_0 = \frac{P_g v}{P_g + d} \quad (4.3)$$

By substituting the expression for  $v_0$  in Eqn. 4.3 into Eqn. 4.2, a result for the object distance  $u_0$  can be derived

$$u_0 = \frac{P_g v F}{P_g v - F(P_g + d)} \quad (4.4)$$

which relates the object depth to the disparity in the image,  $d$ , the pupil separation distance,  $P_g$ , the focal length,  $F$ , and the pupil plane to image distance,  $v$ .

Given an image of a scene containing multiple on axis objects generated by any optical configuration, i.e. any distance  $v$ , the object depth can be estimated. A change in  $v$  equates to a change of optical focus within the system. Functionality of the double pupil model relies on the disparity between the imaged object point from each pupil of the pupil plane which is captured in a single image.

Disparity generated at the image plane is dependent on three model parameters,  $F$ ,  $P_g$ , and  $v$ . All three properties must be known to recover scene depth from a single image. The focal length and pupil separation distance are fixed input parameters to the system, therefore the sensor depth must be calculated for the optical configuration in use. This is achieved through a calibration process (see Section 4.3). Assuming a system has been calibrated for  $v$ , the last remaining challenge in terms of the practicality of the DPM is that of feature detection and matching. Since an object point is imaged by each of the pupils separately, in order to calculate the image disparity, the imaged points generated by each pupil must be matched. The scale invariant feature transform (SIFT) developed by Lowe (2004) is used to locate features and a standard matching algorithm is used to match the imaged points. This is a robust feature detection algorithm and generally accepted as the gold standard within the computer vision community.

A closer examination of Figure 4.1 reveals that the DPM was developed with the optical system focusing the image point between the image plane and the lens element. The model could equally have been derived with the system focusing the image point behind the image plane resulting in a sign change of the disparity in Eqn. 4.4. This change in sign of the image disparity is critical to overcome the problem of depth estimation when the magnitude of the measured disparities are equal. Such ambiguities arise when an optical configuration is focused at a fixed distance and the disparities being measured correspond to object points which are equidistant on both sides of the focus plane (as indicated in Figure 4.2). This problem has been avoided in the computer vision community by assuming that all object points are on a single side of the focus plane. The cost of this assumption is that the optical system is constrained to focusing at infinity and thus loses flexibility of optical configurations.

The DPM overcomes the depth ambiguity by matching image points to the pupil through which they have been projected. This allows for changes in sign

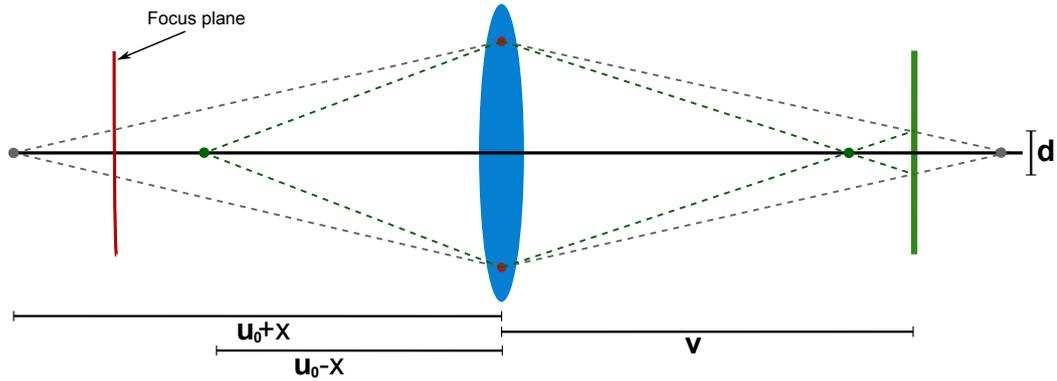


Figure 4.2: Equidistant object points.

of the disparity to be identified and thus distinguishes between object points which lie in front of the focus plane and object points which lie behind the focus plane. Pupil and image point matching is accomplished by introducing colour filters to the pupils in the pupil plane. By choosing appropriate wavelengths, the image captured by the optical system can be separated into two images - one from each pupil. This process also improves the feature detection and matching algorithm. Since each pupil can be separated into its own image, disparities of smaller magnitudes are easier identified. This would be a difficult task to detect and match in a single image. The proposed solution to the depth ambiguity ensures that the DPM maintains flexibility of optical configurations within the imaging system.

### 4.1.2 Simulated Experiments

Experiments were carried out using simulated data to verify and evaluate the DPM in terms of the overall accuracy of its depth estimates. Evaluation of the depth estimates in the presence of pixel noise is performed along with a characterisation of the model in terms of the sensitivity of its parameters. The initial model verification is performed using the optical design software package Zemax. The second experiment is conducted in order to identify which parameters within the model are most sensitive to perturbations and thus require greater accuracy when being estimated. The final experiment evaluates the accuracy of the DPM when estimating depth from a scene with varying input parameters and image noise. The model characterisation and depth estimation experiments are performed using a simulated camera with a pixel pitch of  $4.5\mu m$ , and sensor length and width of  $18mm$  and  $13.5mm$

Table 4.1: DPM Zemax verification results for near, mid and far focus settings.

$u_0$ (mm)	Disparity (mm)			Estimated Depth (mm)		
	Near focus	Mid focus	Far focus	Near focus	Mid focus	Far focus
500	0	-2.9868	-3.18	500	500	500
1500	2.7309	-0.6637	-0.8833	1500	1500	1500
2500	3.2771	-0.1991	-0.424	2500	2500	2500
3500	3.5111	0	-0.2771	3500	3500	3500
4500	3.6412	0.1106	-0.1177	4500	4500	4500
5500	3.7239	0.181	-0.0481	5500	5500	5500
6500	3.7812	0.2297	0	6500	6500	6500

respectively. The corresponding lens diameter of the simulated imaging system is set to  $25\text{mm}$  as this is the physical diameter of the real lenses used in Section 4.4.3. The object depths are set between  $300\text{mm}$  and  $8,000\text{mm}$  in steps of  $50\text{mm}$  which covers the near/mid/far focus regions of a conventional imaging system.

### Zemax Verification

Verification of the DPM is performed with the Zemax software. Simulating the DPM from first principles in Zemax is accomplished by specifying a paraxial surface (thin lens) with a diameter of  $25\text{mm}$  and focal length  $85\text{mm}$ . A custom pupil plane is designed containing a double pupil with a centred transverse separation distance of  $20\text{mm}$  and pupil diameters of  $1\text{mm}$ . The optical design of the system is shown in Figure 4.3(a). Another input parameter to the system is the wavelength of light, which is set to the primary wavelengths of the visible spectrum ( $0.486\mu\text{m}$ ,  $0.587\mu\text{m}$ ,  $0.656\mu\text{m}$ ). Axial objects are set at distances of  $500\text{mm}$  to  $6,500\text{mm}$  in steps of  $1,000\text{mm}$ . Three focus distances are taken to verify that the equal magnitude disparity condition is solved by the DPM. Thus,  $v$ , corresponds to focus distances at  $8,500\text{mm}$ , mid-range ( $3,500\text{mm}$ ) and near-focus ( $500\text{mm}$ ). The aim of this simulation was to verify the correctness of the DPM using the industry standard optical design software. Figure 4.3(b) displays the output on the image plane of the DPM optical system. This function of Zemax is typically used to analyse the PSF for a given optical system. In this instance, it yields the image disparity as a result of the double pupil on the pupil plane. For the purpose of measuring the disparity, the rays displayed in Figure 4.3(b) are cut-off at the centroid of

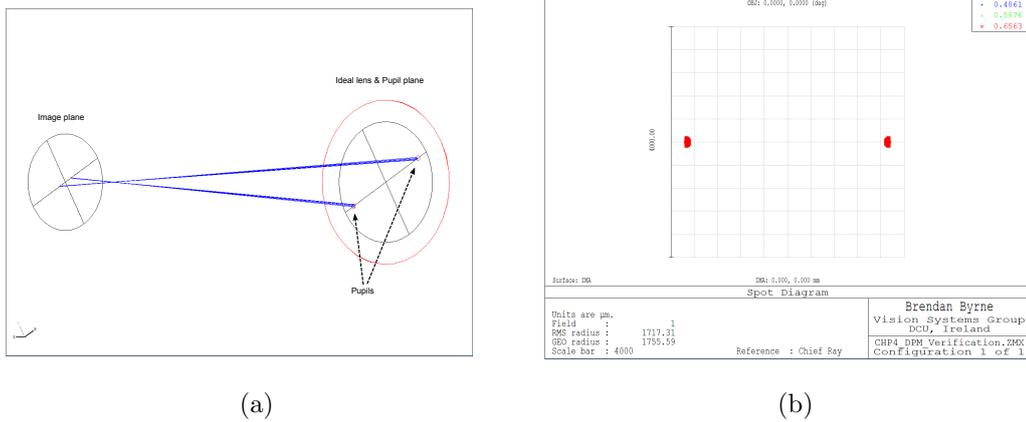


Figure 4.3: (a) DPM Zemax Verification Layout; (b) Point Spread Function of the DPM.

the resulting image points on the image plane. This is necessary to calculate the geometric distance between the imaged points within Zemax. For this particular example, the geometric radius is  $1755.59\mu\text{m}$ , thus the disparity is twice this figure. Table 4.1 presents the results of the Zemax experiment. Since an ideal optical system is simulated, there is no noise present in the system. As a result, the estimated depth for each object point ( $u_0$ ) is found to be exact and thus verifies that the DPM is indeed a valid imaging model. In addition, these results also highlight that the imaging model overcomes the problem of distinguishing between equidistant object distances from the focus plane due to the sign of the disparity.

## Model Characterisation

An analysis on the sensitivity of the DPM parameters is a key factor in determining how accurately the model can estimate depth. There are four input parameters to the DPM,  $F$ ,  $P_g$ ,  $v$ , and  $d$ . Two of these are known: The focal length which is a given optical property of a lens, and the pupil gap which is predefined via a manufacturing process. The remaining parameters,  $v$  and  $d$ , are determined via calibration and from the image sensor respectively. This analysis is completed in two stages. Initially the fixed parameters are examined in terms of the overall impact they have on the model. Subsequently, the measured parameters are analysed and an overall characterisation of the model parameters is presented.

The choice of focal length for an imaging system varies depending on the application. In terms of practicality, generally as the focal length increases with respect to the lens diameter, an increase in object magnification is observed in the image coupled with a decreased field of view. This observation is also true with respect to the DPM. The influence of increasing or decreasing the focal length in the DPM results in a magnification or increase/decrease in the object disparities generated by the imaging system. Therefore the most significant property of  $F$  in terms of the DPM is that it either increases or decreases disparity, depending on its optical power. The limits to which it can be increased or decreased depends firstly on the physical size of the imaging sensor being used, and secondly, on the intended scene to be captured i.e. if a relatively large field of view is required, a smaller focal length should be chosen. The significance in choice of the pupil gap  $P_g$  is solely dependent on other fixed properties of the imaging system such as the lens diameter and focal length. The observed effect of increasing or decreasing the pupil gap in the DPM equates to a scaling in the magnitude of the image disparities. In a similar manner to the practical aspects which influence the choice of  $F$ , these properties equally influence the choice of  $P_g$ .

Although the fixed parameters of the DPM are dependent on optical and physical properties of the imaging system, the sensitivity of the depth estimation to perturbations in these parameters is examined. Table 4.2 presents the measurement accuracy for each fixed parameter in the DPM as well as their sensitivity to the depth estimation process. The measurement accuracy of the parameters presented in Table 4.2 reflects the achievable resolution in respect to the experiments carried out in Section 4.4.3. The resolution to which the focal length and pupil gap are measured is in the order of micrometres. Therefore the net error propagated through the system to the depth estimation is of a small magnitude, particularly in the case of perturbations in the focal length. Figures 4.4(a)-(c) show the error profiles in the depth estimation process for variation in the focal length at the measurement resolution. A range of focal lengths between  $35mm$  and  $200mm$  in steps of  $5mm$  are tested, which are within the range of standard focal lengths used in conventional imaging systems. A single pupil gap of  $10mm$  is chosen (for 4.4(a)-(c)) as variations in  $P_g$  do not affect the depth sensitivity to  $F$  (this is a result of the proportional increase/decrease of disparity with pupil gap separation distance). The maximum depth estimation error due to perturbations in  $F$  is less than  $0.1mm$  across the three focus

Table 4.2: DPM Parameter measurement accuracy and sensitivity.

DPM Parameter	Measurement Accuracy	Depth Sensitivity
$F$	$10\mu m$	$\frac{(P_g v)^2}{(P_g v - F P_g - F d)^2}$
$P_g$	$7.8\mu m$	$\frac{-F^2 dv}{(P_g v - F P_g - F d)^2}$
$v$	Calibration	$\frac{-P_g F^2 (P_g + d)}{(P_g v - F P_g - F d)^2}$
$d$	Sub-pixel	$\frac{P_g v F^2}{(P_g v - F P_g - F d)^2}$

settings (sensor depths) for each focal length. Figures 4.4(d)-(f) present the error in  $u_0$  estimation due to perturbations in the pupil gap  $P_g$ . Three pupil gap separation distances are chosen,  $6mm$ ,  $10mm$ , and  $20mm$  along with a fixed focal length of  $85mm$ . The focal length remains fixed for these experiments as variations in this parameter do not affect the errors in depth estimation due to perturbations in  $P_g$ . This is a result of the proportional change in disparity with focal length. The maximum error induced in the depth estimation due to a unit variation in  $P_g$  is  $250mm$  at a depth of  $8,000mm$  which represents a 3.1% error (Figure 4.4(d)). This error is reduced with the increase in pupil gap separation distance. At a separation distance of  $20mm$ , this error is reduced to 0.08% at the same object depth (Figure 4.4(f)). There are a number of conclusions to be drawn from these simulations of the fixed model parameters. Firstly, the model is applicable to any focal length within the constraints of the given physical sensor dimensions. Secondly, depending on the lens diameter of the imaging system, a larger pupil gap separation distance is favorable as errors in measurement of larger gap distances induce less errors in the final depth estimates compared to smaller  $P_g$  settings.

A deeper understanding of the measured parameters of the DPM is best gained by initially analysing the image disparity profile of the DPM imaging system. Figure 4.5 presents the disparity profile for a lens with focal length  $85mm$  and pupil separation distance  $10mm$  using simulated data. The profile is represented in pixel units across all focus settings at every object depth. As previously noted, changing  $P_g$  has the effect of scaling the image disparity. Similarly, changing the focal length simply scales the image disparities. However, the disparity profile remains constant across all sensor depths regardless of changes in  $F$  and  $P_g$ . There are two distinct image disparity profiles within the DPM, one representing a near-focused system and the other a far-focused system. The most significant difference is that image disparities increase with object depth

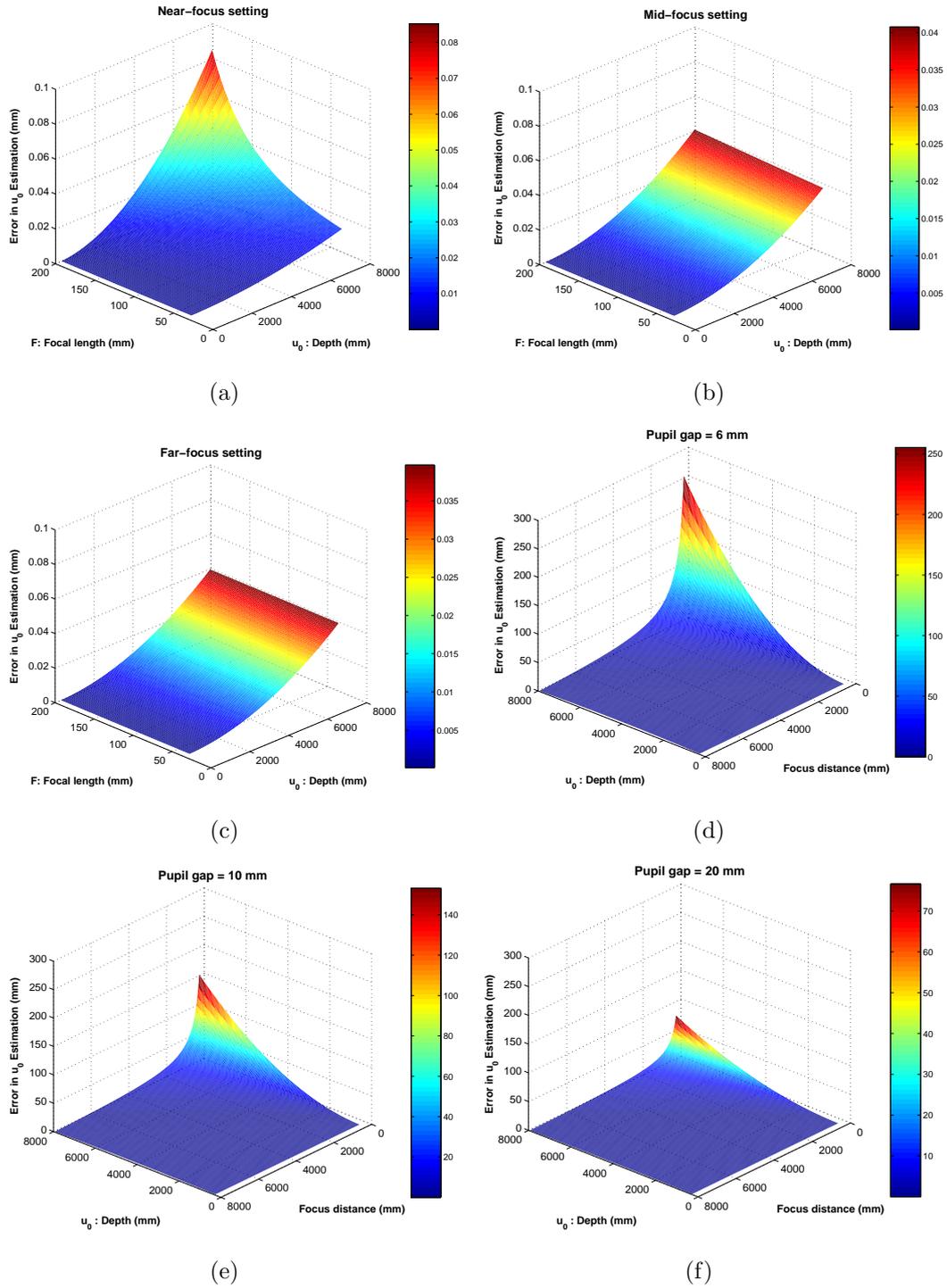


Figure 4.4: DPM Depth estimation sensitivity to  $F$  and  $P_g$  (a)  $F$ : Near-focus setting; (b)  $F$ : Mid-focus setting; (c)  $F$ : Far-focus setting; (d)  $P_g$ : For  $P_g = 6mm$ ; (e)  $P_g$ : For  $P_g = 10mm$ ; (f)  $P_g$ : For  $P_g = 20mm$

for a near-focused system and decrease with depth for a far-focused system. This plays an important role in choosing a sensor depth for the DPM. In order to achieve accurate depth estimates, the magnitude of image disparities must be independent at every object depth. Additionally, each disparity must be distinct within the resolution of pixel measurement in the imaging system. On examination of Figure 4.5, the areas of the profile which exhibit less sensitivity (as the object depth increases) result in smaller differences in magnitudes of image disparities. If the differences in disparity magnitudes for distinct object depths is less than the measurement resolution of the imaging system, then the DPM will be unable to discriminate between these object depths. The measurement resolution of the DPM is to sub-pixel accuracy.<sup>1</sup> Therefore unless two image disparities are identical at a sub-pixel level, the DPM will accurately estimate the corresponding depths. With the introduction of noise to an imaging system, perturbations in the measurement of image disparities can lead to errors in depth estimates, particularly as object depth increases and the disparity difference decreases. Consequently, greater sensitivity of the image disparity profile is desirable at larger object depths. With the two distinct profiles of Figure 4.5, this criteria is better served by larger sensor depths with respect to the focal length i.e. the near-focused system. Although both profiles exhibit reduced sensitivity at larger object depths, there is a greater range of disparity and differences in disparity with a near-focused system in comparison to a far-focused configuration.

The implications of the disparity profile are visible in the sensitivity analysis performed for the sensor depth parameter  $v$ . Figures 4.6(a)-(c) show the sensitivity of the estimated depth to a perturbation in  $v$  of a magnitude in the order of the maximum resolution to which  $v$  can be estimated (see Section 4.3). The simulation was performed with a fixed  $P_g$  of  $10mm$  and three focal lengths ( $35mm$ ,  $85mm$ , and  $200mm$ ) representing short, mid and long focal lengths (changes in  $P_g$  have no effect on sensitivity of  $u_0$  to  $v$  due to the proportionality between  $P_g$  and  $d$ ). The results indicate that there is less error in the estimated depth for systems which have larger sensor depth values (shorter focus distances). This is in agreement with the disparity profile which illustrates that a larger range of disparity is achieved with near-focused configurations, thus the induced error due to  $v$  perturbations has less effect in the

---

<sup>1</sup>Achieved by calculating the centroid of the resulting imaged point emanating from each pupil

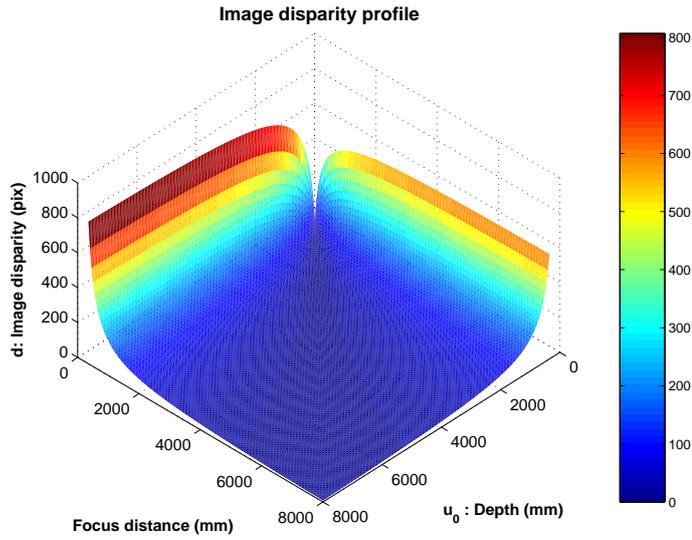


Figure 4.5: Image disparity profile of the DPM.

depth estimation at these configurations. In terms of choosing an operating focal length, practical aspects must be taken into consideration. The larger the focal length, the bigger the disparities and in turn, less error is propagated through to the depth estimation due to errors in  $v$ . The main disadvantage of choosing larger focal lengths is the increased magnification of objects on the image plane which also yields a smaller field of view. Therefore a mid-range focal length is most suitable for the DPM due to the minimisation of induced error 0.2% error at 8m compared to 1.6% at the short focal length (35mm) (see error at 8m in Figures 4.6(a) and (b)).

The sensitivity of the depth estimation to variation in  $d$  follows the observed trend in the disparity profile. As such, with a large sensor depth, the induced error in depth estimation is less compared with the far focused system. Figures 4.6(d)-(f) present the simulated data for a short, mid and long focal lengths for all focus settings corresponding to the depth range within the experiment. Similarly to the sensor depth analysis, the long focal length (200mm) yields the least sensitive depth estimates. The medium and short focal lengths induce errors of approximately 0.27% and 0.88% at the largest sensor depth setting. Changes in  $P_g$  have the effect of increasing/decreasing the magnitude of the image disparity. The results in Figures 4.6(d)-(f) represent a pupil gap of 10mm. The net result of increasing  $P_g$  is a scaling of the image disparity magnitudes which in turn induces less error in the final depth estimates. With  $P_g$  set at 10mm the largest error in depth estimation is 1% at 8m (with a far-focused configuration and  $F$  at 35mm). This is reduced significantly by

increasing  $P_g$  as well as choosing a near-focused configuration (0.4% at 8m with  $P_g$  set at 20mm).

An insight into the sensitivities of the DPM has been gained through performing the model characterisation. The input parameters are divided into fixed and varying parameters. The choices for focal length and pupil gap distance (fixed parameters) play an important role in defining the accuracy of the DPM depth estimation. These parameters are limited by the physical properties of the imaging system. Ideally the pupil gap distance should be maximised within the diameter of the lens. Similarly, the focal length should be chosen at a distance which will yield the largest range of object depths for which the corresponding image disparities remain within the active area of the image sensor. Accuracy of the varying parameters depends largely on the resolution to which they can be measured. In the case of the image disparity, this can be measured to sub-pixel accuracy which in turn leads to accurate measurements of object depth. The sensor depth, which is a calibrated parameter, also affects the DPM disparity profile. The DPM can be optimised to achieve depth estimation over a large range based on the sensor depth. If a large sensor depth is chosen, the optical system is in a near-focused configuration. It is this configuration that yields the most favorable disparity profile for large range depth estimation.

### Depth Estimation

Robustness of the DPM depth estimation was evaluated using additive Gaussian pixel noise to simulate perturbations in the measurement of image disparities, and to a lesser extent, image sensor noise. The simulated camera sensor size was  $18mm \times 13.5mm$  with a pixel pitch of  $4.5\mu m$ . The lens diameter was set to  $25mm$  with experiments carried out for focal lengths covering the near, mid, and long focal ranges ( $35mm$ ,  $85mm$ ,  $200mm$ ). Object depths were set within the range  $300mm$  to  $8,000mm$  in steps of  $50mm$ . Two pupil gap separation distances were simulated ( $15mm$  and  $22.5mm$ ) as well as two optical configurations for sensor depth corresponding to near-focus and far-focus. Figure 4.7(a)-(f) and Figure 4.8(a)-(f) each represent a simulated camera of a given  $F$ ,  $P_g$  and  $v$  in which each point on the error surface is the mean of 1000 simulated experiments. Additive noise with a SD between 0 and 5 pixels in steps of 0.1 is applied to each image before the depth is estimated.

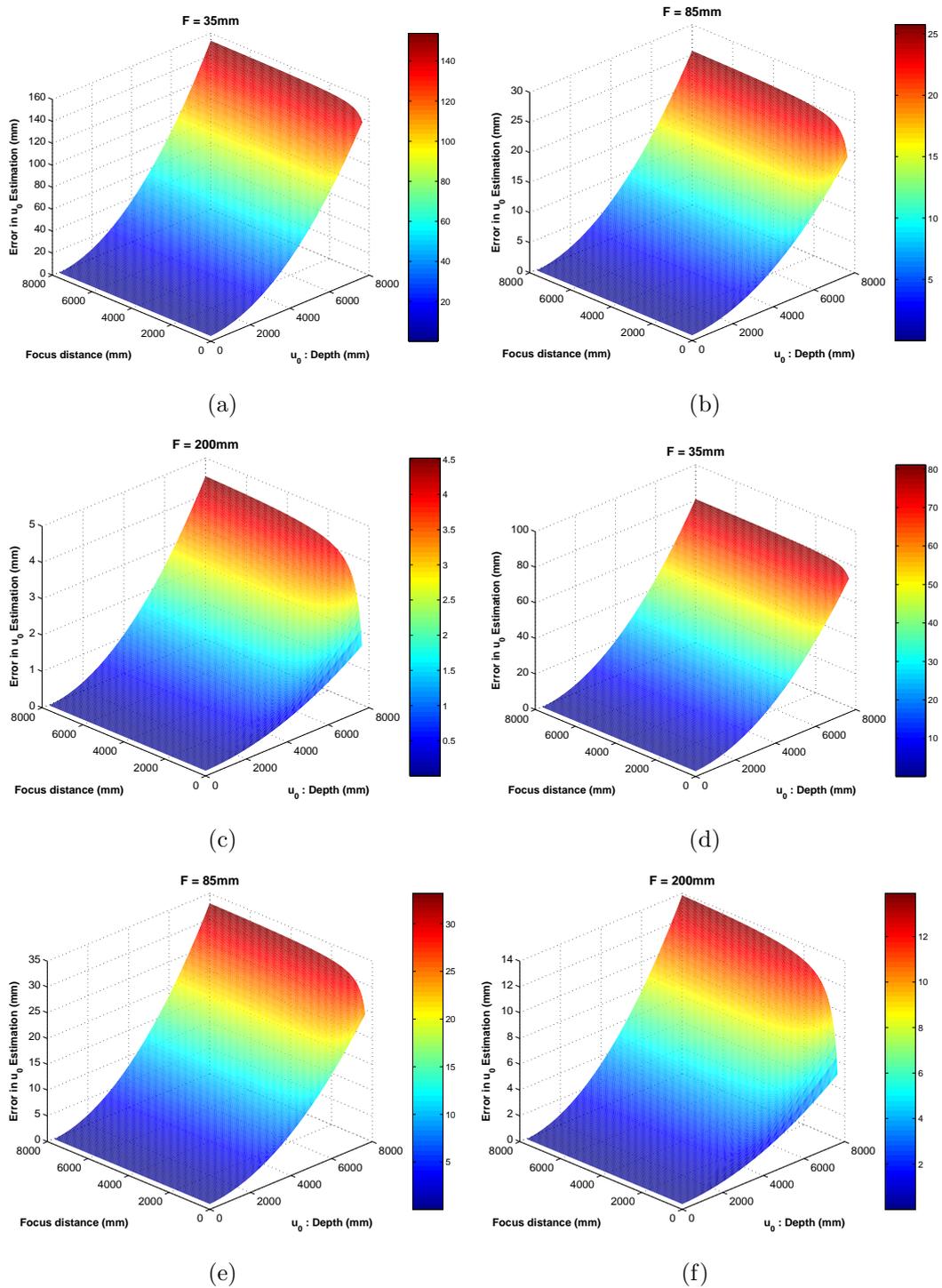


Figure 4.6: DPM Depth estimation sensitivity to  $v$  and  $d$  (a)  $v$ :  $F = 35mm$ ; (b)  $v$ :  $F = 85mm$ ; (c)  $v$ :  $F = 200mm$ ; (d)  $d$ :  $F = 35mm$ ; (e)  $d$ :  $F = 85mm$ ; (f)  $d$ :  $F = 200mm$ .

Table 4.3: DPM depth estimation results at 8,000mm across various input settings with additive Gaussian pixel noise (0 – 5 pixels).

Input Parameters	$F(mm)$	35mm				85mm				200mm			
	$P_g(mm)$	15		22.5		15		22.5		15		22.5	
	$v$ (focus)	Near	Far	Near	Far	Near	Far	Near	Far	Near	Far	Near	Far
Depth Estimation	Mean (mm)	102.13	114.93	67.68	76.09	33.97	47.22	22.56	31.63	6.71	19.98	4.47	13.40
Error (mm)	SD (mm)	31.70	35.59	20.87	23.21	10.51	14.97	6.93	9.67	2.07	6.15	1.38	4.14

As discussed during the DPM characterisation, the profile of image disparities becomes less sensitive as object depth increases. This results in smaller differences in magnitudes of image disparities at the larger depths. For instance, there is no experimental value in adding  $\sim 15$  pixels of noise to an image disparity of similar magnitude. Therefore a normalisation is performed to ensure that the added Gaussian noise in each experiment is apportioned correctly. This consists of re-scaling the noise proportional to the object depth. A reference object depth of 750mm is chosen as the scaling factor.

The results in Table 4.3 show the mean estimated error and its SD for a single object depth of 8,000mm. Each simulated configuration contains additive Gaussian noise with SD between zero and five pixels. Since the image disparities are least sensitive at this object depth, these figures represent the worst case scenario depth estimates across all simulations. The depth error varies from 114.93mm ( $F : 35, P_g : 15, v : far$ ) to 4.47mm ( $F : 200, P_g : 22.5, v : near$ ) depending on the camera configuration. Figures 4.7(a)-(f) and 4.8(a)-(f) are presented with the near-focus configurations in (a), (c), and (e) while the far-focus configurations are shown in (b), (d), and (f). It is clear that the near-focus configurations outperform the far-focus settings across all DPM input parameters. This is in agreement with the observations made on the characteristics of the DPM. The image disparity profile of the near-focus configuration yields higher image disparity sensitivity across the object depth range, which in turn, is more robust to the additive noise. An increase in the pupil gap distance, which results in larger image disparities, similarly allows more accurate depth estimates. Further refinement in the accuracy is observed with an increase in focal length of the camera lens. This has the effect of magnifying the image disparities and thus increasing their magnitudes.

The results presented in this section demonstrate the robustness and accuracy of the DPM in estimating depth over a large range of object depths. Depending on the choice of input parameters, different levels of accuracy can be

achieved. The largest error is seen at the shortest focal length. In a far-focused configuration, the mean error is  $114.93\text{mm}$  (SD  $35.59\text{mm}$ ) which represents an overall depth estimation error of approximately 1.43% (SD 0.4%). If longer focal lengths are chosen, accuracy of less than 1% can be achieved in the depth estimation. This level of accuracy enables flexibility in the choice of input parameters to a DPM imaging system.

## 4.2 Shifted Multi-Pupil Camera Model

The multi-pupil camera model in the form of the DPM has been presented with characteristics suitable for depth estimation, and the conducted simulations demonstrate its good estimation accuracy. However, the limitation of positioning the pupil plane either within the lens (during manufacturing) or directly in front of the lens as an approximation, remains problematic when considering integration of the DPM into consumer optical systems. This limitation is overcome with the introduction of the Double Pupil Shifted Imaging Model (DPSM). Generalisation of the axial pupil plane location, within the imaging system is the fundamental property of the DPSM. Flexibility in positioning the pupil plane overcomes the practical issues raised with the DPM.

The general imaging systems considered in this chapter consist of an image sensor, a thin lens, and a pupil plane. If the sensor depth is known (calibrated), the location of the lens and pupil plane relative to the image sensor are also known. The key difference with the DPSM is that the pupil plane is no longer constrained by placement at the lens aperture. The amount by which the pupil plane can shift depends on the sensor depth of the imaging system and the physical diameter of the lens. For a given pupil gap separation distance, the net effect of shifting the pupil plane is a magnification of the image disparities. This suggests that a larger pupil gap is being observed at the lens plane, which, is the case if rays are traced from the object point through the shifted plane to the lens. Rather than trying to estimate the apparent change in the pupil gap each time through a DPM system, an additional parameter is introduced in the form of an (optical) axial shift. This represents the distance between the pupil plane and the lens optical centre. If this distance is known along with the pupil gap of the pupil plane, the principles of the DPM, in terms of depth dependent image disparity, can be applied to the DPSM. Addition of

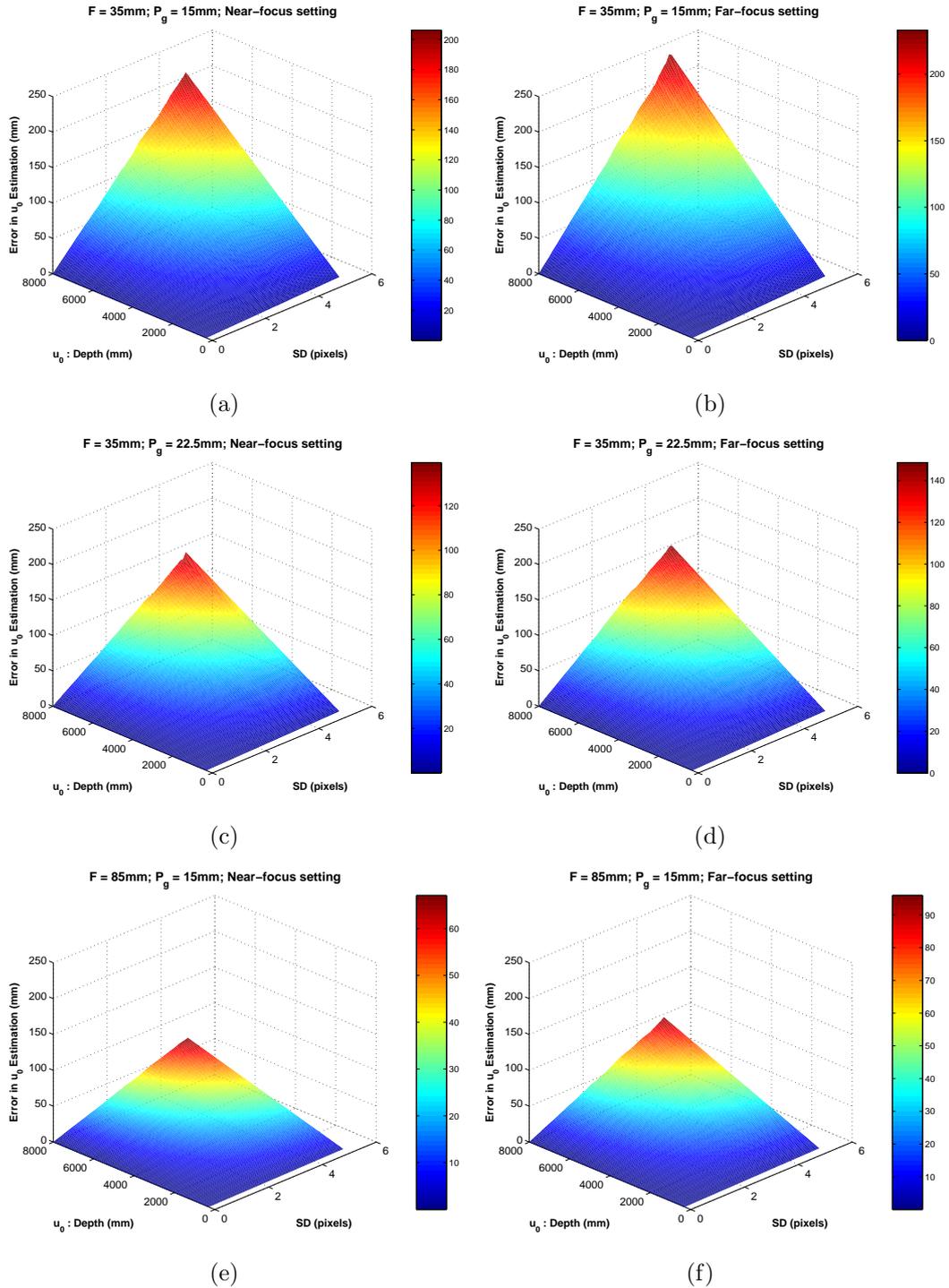


Figure 4.7: DPM Depth estimation in the presence of pixel noise across a range of input parameters (a)  $F: 35\text{mm}, P_g: 15\text{mm}, v: \text{near-focus}$ ; (b)  $F: 35\text{mm}, P_g: 22.5\text{mm}, v: \text{far-focus}$ ; (c)  $F: 35\text{mm}, P_g: 22.5\text{mm}, v: \text{near-focus}$ ; (d)  $F: 35\text{mm}, P_g: 22.5\text{mm}, v: \text{far-focus}$ ; (e)  $F: 85\text{mm}, P_g: 15\text{mm}, v: \text{near-focus}$ ; (f)  $F: 85\text{mm}, P_g: 15\text{mm}, v: \text{far-focus}$

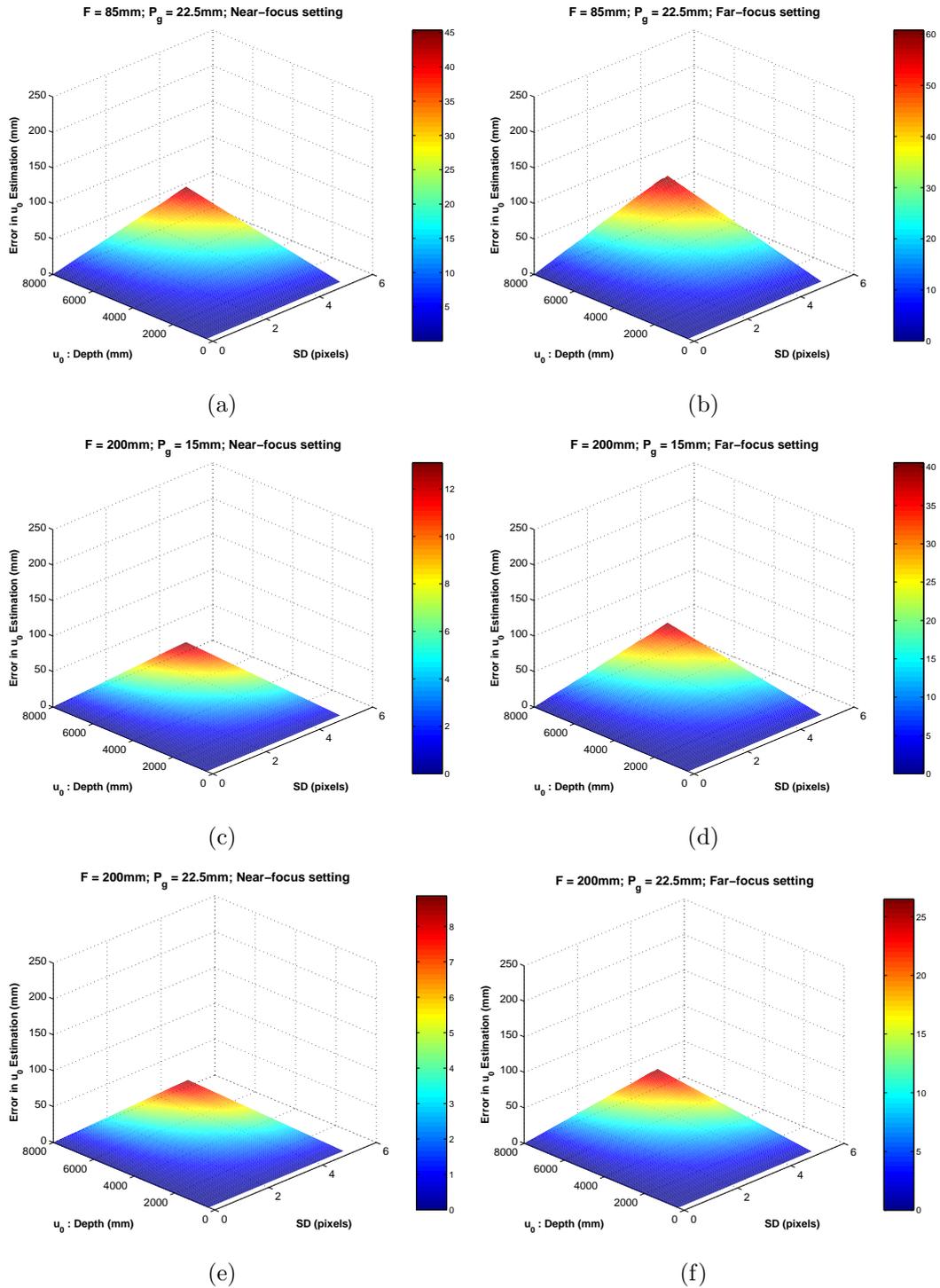


Figure 4.8: DPM Depth estimation in the presence of pixel noise across a range of input parameters (a)  $F: 85\text{mm}$ ,  $P_g: 22.5\text{mm}$ ,  $v$ : near-focus; (b)  $F: 85\text{mm}$ ,  $P_g: 22.5\text{mm}$ ,  $v$ : far-focus; (c)  $F: 200\text{mm}$ ,  $P_g: 15\text{mm}$ ,  $v$ : near-focus; (d)  $F: 200\text{mm}$ ,  $P_g: 15\text{mm}$ ,  $v$ : far-focus; (e)  $F: 200\text{mm}$ ,  $P_g: 22.5\text{mm}$ ,  $v$ : near-focus; (f)  $F: 200\text{mm}$ ,  $P_g: 22.5\text{mm}$ ,  $v$ : far-focus

the shifting parameter thus generalises the DPM camera model.

### 4.2.1 Double Pupil Shifted Imaging Model

The model parameters of the DPSM are presented in Figure 4.9. The imaging system consists of a sensor at a distance  $v$  from the lens, a pupil plane which is at a distance  $z_p$  on the object side of the lens, and two imaging pupils with separation distance  $P_g$ . The model is developed with the optical configuration set to  $v > v_0$ . This could equally have been developed with  $v < v_0$ . Similarly to the DPM, the depth ambiguity is overcome by applying the negative disparity conditions. The object point is on-axis and at a distance  $u_0$  from the lens with a conjugate image distance  $v_0$ . Additionally, like the DPM, the pupil plane generates a depth dependent image disparity based on the pupil gap and the sensor depth. Within the DPSM framework, the pupil plane is shifted  $z_p$  towards the object point. Axial shifts occur on the object side of the lens only. Axial shifts towards the image sensor would require knowledge of the refractive properties of the lens which are not desirable for a general camera model as outlined in Chapter 2.

On closer examination of the image disparity  $d$ , an increase in  $z_p$  will result in an observable increase in  $d$ . The increased disparity can also be attributed to an apparent increased pupil gap of  $E_p$  situated at the lens plane. Therefore, if an expression for the distance  $E_p$  is derived, the imaging system would be reduced to a pupil gap on an apparent pupil plane situated in the lens. This would allow the calculation of object depth based on the image disparity in a similar fashion to the DPM.

An expression for  $E_p$  can be developed based on the similar triangles formed by the rays emanating from the object point through the pupils, lens and eventually intersecting the image plane. The triangle with the base  $E_p$  and the object point at its apex is similar to the triangle that shares the same apex with  $P_g$  at its base. By comparing these similar triangles, an expression for the object depth,  $u_0$ , is formed based on the apparent pupil gap  $E_p$ , the actual pupil gap  $P_g$ , and the axial shift of the pupil plane  $z_p$ .

$$u_0 = \frac{-E_p z_p}{P_g - E_p} \quad (4.5)$$

Based on the similar triangles with  $d$  and  $E_p$  as a base and the conjugate image

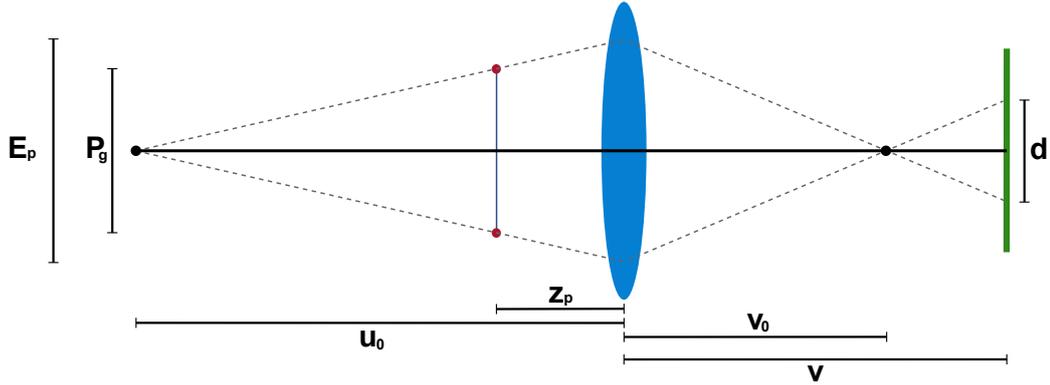


Figure 4.9: Double Pupil Shifted Model.

point as a shared apex, the following expression is derived for  $u_0$  (the DPM model).

$$u_0 = \frac{E_p v F}{E_p v - F(E_p + d)} \quad (4.6)$$

By setting Eqn. 4.5 equal to Eqn. 4.6 an expression for  $E_p$  is found in terms of  $F$ ,  $d$ ,  $P_g$ ,  $v$ , and  $z_p$ .

$$E_p = \frac{P_g v F - F d z_p}{F z_p + v F - v z_p} \quad (4.7)$$

This expression for  $E_p$  is substituted back into Eqn. 4.6 which yields the DPSM depth estimation.

$$u_0 = \frac{P_g v F - F d z_p}{P_g v - F(P_g + d)} \quad (4.8)$$

Therefore, similarly to the DPM, knowledge of the focal length, sensor depth, image disparity, and pupil gap, coupled with the new axial shift parameter allows object depth to be calculated. The DPSM is a natural extension of the DPM model as can be seen by setting the  $z_p$  parameter to zero, in which case the resulting expression is the depth equation for the DPM.

Input parameters to the DPSM can be classified as fixed or varying. The fixed input parameters are the focal length of the lens and the pupil gap separation distance. Typically these parameters are chosen off-line to achieve the most desirable disparity profile for object depth calculation. The identification and separation of object points emanating from particular pupils is achieved by adopting the same strategy as was applied to the DPM. The varying parameters are the sensor depth and the axial shift  $z_p$ . These parameters must be calculated without any prior knowledge, hence a calibration algorithm for performing such a task is proposed in Section 4.3.3. The simultaneous calibration

Table 4.4: DPSM Zemax verification results for various pupil plane shifts.

$u_0$ (mm)	Disparity (mm)			Estimated Depth (mm)		
	$z_p = 0mm$	$z_p = 25mm$	$z_p = 50mm$	$z_p = 0mm$	$z_p = 25mm$	$z_p = 50mm$
500	0	0	0	500	500	500
1500	2.7309	2.7772	2.8251	1500	1500	1500
2500	3.2771	3.3102	3.3440	2500	2500	2500
3500	3.5111	3.5364	3.5621	3500	3500	3500
4500	3.6412	3.6616	3.6821	4500	4500	4500
5500	3.7239	3.7410	3.7582	5500	5500	5500
6500	3.7812	3.7959	3.8106	6500	6500	6500

of  $v$  and  $z_p$  is an important result as it lends the DPSM model well to general optical configurations. This provides a flexible solution to acquiring accurate depth information from a multitude of multi-pupil imaging systems.

### 4.2.2 Simulated Experiments

Evaluation of the DPSM is carried out using synthetic data on two fronts. Initially the model validation is performed using Zemax. A full characterisation of the fixed and varying parameters is subsequently performed. The final experiment benchmarks the accuracy of the DPSM in retrieving depth information from the acquired images in the presence of varying levels of noise. Experiments in this section are conducted using a simulated camera with a pixel pitch of  $4.5\mu m$ , and a sensor length and width of  $18mm$  and  $13.5mm$  respectively. The simulated lens diameter is set to  $25mm$  and object depths are set in the range  $300mm$  to  $8,000mm$  in steps of  $50mm$ .

#### Zemax Verification

The DPSM is verified using industry standard optical design software in the form of Zemax. The optical system is configured as presented in Figure 4.10. A paraxial surface models the ideal lens and is set with a focal length of  $85mm$  and diameter  $25mm$ . A fixed pupil gap separation distance of  $20mm$  and a single near-focus setting are used in the simulations. The pupil plane is a custom surface containing the double pupil. This experiment is conducted with the pupil plane in three separate locations: with zero translation,  $25mm$  translation, and  $50mm$  translation. The results of the Zemax simulations are outlined in Table 4.4. The image disparity is calculated by measuring

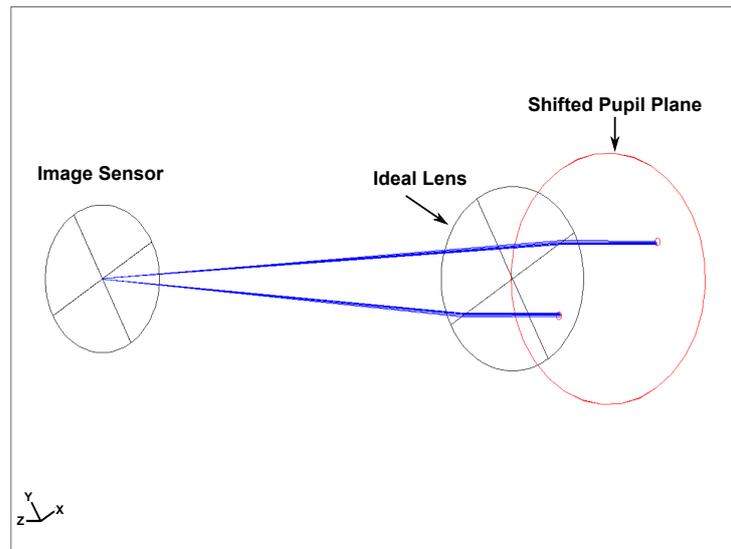


Figure 4.10: DPSM Zemax verification layout

the distance between the centroids of each object point's image through its respective pupil. With  $z_p = 0mm$  the DPSM is reduced to a DPM as can be verified from the results presented in Table 4.1. A magnification of the image disparities is observed as the pupil plane offset is increased to  $25mm$  and  $50mm$ . In each case the depth estimation is verified to be exact, within a noiseless environment, thus confirming the DPSM as a valid model.

### Model Characterisation

Sensitivity of the DPSM parameters is examined in two stages. Initially the fixed input parameters are analysed followed by a characterisation of the varying and calibrated parameters. The addition of the  $z_p$  parameter, in comparison to the DPM, plays a significant role in the behavior of the DPSM model. The same criteria applies to the DPSM analysis as applied to the DPM examination in terms of the practical constraints imposed by sensor size, lens diameter and magnification due to focal length. The resolutions to which the fixed input parameters are measured are outlined in Table 4.5. The measurement resolution of the focal length is in the order of microns therefore perturbations of this scale have little effect on the depth estimation accuracy

Analysis of the the pupil gap separation distance reveals that the addition of shifting the pupil plane has a multiplicative effect on the induced error in the depth estimation. The effect of increasing the pupil gap separation distance has

Table 4.5: DPSM Parameter measurement accuracy and sensitivity.

DPM Parameter	Measurement Accuracy	Depth Sensitivity
$F$	$10\mu m$	$\frac{(P_g v)^2 - P_g v d z_p}{(P_g v - F P_g - F d)^2}$
$P_g$	$7.8\mu m$	$\frac{-F^2 d (v + v z_p - z_p)}{(P_g v - F P_g - F d)^2}$
$v$	Calibration	$\frac{P_g F (d z_p - F d - P_g F)}{(P_g v - F P_g - F d)^2}$
$d$	Sub-pixel	$\frac{P_g F (F z_p - v z_p - F v)}{(P_g v - F P_g - F d)^2}$
$z_p$	Calibration	$\frac{F d}{(P_g v - F P_g - F d)}$

been discussed in the DPM analysis. The same concept applies for the DPSM, therefore larger separation distance induces less error in the depth estimation, and therefore it is set fixed at  $20mm$  for the  $P_g$  analysis. Varying the focal length has no effect on the perturbations of  $P_g$  with respect to  $u_0$ . This is a result of the disparity and sensor depth varying in proportion with changes in  $F$ . Therefore a fixed  $F$  of  $85mm$  is chosen. Figures 4.11(a)-(c) present the results of the sensitivity analysis of the  $P_g$  parameter for three axial shifts of the pupil plane. A zero shift is included for comparison with the DPM (Figure 4.11(a)), along with shifts of  $10mm$  and  $20mm$  in Figures 4.11(b)-(c) respectively. With the pupil plane situated at the lens, the DPM exhibited sensitive attributes to  $P_g$ , however these were of a small magnitude (less than 1% error at  $8m$ ). Shifting the pupil plane  $10mm$  from the lens induces errors which are an order of magnitude above the zero shift errors. This profile is also visible for a  $20mm$  shift of the pupil plane with a further increase of sensitivity to  $P_g$ . On closer inspection of the sensitivity profiles, it is evident that the near focused system induces the largest error in the depth estimation process. By examining the practical implications of optical configurations and the disparity profile of the system being modelled, the causes of this increased sensitivity can be rationalised. As the pupil plane shifts further into object space, the rays emanating from object points are projected through the lens at a greater distance from its centre. Therefore, error in the estimation of  $P_g$  for large shifts of the pupil plane into object space will result in larger disparity error on the image plane. Naturally, in a near-focus configuration the disparity profile is increased, therefore, the error is magnified for these configurations. The most significant result from the analysis of  $P_g$  is that a near focused optical system is not the ideal configuration for the DPSM for large depth ( $> 5m$ ) range estimation.

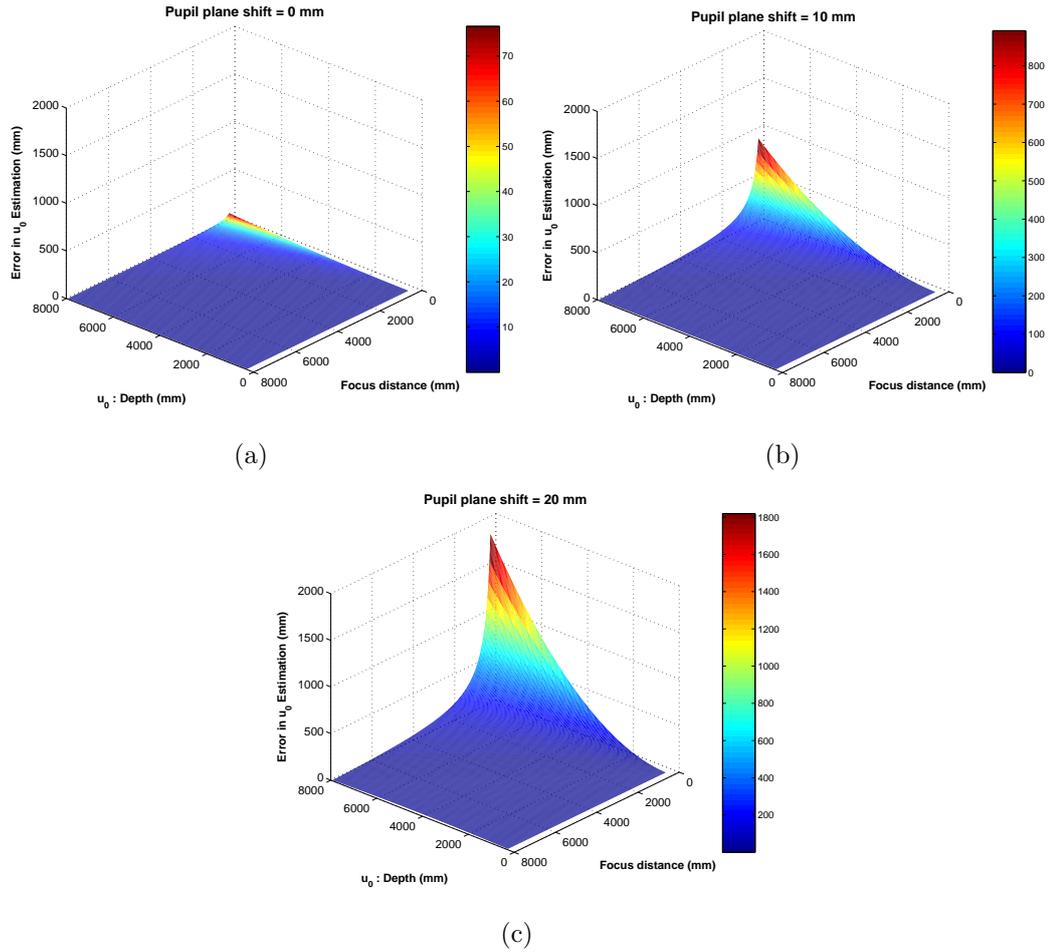


Figure 4.11: DPSM sensitivity of depth estimation to perturbations in  $P_g$  (a)  $z_p$ : 0mm; (b)  $z_p$ : 10mm; (c)  $z_p$ : 20mm.

The disparity profile of the DPSM given in Figures 4.12(a)-(c) represent pupil plane axial shifts of 0, 25mm and 50mm (zero shift equates to the DPM model). The conclusions drawn from the DPM analysis were that all configurations led to decreased sensitivity of the image disparities as the object depth increased. A near-focused configuration is favourable with the DPM due to its slightly increased sensitivity and broader range of disparities at larger object depths. Shifting the pupil plane axially changes the disparity profile of the nearer sensor depths. It is evident in Figures 4.12(b)-(c) that the profiles of the far-focused configurations are increasing in sensitivity, thus, a broadening of the image disparities is observed at the higher object depths. As a result, the disparity profile of the DPSM is not restricted to a near-focused configuration to achieve robust depth estimations. The DPSM image disparity profile in a near-focused configuration meets the same criteria as the DPM profile in terms of a broad magnitude of disparities at the larger object depths. However, in

light of the sensitivity issues of the pupil gap distance in these configurations, a mid to far focused setting is more desirable to avoid large errors in the depth estimation process. The increasing sensitivity of  $P_g$  to particular optical configurations is due to the magnitude of the error propagated through the system coupled with the sensor depth setting. As outlined, regardless of the sensor depth setting, there is reduced sensitivity of image disparities at large objects depths. Therefore, perturbations in  $P_g$  which result in large changes in the magnitude of image disparities, particularly at larger object depths, results in large depth estimation errors.

An imaging system in a near-focus configuration magnifies image disparity error due to  $P_g$  perturbations, particularly at large objects depths. This is a result of the rays passing through the system intersecting the optical axis at a larger distance from the image sensor. Large errors are not observed at the lower object depth estimates due to high sensitivity of the disparity profile in this region of object depths. In contrast, imaging systems in a far-focused configuration force incoming rays to cross the optical axis closer to the image sensor. Subsequently, perturbations in  $P_g$  result in disparity errors of similar scale to the DPSM disparity profile at larger object depths. Therefore, the perturbations in  $P_g$  propagate less error through the system which results in more accurate depth estimates.

The sensor depth parameter,  $v$ , and the pupil plane axial shift parameter,  $z_p$ , are both estimated via the calibration algorithm developed and presented in Sec. 4.3. The sensitivity analysis performed on the DPM in relation to the sensor depth parameter identified that changes in  $P_g$  had little effect on the depth estimation process. Similarly, it was shown if larger focal lengths were chosen, increased accuracy in the depth estimations process was achieved. Both of these results are equally applicable to the DPSM, therefore Figures 4.13(a)-(c) present the results for the sensitivity of the DPSM to  $v$  for a fixed  $F$  of  $85mm$  and a fixed  $P_g$  of  $20mm$  for three separate axial shifts ( $z_p : 10mm, 20mm, 30mm$ ). The perturbation in  $v$  remains constant across all axial shifts of the pupil plane. Thus,  $v$  is not sensitive to changes in  $z_p$  in the context of depth estimation. In each instance the maximum error induced in the depth estimation is  $25mm$ , at an object depth of  $8,000mm$ .

An expression for the sensitivity of the DPSM to perturbations in  $z_p$  is given in Table 4.5. The induced perturbation is in the order of millimeters. Similarly

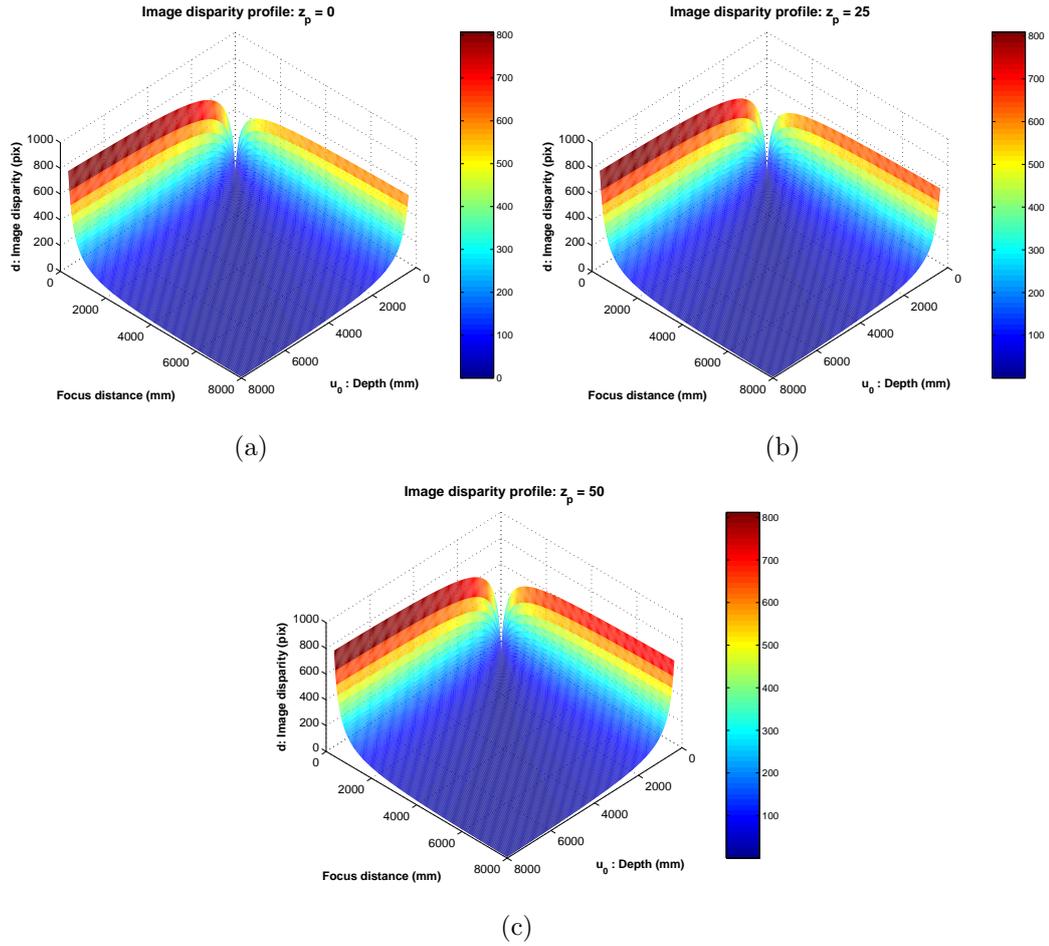


Figure 4.12: DPSM image disparity profile. (a)  $z_p$ : 0mm; (b)  $z_p$ : 25mm; (c)  $z_p$ : 50mm.

to the sensitivity of  $v$ , the pupil gap distance has a negligible effect on the depth estimation process in the presence of perturbations in  $z_p$ . Additionally, changes in focal length do not increase the depth estimation error. This is due to the proportional change in disparity with  $P_g$  and  $F$ . Therefore the input parameters of the simulations presented in Figures 4.14(a)-(c) are for a single focal length of 85mm and fixed pupil gap distance of 20mm. The only parameter which affects the sensitivity of the axial shift is the sensor depth. It is clear from the error profile that the near-focused configurations are performing worse than the mid to far focused settings. Albeit, the magnitude of the induced error which is roughly 25mm at 8m, remains constant across the range of axial shifts. Based on this analysis, choosing a near-focused configuration would not be severely affecting the overall depth estimation process. However, taking the sensitivity of the pupil gap distance into consideration, the DPSM is best suited to a medium or medium-long focus distance.

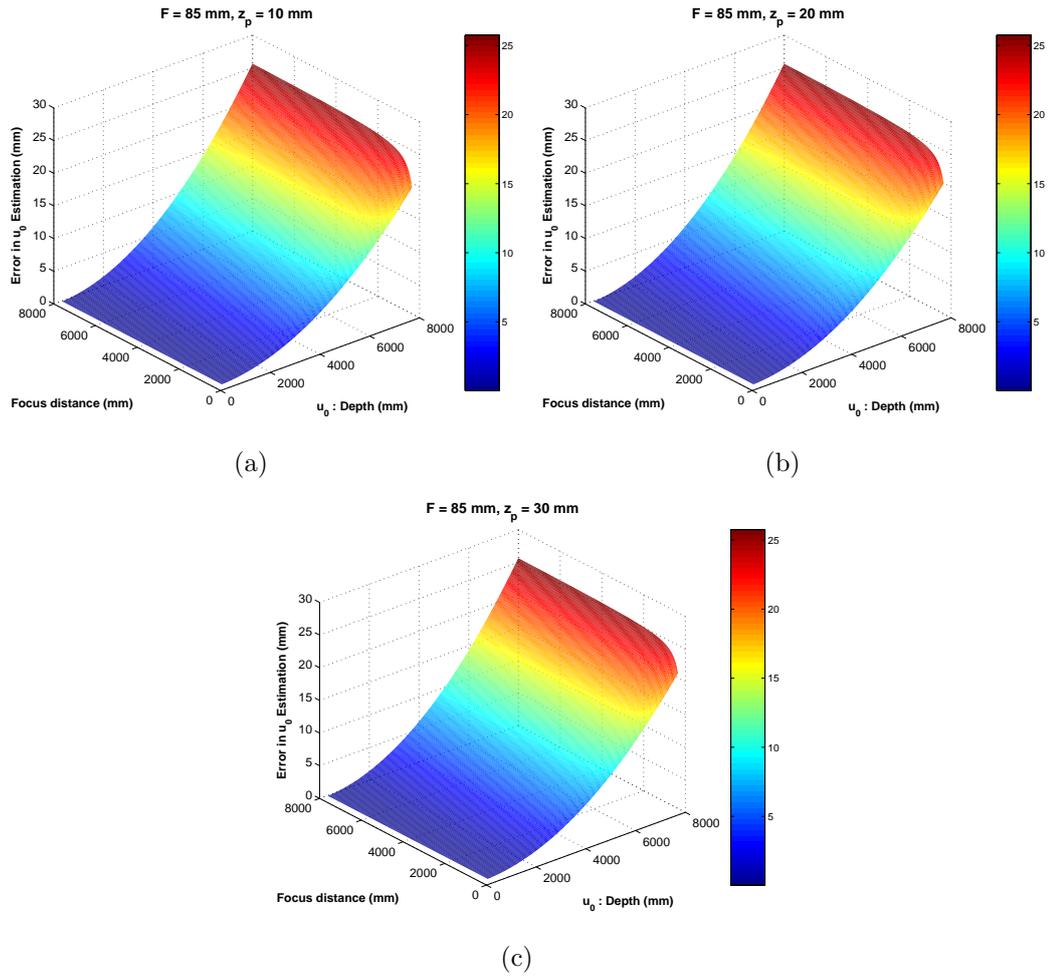


Figure 4.13: DPSM sensitivity to perturbations in  $v$ . (a)  $z_p$ : 10mm; (b)  $z_p$ : 20mm; (c)  $z_p$ : 30mm.

The characterisation of the DPSM leads to a number of conclusions. Firstly, if the axial shift of the pupil plane is set to zero, the model is equivalent to a DPM imaging system. The same physical limitations as apply to the DPM apply to the DPSM in terms of focal length magnification affecting sensor size and pupil gap distance within the diameter of the imaging lens. As the pupil plane is shifted linearly on the object side of the system, certain properties of the imaging system become very sensitive to perturbations. The pupil gap distance on the shifted pupil plane is one such parameter. As the pupil plane shift increases, a magnification of the depth estimation error is observed in near-focused configurations. It is this property which defines the ideal system properties of the DPSM. Even though the remaining DPSM properties are less sensitive in near focus configurations (with the exception of  $z_p$ ), the difference in the magnitude of induced errors for these parameters, in comparison to mid-far focused settings, are of a small magnitude in the overall system. Therefore

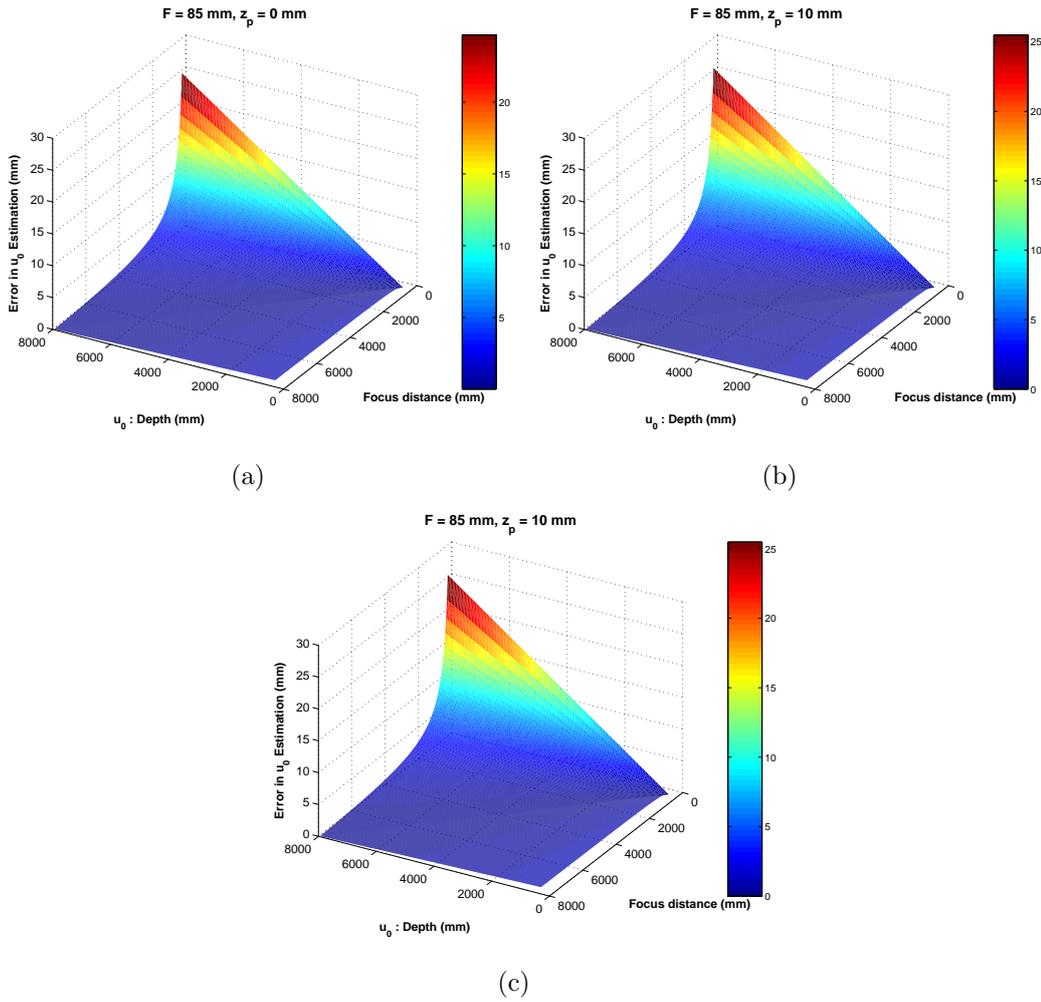


Figure 4.14: DPSM sensitivity to perturbations in  $z_p$ . (a)  $z_p: 0mm$ ; (b)  $z_p: 10mm$ ; (c)  $z_p: 20mm$ .

the ideal DPSM settings for achieving robust depth estimation are constrained by the focus distance. Similarly to the DPM, a maximised pupil gap distance within the lens diameter along with a mid range focal length are the ideal DPSM settings. In this configuration, the DPSM is well suited to accurately estimating depth over a large object distance range.

## Depth Estimation

Simulations are conducted to identify the accuracy of the depth estimation of the DPSM. Additive Gaussian pixel noise is introduced to the experiment to simulate perturbations in the input parameters and to evaluate the robustness of the DPSM. Normalisation is applied to the additive noise in the same manner as was applied in Section 4.1.2. Figures 4.15(a)-(f) represent the simu-

Table 4.6: DPSM depth estimation results at  $8,000mm$  across various input settings with additive Gaussian pixel noise (0-5 pixels SD).  $P_g$  is fixed at  $22.5mm$ .

Input Parameters	$F(mm)$	$35mm$						$85mm$						$200mm$					
	$z_p(mm)$	5		15		25		5		15		25		5		15		25	
	$v$ (focus)	Mid	Far	Mid	Far	Mid	Far	Mid	Far	Mid	Far	Mid	Far	Mid	Far	Mid	Far	Mid	Far
Depth Estimation	Mean (mm)	74.89	75.45	72.96	73.66	71.60	71.66	30.28	31.13	29.36	30.35	28.86	29.63	12.31	13.23	12.05	12.91	11.74	12.51
Error (mm)	SD (mm)	23.15	23.08	22.57	22.73	22.22	22.22	9.36	9.61	8.96	9.38	8.89	9.21	3.82	4.08	3.73	3.96	3.61	3.83

lations performed for a camera with focal length  $85mm$  and pupil gap distance of  $22.5mm$ . The plots in (a), (c), and (e) are the results for a mid-focused configuration while (b), (d), and (f) present a far-focused system. Near-focus configurations were not considered due to their high sensitivity to perturbations in the  $P_g$  parameter. Three pupil plane shifts have been examined for these camera settings: a  $5mm$ ,  $15mm$  and  $25mm$  shift. Pupil plane axial shifts larger than the diameter of the imaging lens are not simulated. Pupil plane shifts beyond  $25mm$  lead to axial rays emanating from the object points not passing through the lens. This is due to the angle at which the axial ray meets the shifted pupil plane.

The most significant result of these simulations is that shifting the pupil plane does not reduce the accuracy of the DPSM depth estimation. The maximum error observed across all shifts for an  $85mm$  focal length is  $60mm$  at an object depth of  $8,000mm$ . This equates to less than 1% error in the presence of a 5 pixel SD in noise. The DPSM characterisation identified a change in the disparity profile as the pupil plane was shifted axially. This observation is confirmed in the results presented in Figures 4.15(a)-(f). As  $z_p$  increases the far-focused configuration error reduces. This effect can also be seen in the error of the mid-focused setting. The reduction is a result of a broadening of the image disparity profile at the larger object depths.

Results for DPSM depth estimation, at an object depth of  $8m$ , for a number of input settings is given in Table 4.6. A single pupil gap distance of  $22.5mm$  was chosen for these tests. Results are given for three focal lengths ( $35mm$ ,  $85mm$ ,  $20mm$ ) and three pupil plane shifts ( $5mm$ ,  $15mm$ ,  $25mm$ ). As expected, the mean error and its SD improve as the focal length increases. This is due to the increased magnitude of image disparities which become less sensitive to pixel perturbations. A similar effect is observed with the increase in pupil plane shift. For each focal length, the best results are obtained with the maximum pupil plane shift. Overall, consistent depth estimates, with error less than 1%

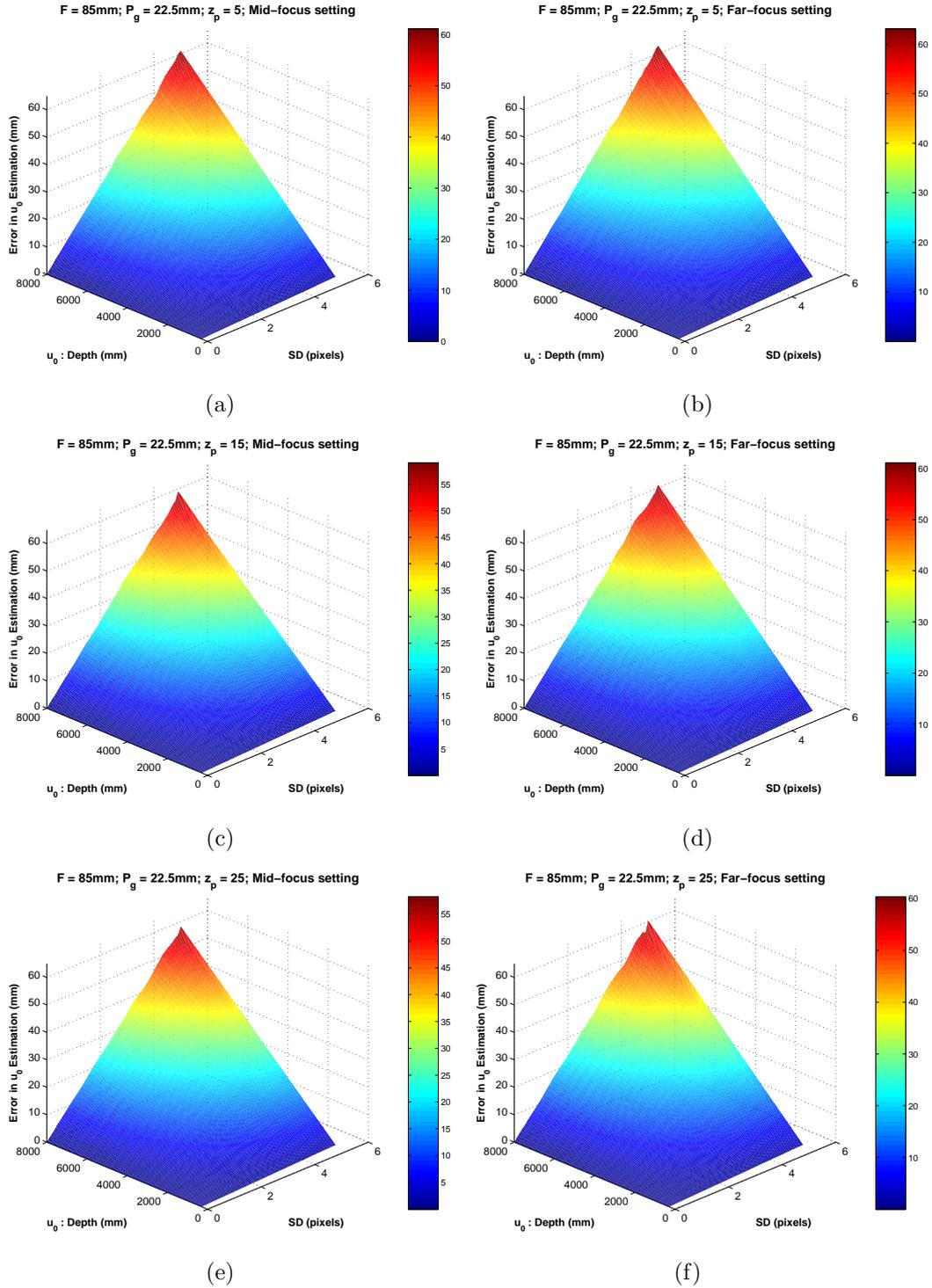


Figure 4.15: DPSM depth estimation in the presence of input pixel noise for various focus settings and pupil plane shifts. (a)  $z_p$ :  $0\text{mm}$ , mid-focus; (b)  $z_p$ :  $5\text{mm}$ , far-focus; (c)  $z_p$ :  $15\text{mm}$ , mid-focus; (d)  $z_p$ :  $15\text{mm}$ , far-focus; (e)  $z_p$ :  $25\text{mm}$ , mid-focus; (f)  $z_p$ :  $25\text{mm}$ , far-focus.

across all settings, are achieved at  $8m$  which is the most sensitive depth region.

This section has validated the DPSM camera model and tested its accuracy in estimating depth from a single image. The conducted experiments have shown that good accuracy is achieved across all pupil plane shifts and focal lengths of the imaging system. The pupil gap distance has been identified as a significant parameter as it induces a level of error to the overall system depending on the chosen optical configuration of the DPSM. Experiments show that mid to far focused configurations are favourable due to their suppressive effect on the induced pupil gap noise. The key benefit of using the DPSM camera model for depth estimation is its pupil plane shifting properties. This offers invaluable flexibility in optical system design.

### 4.3 Calibration of the Pupil Model

The calibration of a camera is the fundamental step to reconstructing Euclidean information from images. Standard pinhole calibration methods were introduced in Chapter 2 and an image network approach to calibrating pinhole cameras was proposed in Chapter 3. The applicability of standard pinhole calibration to the DPM and DPSM camera models is investigated in this section. However, there is a fundamental flaw in applying pinhole calibration methods to the double pupil models. Pinhole methods are only applicable to a (single) centrally located pupil on the pupil plane. Thus, in order to apply the pinhole calibration methodology to the DPM or DPSM, the pupil plane must be physically modified or removed to perform the calibration. Inherently, this introduces uncertainty to the accuracy of the calibration and subsequently to the accuracy of the depth estimation. Therefore, new calibration methods are proposed for the recovery of the sensor depth parameter for the DPM, and the DPSM camera models. Simulated testing is performed to compare and evaluate the new methods to the standard pinhole methods in recovering the sensor depth parameter.

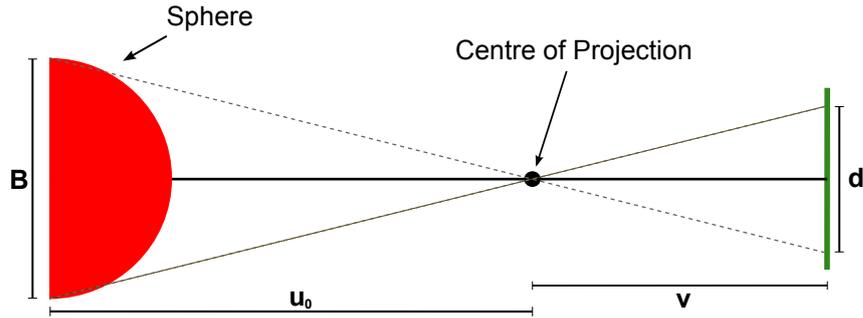


Figure 4.16: Calibration using a spherical object at a known distance.

### 4.3.1 Standard Pinhole Calibration Methods

The work presented on the characterisation of the DPM and DPSM has shown that  $v$  must be estimated accurately to achieve reliable depth estimates. When considering a calibration procedure to apply to the DPM and DPSM, it is the sensor depth parameter,  $v$ , which must be estimated most accurately. Standard pinhole calibration methods such as those proposed by Sturm and Maybank (1999) and Zhang (2000) involve imaging planar targets with varying pose while the work presented in Chapter 3 adds an additional step to these methods by specifying the planar pose of the targets for increased accuracy. In the case of planar calibration, all camera parameters are estimated. For the DPM and DPSM we only require the sensor depth since it is the only parameter needed to estimate object depth. Therefore, an alternative pinhole calibration method is also implemented in which only  $v$  is estimated based on imaging a spherical target at a known object distance.

The requirements for planar camera calibration have been well documented in Chapter 3. The spherical object calibration setup is shown in Figure 4.16. A spherical object of known diameter,  $B$ , is placed at a known distance  $u_0$  from the centre of projection. An image of the sphere is formed on the image plane where its diameter can be measured in pixels and subsequently converted to millimeters (the pixel to millimeter conversion is available from the camera manufacturer). Then by similar triangles, an expression for the sensor depth parameter,  $v$ , is formed.

$$v = \frac{du_0}{B} \quad (4.9)$$

Both pinhole calibration methods discussed in this section are valid to calibrate the DPM and DPSM if the pupil plane can be modified in the case of the DPM and, modified and shifted axially in the case of the DPSM. Simulated

testing is performed in this section to evaluate the applicability and accuracy of the pinhole calibration methods, under the ideal conditions of pupil plane placement and modification, to the DPM and DPSM camera models. Due to the nature of the physical constraints imposed by pinhole calibration methods, it is desirable that a calibration method, from a practical perspective, would not require the physical placement or movement of any internal components to the imaging system. This is the objective in forming the DPM and DPSM calibration methods.

### 4.3.2 DPM Calibration Method

The sensor depth parameter has been identified as an integral parameter to both multi-pupil imaging models. The physical constraints in applying standard pinhole methods to these models warrants consideration due to the sensitivity of the  $v$  parameter (as discussed in Section 4.1 and 4.2). Practicality must also take precedence when developing a new calibration algorithm. Less user input requirements equates to less error in the practical implementation of the calibration and thus less error in the  $v$  estimation. It is within these constraints that the DPM calibration algorithm is developed. Ideally the basic input requirements of the calibration algorithm should be a single image captured without the need for any modification to the pupil plane location or configuration. The parameters which are available off-line are the focal length,  $F$ , the pupil gap distance,  $P_g$ , and the pixel pitch which allows the conversion of image disparities from pixels to millimeters.

On examination of the DPM model (Eqn. 4.4) it is evident that with knowledge of the off-line parameters it is possible to form an expression for the sensor depth. However this expression contains the object depth which, ideally, should be independent of the calibration process. Therefore the proposed DPM calibration process makes use of two object depths in which only the translational component between both is known. This benefits the practitioner as there is no need to measure objects at various depths, in which experimental errors are likely to arise. It allows the use of accurate equipment such as a translation stage which allows the measurement of translational components to a resolution of  $0.1mm$ . Additionally this calibration process can be performed by capturing a single image that contains both object points. Furthermore, the pupil plane remains unmodified in the proposed method, resulting in a more

natural calibration process for the DPM in comparison to applying pinhole techniques.

The proposed calibration setup is presented in Figure 4.17. The DPM imaging system is defined with pupil gap distance  $P_g$  for a lens of given focal length  $F$ . The sensor depth,  $v$ , can be set to any distance for the calibration process. A large sensor depth, which sets the system to a near focus configuration, has been shown to be ideal for accuracy in the depth estimation (see Figures 4.6(a)-(c)). Therefore the DPM should be configured in a near focus configuration for the calibration procedure. A calibration target is set up which consists of two feature points separated along the optical axis by a known distance  $u_z$ . In this configuration, the DPM images both object points as image disparities,  $d_0$  and  $d_1$ . By using the off-line knowledge of focal length, pupil gap distance, calibration target properties, and pixel pitch along with the image disparities  $d_0$  and  $d_1$ , an expression can be developed for the sensor depth parameter. In its basic form the relationship can be verified as  $u_1 = u_0 + u_z$  where  $u_0$  and  $u_1$  correspond to both calibration target object depths. Substituting the expression for DPM depth (Eqn. 4.4) into this relationship yields

$$\begin{aligned} \frac{P_g v F}{P_g v - F P_g + F d_0} + u_z &= \frac{P_g v F}{P_g v - F P_g + F d_1} \\ \frac{P_g v F + u_z (P_g v - F P_g - F d_0)}{P_g v - F P_g + F d_0} &= \frac{P_g v F}{P_g v - F P_g + F d_1} \end{aligned} \quad (4.10)$$

which is an expression independent of object depths and dependent on the known imaging system parameters and on the sensor depth. By further manipulation, an expression for the sensor depth can be found in the form of a polynomial equation.

$$v^2 \alpha + v \beta + \gamma = 0 \quad (4.11)$$

where

$$\begin{aligned} \alpha &= P_g^2 \\ \beta &= F^2 P_g (d_0 - d_1 - 2u_z P_g) - F P_g u_z (d_0 + d_1) \\ \gamma &= F^2 u_z (P_g^2 + P_g d_0 + P_g d_1 + d_0 d_1) \end{aligned} \quad (4.12)$$

A solution for the sensor depth  $v$  can be found from Eqn. 4.11. As it is a second order polynomial, two solutions for  $v$  are found. However, a basic test

on both solutions confirms that only one value is in fact a valid solution. The bounds on the sensor depth test are that the minimum value for  $v$  must be no less than the given focal length. This equates to an imaging system that is at least in a configuration focused at infinity. The calibration equations developed in this section are formed with an imaging system in a near focused configuration i.e.  $v > v_0$ . The DPM calibration algorithm remains consistent across all optical configurations as long as the negative disparity condition is upheld (see Section 4.1.1). As a result, the imaging system can be set in any optical configuration (within the above bounds) for the calibration process.

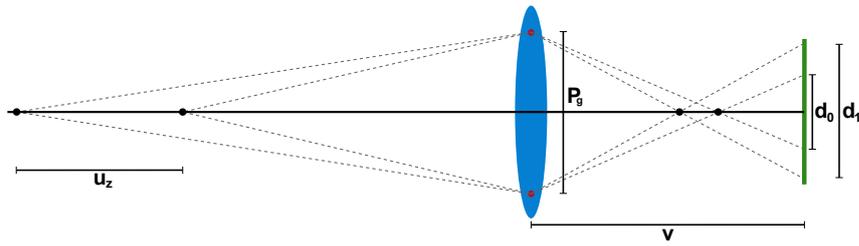


Figure 4.17: DPM calibration setup.

### Simulated Experiments

Simulated experiments are performed to validate the DPM calibration method. The first experiment examines the DPM calibration setup in terms of varying target object depths and thus the translation between both object points. The second experiment compares the DPM calibration method to the standard pinhole methods. Evaluation is quantified in terms of accuracy in  $v$  estimation and accuracy in the corresponding estimated object depths.

The results for varying the calibration target properties, in terms of translation between objects at various depths, are presented in Figures 4.18(a)-(f). The simulated camera has an imaging sensor of dimensions  $18mm \times 13.5mm$  in a near focused configuration in a system of focal length  $85mm$  with pupil gap  $20mm$ . Object depths are set in the range  $300mm$  to  $8,000mm$  in steps of  $50mm$ . The translational component  $u_z$  is set in the range  $100mm$  to  $2,000mm$  in steps of  $10mm$ . Each experiment takes an object depth within the range as the initial DPM calibration target. For each initial target, a translational component  $u_z$  is added to generate the second object target. Subsequently, the calibration is performed and  $v$  is estimated (Figures 4.18(a), (c), and (e))

along with an estimation of actual object depth across the entire range using the calibrated  $v$  (Figures 4.18(b), (d), and (f)). Three levels of pixel noise are added to the image disparities to simulate image noise and evaluate the robustness of the DPM calibration.

As expected, in areas where  $v$  is poorly estimated, the object depth estimation is similarly inaccurate. This is in agreement with the analysis of the DPM in Section 4.1.2. The areas which exhibit most sensitivity are those where small translational components are applied at medium to long object depths. These cases induce large error in  $v$  and subsequently object depth due to the nature of the disparity profile of the DPM. Since the medium to long range depth corresponds to the most sensitive area of the DPM disparity profile, any error in measurement of the image disparity will be magnified and propagated through the DPM system. Figures 4.18(b), (d) and (f) show how this error is magnified as the SD of the pixel noise is increased. Therefore, careful consideration should be given to choosing the calibration target translational component and to a lesser extent, the initial calibration target depth. There is a large area of robust estimation as the translational component is increased coupled with a general increase in object depth. Calibrations performed under these conditions resulted in the most accurate depth estimates even in the presence of large amounts of noise. This trend in the calibration process was observed across all focal settings and optical configurations. In a similar manner to the DPM disparity profile, slightly more accurate results are achieved in near focused configurations as opposed to far-focused configurations.

The second experiment examines how applicable pinhole calibration methods are to the DPM and examines how they compare in terms of accuracy with the DPM calibration method. The same simulated camera was used for each calibration method to ensure a fair comparison was achieved. The camera has a lens with focal length  $85mm$  and sensor size  $18mm \times 13.5mm$ . The optical system was set to a near focused configuration at  $300mm$  thus yielding a large sensor depth for calibration in each method. The methodology of the pinhole approach is to initially calibrate with a single centred pupil on the pupil plane. This calibrated sensor depth is then used to calculate depth estimated from a DPM imaging system employing a double pupil with a  $P_g$  of  $20mm$ . Gaussian pixel noise with varying SD between 0.1 and 5 pixels is added to the calibration images, before the feature extraction process, to evaluate the robustness of each calibration method. In order to ensure that the additive noise equally

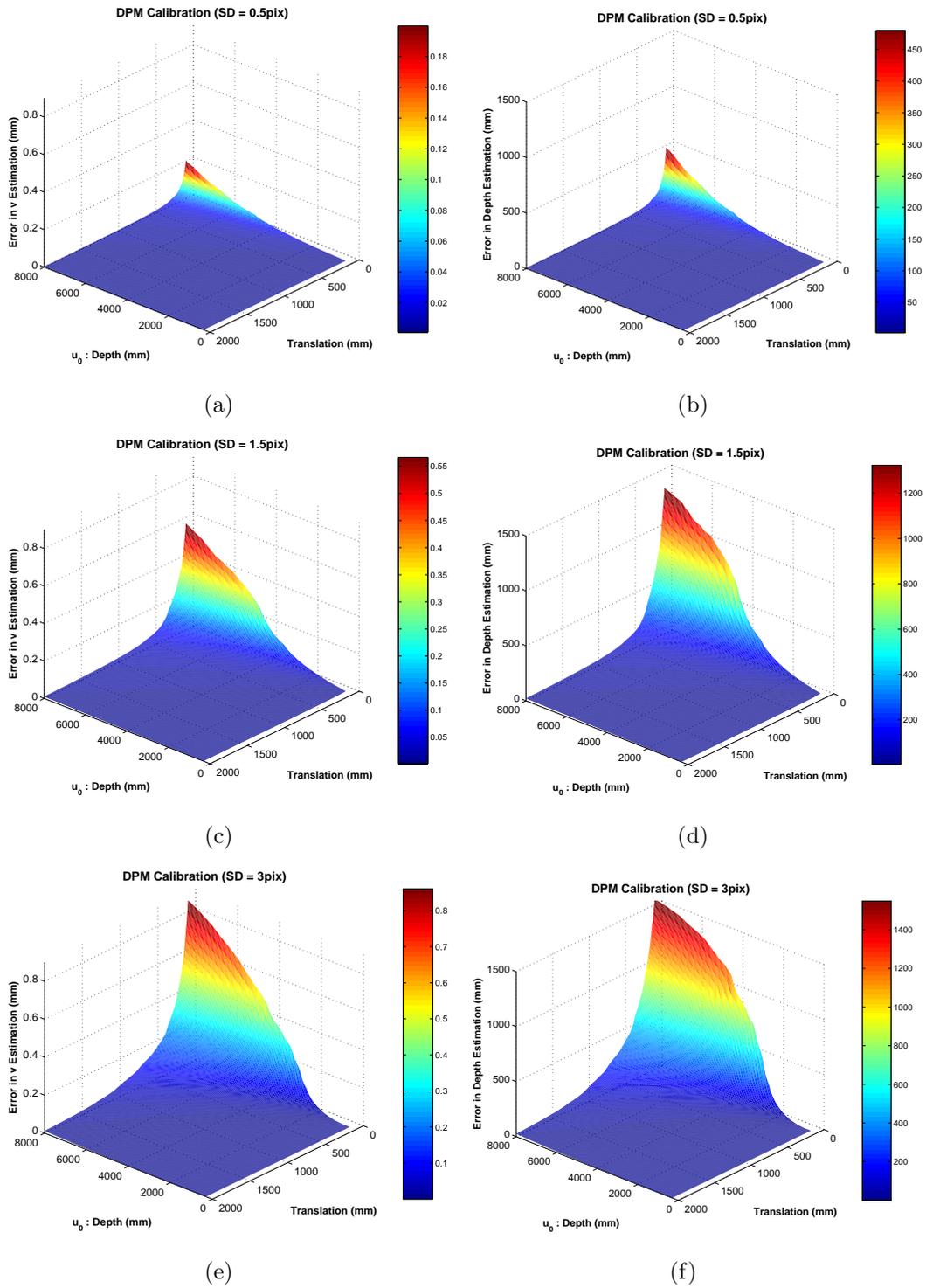


Figure 4.18: DPM Calibration accuracy in sensor depth along with the corresponding depth estimation accuracy error. (a) Accuracy of  $v$  estimation SD = 0.5 pixels ; (b) Accuracy of depth estimation SD = 0.5 pixels; (c) Accuracy of  $v$  estimation SD = 1.5 pixels ; (d) Accuracy of depth estimation SD = 1.5 pixels; (e) Accuracy of  $v$  estimation SD = 3 pixels ; (f) Accuracy of depth estimation SD = 3 pixels

affects each calibration method, the calibration targets are synthesised to cover the same area of the image sensor in each experiment. The first pinhole calibration method employed is that of Zhang (2000) with the image selection strategy for a two IN as outlined in Chapter 3. The second method applied is the pinhole calibration via a spherical object. The input parameters for the DPM calibration are guided by the simulated experimentation conducted and presented in Figures 4.18(a)-(f) so that the areas of highest sensitivity are avoided. That is, calibration targets located greater than 5m from the camera. A translational distance of  $1,000mm$  is chosen for a target at an initial depth of  $1,600mm$ . For each level of noise introduced to the imaging system, 100 trials with random noise at this level are undertaken, and the mean calibrated sensor depth is calculated. The calibrated sensor depth values are subsequently used to estimate object depths in the range  $300mm$  to  $8,000mm$ , which are used to quantify the accuracy of the calibration.

The results in Figures 4.19(a)-(c) present the mean error across all estimated object depths for each level of noise introduced into the imaging system. The SD of the depth error is given at each level as a bar plot. On closer examination of the results, the magnitude of the error for planar calibration ( $417mm$ ; SD of  $298mm$ ) is far greater than the error in the spherical and DPM methods. The proposed DPM calibration outperforms the spherical method with a maximum mean error across all depths of  $31.25mm$  (SD  $25mm$ ) compared to  $83.6mm$  (SD  $62mm$ ) for the spherical method. A more detailed experiment is conducted for each calibration method which presents the mean error at each object depth. The results for this experiment are presented in Figures 4.20(a)-(c) which correspond to the planar, spherical, and DPM methods respectively. This experiment highlights the robustness of the proposed DPM method in comparison to the planar and spherical methods. The magnitude of the error in the depth estimation of the planar method indicates that the calibration of the sensor depth is not of adequate accuracy to be suitable for the DPM camera model. Improvement in the planar accuracy could be achieved by increasing the number of input images to the calibration process, however this would greatly increase the input requirements of the user and lessen the flexibility of the calibration procedure. There is a large improvement with the spherical calibration in comparison to the planar approach. The maximum depth estimation error is  $\sim 230mm$  at an object depth of  $8,000mm$  compared to  $\sim 1,200mm$  with the planar method at the same object depth. The primary

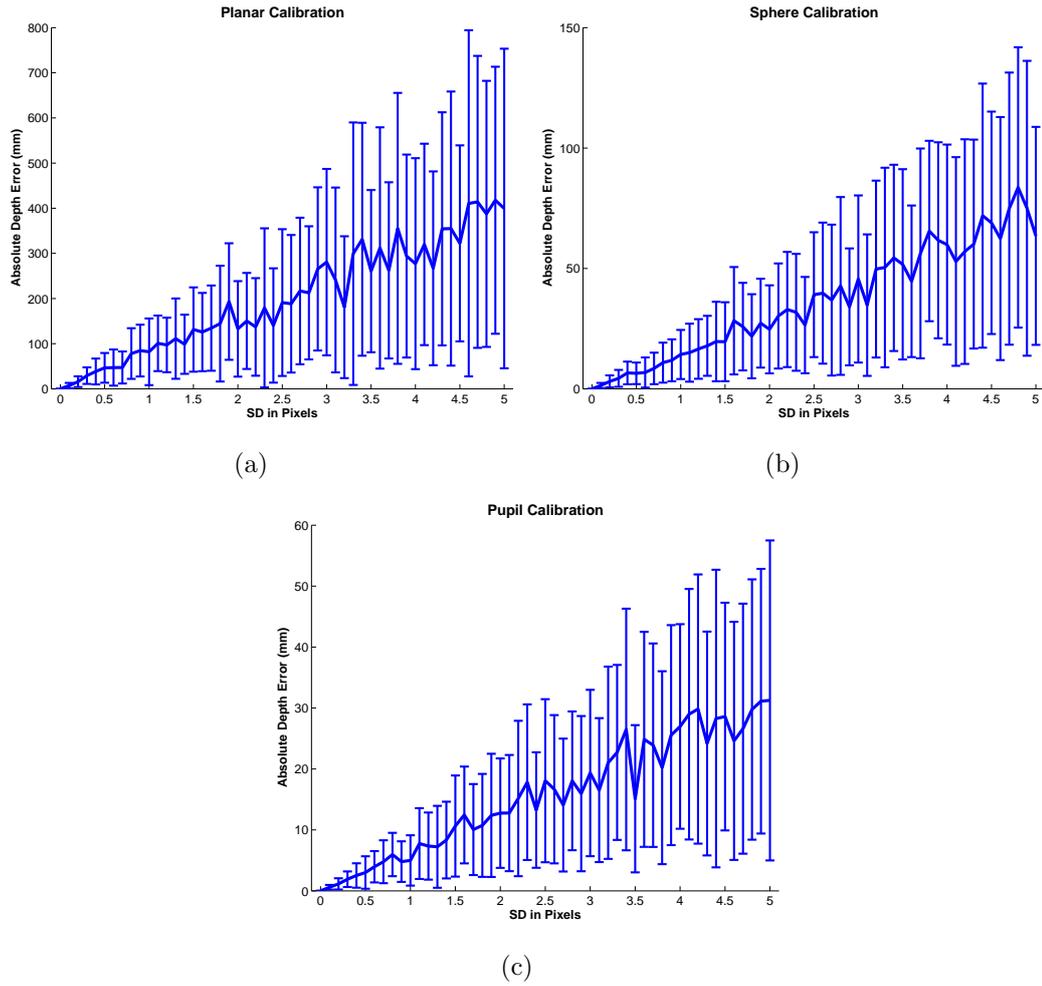


Figure 4.19: DPM Calibration method comparison. (a) Planar calibration; (b) Sphere pinhole calibration; (c) Proposed DPM calibration.

drawback of the spherical approach is that exact measurements are required for object distance as well as an accurate spherical calibration target. The depth estimation error observed as a result of the proposed DPM calibration is  $\sim 90\text{mm}$  at  $8,000\text{mm}$ . One aspect that was not simulated in these experiments was the additional noise element that is applicable to the pinhole calibration techniques as a result of pupil plane movement and replacement. It is expected that this would induce further error.

The conducted experiments in this section have examined the applicability of standard pinhole calibration methods to the proposed DPM camera model. It is shown that these methods lack the accuracy in sensor depth determination to reliably estimate depth in a DPM imaging framework. The proposed DPM calibration algorithm is shown to outperform the standard methods and achieve a high degree of accuracy in depth estimation ( $\sim 1\%$  at  $8,000\text{mm}$ ).

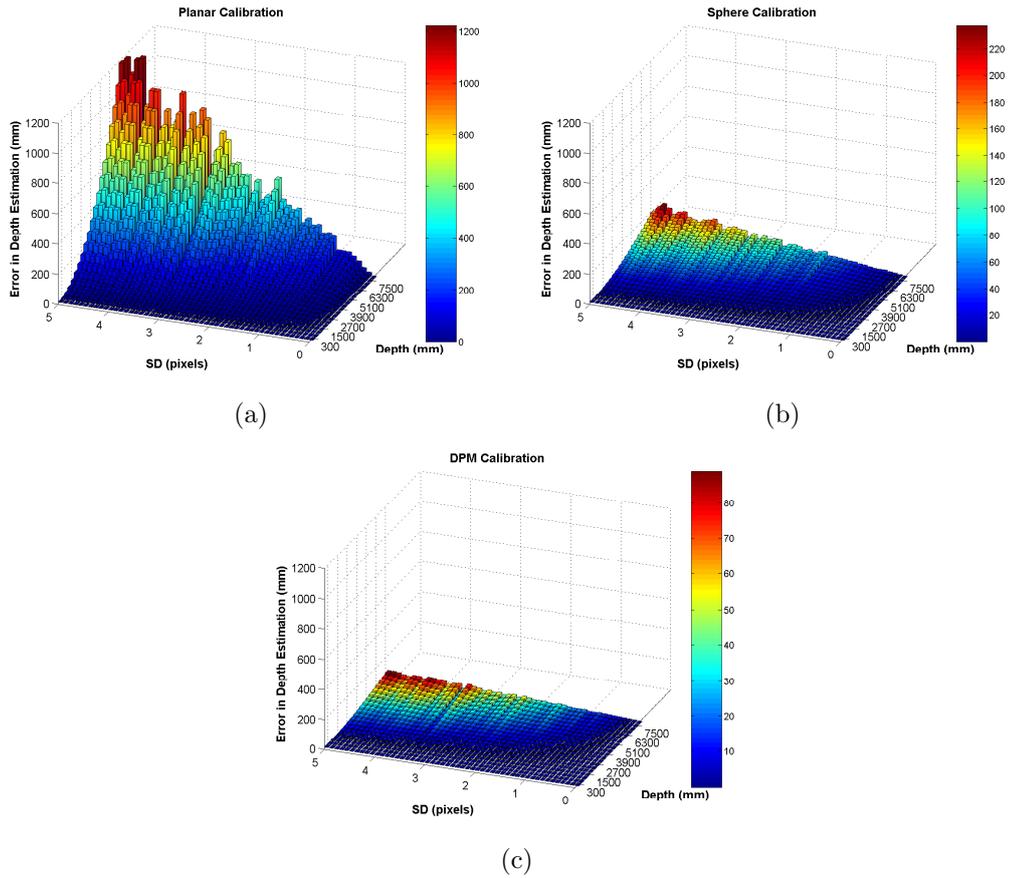


Figure 4.20: Proposed DPM calibration method comparison - depth recovery. (a) Planar calibration; (b) Sphere pinhole calibration; (c) Proposed DPM calibration.

The practical aspect of performing the calibration procedure also favours the DPM approach as it requires less user interaction. Additionally, as it derives naturally from the DPM imaging model, it does not require any modification of the pupil plane location, which is a major source of error in applying the standard pinhole methods.

### 4.3.3 DPSM Calibration Method

The DPSM model differs from the DPM in that it contains an additional model parameter which represents the axial shift of the pupil plane,  $z_p$ . Thus the calibration method developed for the DPM is not applicable to the DPSM as there is an additional parameter that requires calibration. However, by examining the DPSM model equation (Eqn. 4.8), a calibration algorithm can be developed which adopts a similar approach to the DPM calibration procedure

which can calibrate both the sensor depth and the pupil plane axial shift.

As was the case with the DPM calibration, the off-line parameters of the DPSM model in Eqn. 4.8 are the focal length, the pupil gap distance, and the pixel pitch. By choosing a calibration target which consists of two object points separated by a known axial distance ( $u_z$ ), the DPSM equation can be rearranged to form an expression which encapsulates the difference in image disparity due to the axial shift of the object points.

$$\begin{aligned} \frac{P_g v F - F d_0 z_p}{P_g v - F P_g + F d_0} + u_z &= \frac{P_g v F - F z_p d_1}{P_g v - F P_g + F d_1} \\ \frac{P_g v F - F d_0 z_p + u_z (P_g v - F P_g - F d_0)}{P_g v - F P_g + F d_0} &= \frac{P_g v F - F z_p d_1}{P_g v - F P_g + F d_1} \end{aligned} \quad (4.13)$$

where  $d_0$  and  $d_1$  are the image disparities generated by two object points separated axially by the distance  $u_z$ . This expression is further refined into polynomial form

$$\alpha_0 v^2 + \alpha_1 v + \alpha_2 z_p + \alpha_3 v z_p = \gamma_0 \quad (4.14)$$

where

$$\begin{aligned} \alpha_0 &= u_z P_g^2 \\ \alpha_1 &= F P_g (F d_0 - F d_1 - 2 U_z P_g - u_z d_0 - u_z d_1) \\ \alpha_2 &= F^2 P_g (d_0 - d_1) \\ \alpha_3 &= F P_g (d_1 - d_0) \\ \gamma_0 &= u_z F^2 (P_g^2 + P_g d_0 + P_g d_1 + d_0 d_1) \end{aligned} \quad (4.15)$$

As there are two parameters which require estimation, the single expression in Eqn. 4.14 will not suffice to solve both parameters.

Therefore multiple object points translated axially will be required to form multiple equations to solve the calibration parameters.

$$\sum_{i=1}^n \alpha_{i0} v^2 + \alpha_{i1} v + \alpha_{i2} z_p + \alpha_{i3} v z_p = \gamma_i \quad (4.16)$$

where  $n$  is the number of axial translations between object points. In order to uniquely solve the DPSM calibration parameters ( $v$ ,  $z_p$ ), three equations

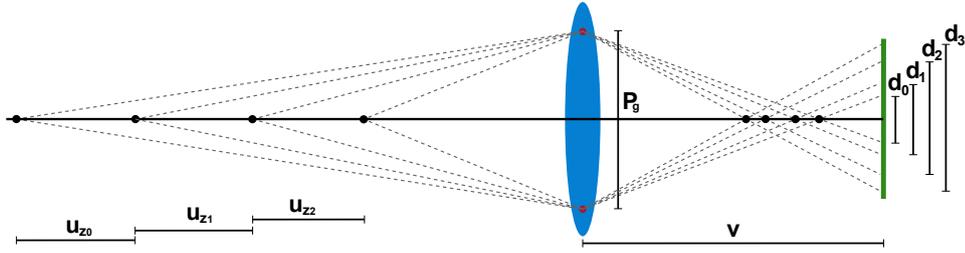


Figure 4.21: Proposed DPSM calibration setup.

are required ( $n = 3$ ). Each new equation,  $i$ , in the system must contribute independently to the overall system of equations to uniquely solve the calibration parameters. Consequently, the calibration setup required for the DPSM camera model is presented in Figure 4.21. It consists of four distinct object points which gives three axial shifts and thus yields three equations. Each pair of disparities, for example  $d_0$  and  $d_1$ , have a corresponding axial shift ( $u_{z0}$ ) and form a new equation in the calibration system. Naturally, the number of object points and resulting axial shifts can be increased to generate a larger system of equations to solve the calibration parameters. Given that the calibration equations are non-linear, a non-linear minimisation technique, in the form of the Levenberg-Marquardt (LM) algorithm, is applied to solve the DPSM calibration equations. This enables the calculation of the sensor depth and pupil plane axial shift parameters. The function to be minimised by the algorithm,  $\mathfrak{F}(v, z_p)$  is,

$$\mathfrak{F}(v, z_p) = \sum_{i=1}^n \alpha_{i0} \hat{v}^2 + \alpha_{i1} \hat{v} + \alpha_{i2} \hat{z}_p + \alpha_{i3} \hat{v} \hat{z}_p - \gamma_i \quad (4.17)$$

which generally converges to a solution for  $v$  and  $z_p$  within 10 – 15 iterations. The proposed DPM method is used as an initial estimate for  $v$ , while the initialisation for  $z_p$  is set to zero.

The likelihood of degeneracies arising during the parameter estimation process is an important consideration for any calibration algorithm and requires investigation. The most obvious case of degeneracy is when there is dependency between the calibration equations within the system. As shown in Figure 4.21, four distinct object points are chosen for the calibration procedure. If a subset of three of these object points were chosen, for example  $(d_0, d_1, d_2)$ , the third equation would have to be generated via object points  $d_0$  and  $d_2$ . This inherently creates a dependency in the calibration equations which leads to failure in the parameter estimation process. Therefore the minimum requirements of the DPSM calibration setup is that there are four distinct object points to

generate the three axial shifts and subsequent equations. Further examination of Eqn. 4.17 indicates that there is a camera configuration which leads to a degenerate solution for the DPSM calibration process. If the imaging system is set to focus at infinity, then  $v = F$ , and the pupil plane axial shift parameter,  $z_p$ , will be eliminated from the calibration equation. This results in failure to estimate the  $z_p$  parameter and thus a failure of the calibration process. Consequently, the sensor depth parameter must not be set to focus at infinity, which is easily avoided when calibrating an imaging system. Additionally, the analysis of Section 4.2.1 led to the conclusion that a mid-far focus configuration was best suited to the DPSM model for estimating depth across a large range. Therefore the optimal calibration settings require that a focused at infinity configuration be avoided.

### Simulated Experiments

Two synthetic experiments are conducted to validate the proposed DPSM calibration method. The first experiment tests the accuracy of the DPSM calibration algorithm in estimating the sensor depth and pupil plane axial shift parameters. This test is conducted with varying input parameters and in the presence of varying input noise to the imaging system. The second experiment evaluates the accuracy of the DPSM calibration algorithm for various input settings. Accuracy is assessed by examining error in the depth estimation using the calibrated parameters across a large object depth range. Experiments are carried out using both a near focused and mid-far focused imaging configuration ( $300mm$  and  $3,000mm$  respectively). The object depth range is set between  $300mm$  and  $8,000mm$ . A pupil gap separation distance of  $20mm$  is chosen (lens diameter is  $25mm$ ) and the focal length of the imaging system is  $85mm$ .

The DPSM calibration setup requires four distinct object depths to generate three axial distances and the subsequent equations for estimating the camera model parameters. The choice for calibration target object distances builds on the experimentation conducted for the DPM calibration procedure in Section 4.3.2. These results (Figures 4.18(b), (d), and (f)) highlighted that once the translational component was large enough (roughly a quarter the object distance) at any object distance, the physical setup of calibration targets was sufficient to obtain an accurate parameter estimation. The addition of the  $z_p$

parameter to the DPSM calibration means that two extra object distances are required, which can be chosen based on the analysis in Figures 4.18(b), (d), and (f). The only effect  $z_p$  has on these error profiles is a broadening of the pixel disparities which marginally reduces the error of the profiles in Figures 4.18(b), (d), and (f). The object distances chosen for the calibration targets in both DPSM experiments are  $500mm$ ,  $1,000mm$ ,  $1,500mm$ , and  $2,000mm$ .

The results for sensor depth and pupil plane axial translation calibration for an imaging system in a near-focus configuration are presented in Figures 4.22(a), (b), and (c). Each point on the error surfaces represents the mean error recorded, for 100 experiments, at a particular  $z_p$  in the presence of noise (with a SD) across all object depths. The sensor depth calibration results are given in Figure 4.22(a). The maximum mean error in  $v$  across all depths is  $40\mu m$  which is observed with the system under a noise level of 5 pixels SD. Results for the estimation of  $z_p$  are shown in Figure 4.22(b). The maximum error is recorded at the translational components of lesser magnitude and is approximately  $90\mu m$  (at  $SD = 5$  pixels). The resulting error in depth estimation using the calibrated parameters is given in Figure 4.22(c). As the SD of the pixel noise is increased, the error increases as expected. An error of  $90mm$  is the maximum depth error observed and corresponds to the least accurate estimation of the DPSM parameters. Results for the same tests with an imaging system in a mid-far focused configuration are given in Figures 4.23(a), (b), and (c). The main difference with this imaging configuration is that  $v$  is estimated marginally more accurately ( $\sim 30 - 33\mu m$  error) while  $z_p$  is less accurate. As the pupil plane axial translation is decreased towards its natural position, at the lens, the imaging system becomes closer to a DPM than the DPSM. Consequently, a mid-far focused system (in a DPM configuration) is more sensitive to the induced noise in the imaging system. Thus the  $z_p$  estimation is less accurate, by approximately an order of magnitude, than the near-focused configuration. However the increase in accuracy of the  $v$  estimation offsets the additional induced error of  $z_p$  as  $v$  is a considerably more sensitive parameter in the model when estimating depth. Thus the estimation of  $z_p$  in the mid-far focus configuration leads to a marginal increase in the overall error in estimated depth. The maximum depth error observed in this configuration is approximately  $20mm$  greater than in the near-focus configuration.

The second simulated experiment in this section evaluates the accuracy of the DPSM calibration for varying input parameters and quantifies the results based

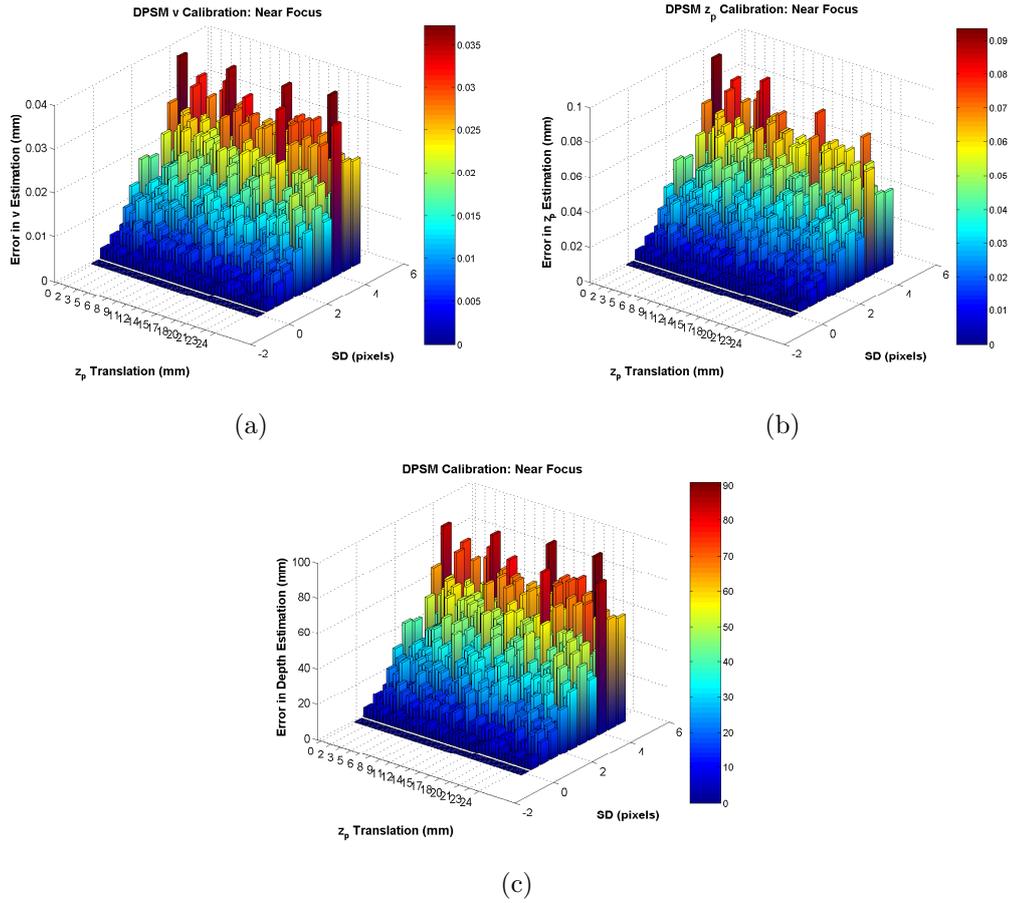


Figure 4.22: DPSM Calibration - parameter estimates in a near-focus configuration. (a) Sensor depth error; (b) Pupil plane offset error; (c) Mean depth error.

on the estimated depth at each object depth. Experiments are carried out for both a near focus and mid-far focused configuration. Three pupil plane axial translations are simulated ( $5mm$ ,  $15mm$ , and  $20mm$ ). Gaussian noise with SD varying between 0.1 and 5 pixels is added to the imaging system to test the robustness of the calibration algorithm. Each data point in the results is the mean depth error of 25 calibrations at that particular level of noise for the given camera settings. The mean estimated depth error at each object depth is presented in Figures 4.24(a), (c), and (e) for the near focus configuration and Figures 4.24(b), (d), and (f) for the mid-far focused system. As expected the least accurate depth estimates occur when the noise level is above 4 pixels SD when estimating object depths beyond  $3,000 - 4,000mm$ . The maximum of the mean errors is approximately  $95mm$ . This error corresponds to a short pupil plane axial translation ( $z_p = 5mm$ ) in a mid-far focus configuration. This error is in agreement with the observation that in a mid-far focused

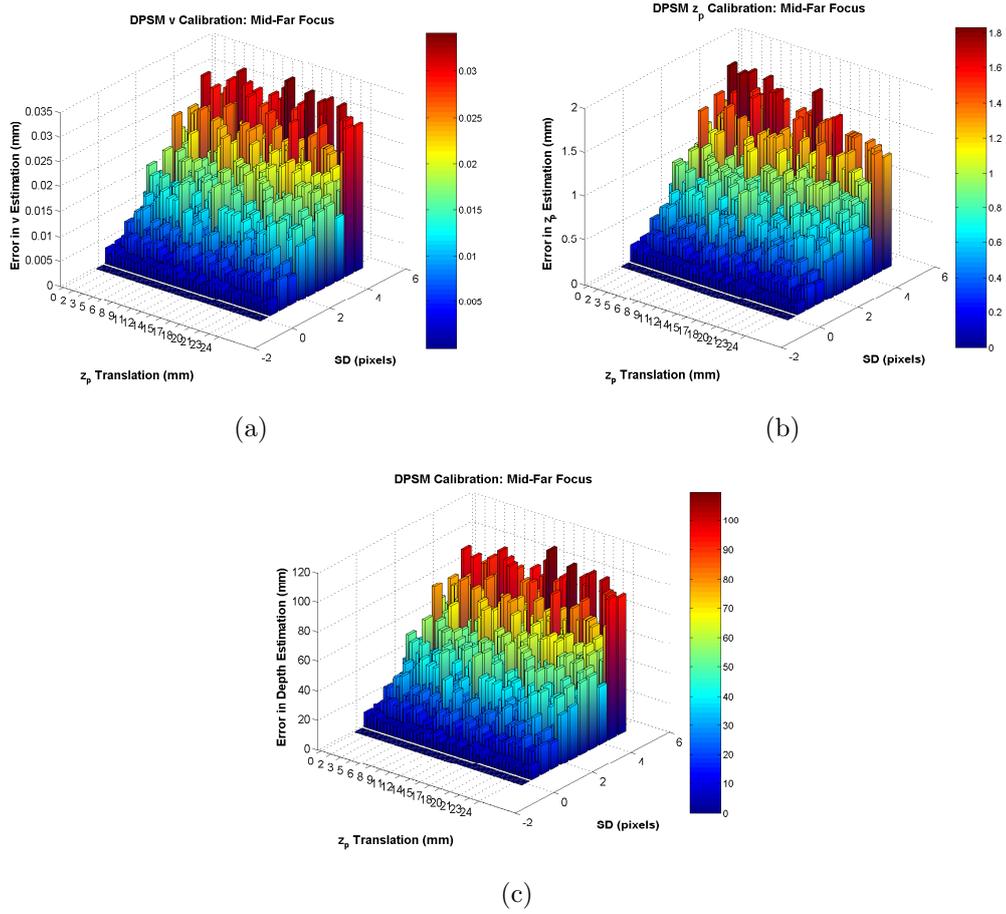


Figure 4.23: DPSM Calibration - parameter estimates in a mid-far focus configuration. (a) Sensor depth error; (b) Pupil plane offset error; (c) Mean depth error.

configuration, the imaging system is more sensitive as it tends towards a DPM (as  $z_p$  decreases). On the other hand, the simulations with a near focus system are less sensitive to the induced noise at  $z_p$ s of smaller magnitudes. As the magnitude of the pupil plane axial shift increases, an improvement in the accuracy of the depth estimation is seen with the mid-far focus configuration. The maximum of the mean error for the mid-far focused system with  $z_p = 20\text{mm}$  is less than  $80\text{mm}$  in comparison with the near focused system for the same settings which has error greater than  $80\text{mm}$ . As  $z_p$  increases, the general trend is that the mid-focus system decreases in error of depth estimates while the near focused system increases. This also confirms the findings in Section. 4.2.1 which showed that the DPSM parameters were more sensitive to perturbations in the image data for near focused systems in comparison to the mid-far focused systems. The corresponding mean and SD across all object depths for the plots in Figures 4.24(a)-(f) are given in Figures 4.25(a)-(f).

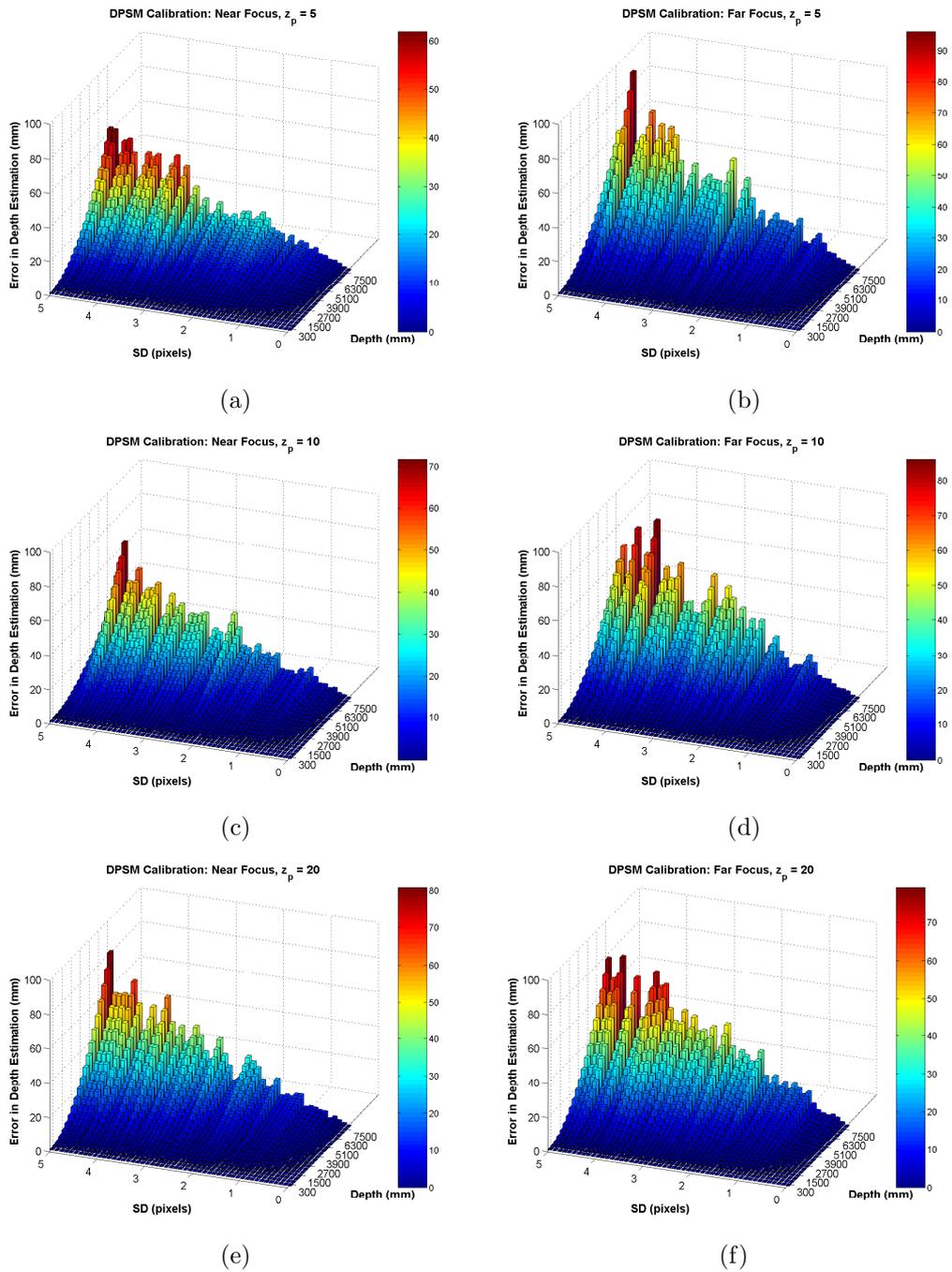


Figure 4.24: DPSM Depth estimation error using calibrated parameters. (a) Near-focus configuration with  $z_p = 5mm$ ; (b) Mid-far focus configuration with  $z_p = 5mm$ ; (c) Near-focus configuration with  $z_p = 10mm$ ; (d) Mid-far focus configuration with  $z_p = 10mm$ ; (e) Near focus configuration with  $z_p = 20mm$ ; (f) Mid-far focus configuration with  $z_p = 20mm$ .

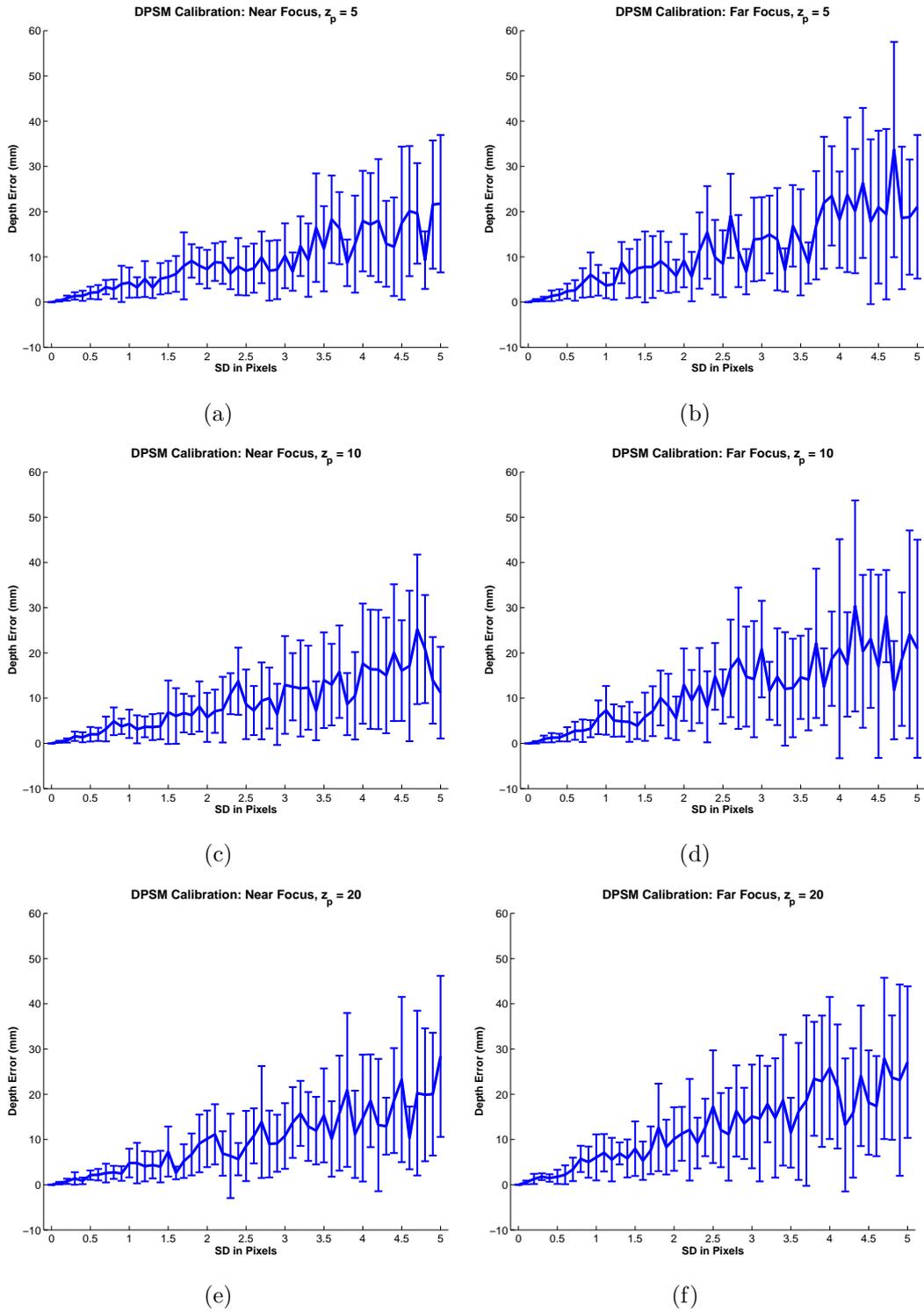


Figure 4.25: DPSM Depth estimation error using calibrated parameters errorbar plots. (a) Near-focus configuration with  $z_p = 5mm$ ; (b) Mid-far focus configuration with  $z_p = 5mm$ ; (c) Near-focus configuration with  $z_p = 10mm$ ; (d) Mid-far focus configuration with  $z_p = 10mm$ ; (e) Near focus configuration with  $z_p = 20mm$ ; (f) Mid-far focus configuration with  $z_p = 20mm$ .

Experiments conducted in this section have examined various aspects of the proposed DPSM calibration algorithm. The initial calibration target depths are chosen based on the analysis of the DPM setup. It was found that provided the initial target depth was greater than  $500mm$ , and the translational component between the targets was at a minimum approximately one quarter of that distance, a reliable parameter estimation is achieved. Naturally, as the magnitude of the translational component increases, the robustness of the parameter estimates also increases. A thorough examination of the accuracy in the parameter estimation was performed and it was determined that a mid-far focus configuration was the optimal setting to perform the DPSM calibration procedure. The increased performance was due to a more accurate determination of the sensor depth parameter, which is the most sensitive parameter within the system. An increase in accuracy of the estimation of the pupil plane axial translation was also observed in the mid-far focus configuration as  $z_p$  increased. This increase in performance as  $z_p$  increased in the DPSM model was also observed in the experiments conducted in estimating object depth using the calibrated DPSM parameters. The worst case depth estimation was recorded in the presence of 5 pixels SD Gaussian noise and had a mean error of  $27.1mm$  (SD  $16.7mm$ ). Application of standard pinhole calibration methods to the DPSM is likely to induce large error. The accuracy would be, at best, similar to that outlined in the experimentation in the DPM section. However, it is likely that these methods would lead to considerably less accurate estimates than outlined in Section. 4.3.2. Due to the axial translation of the pupil plane, calibrating with the standard pinhole methods would require manual placement and measurement of the pupil plane from the natural position to the translated position after the calibration. This further re-enforces the applicability of the DPSM calibration method since it is multi-pupil based, the calibration process is a more natural process and, thus, produces more reliable parameter estimates.

## 4.4 Real Experiments

The DPM and DPSM imaging models are tested using real data with respect to the calibration of the camera and the resulting depth estimation process. Standard camera calibration procedures are conducted and a resulting depth estimation is performed in order to evaluate the DPM and DPSM imaging

models.

#### 4.4.1 Experimental Setup

The camera used for all experiments in this section is the Panasonic Lumix DMC-G1, a Micro Four Thirds (MFT) format camera. The MFT format allows interchangeable lenses without the need for internal mirrors and a pentaprism such as those used in standard DSLR systems. Due to the decreased number of components in the camera, the physical distance between the lens mount and the image sensor is reduced. This allows smaller diameter lenses to be used, which provides an ideal test bed in which to experiment with modified optical components. These ideal conditions are enhanced by having a large image sensor ( $17.3\text{mm} \times 13\text{mm}$ ) which allows images to be captured at  $4,000 \times 3,000$  pixels yielding a pixel pitch of  $4.3\mu\text{m}$ . Image pixel resolutions of this order ensure that the choice of optical components will be the limiting factor in resolving image details. An EO achromatic lens of focal length  $75\text{mm}$  was used for the conducted experiments. EO lenses were chosen for two reasons. Firstly, from a practical aspect, the modification of the pupil plane requires direct access to the primary elements of the imaging lens. Typically, conventional imaging lenses are encased and difficult to disassemble without damaging the lens. Secondly, the optical properties of the EO catalog are available in the Zemax software. Consequently, the optical resolution of the lens is known: 70 lines/mm. The resolvable resolution of the image sensor can be calculated using a machine vision primer ( $500/\text{pixelsize}$ ) Stemmer (2011) and is found to be  $\sim 116$  lines/mm. Therefore the resolution of the implemented system is sufficient to capture detailed image projections through imaging lens.

An additional benefit of using a MFT format is “Live View”. This enables a continuous stream of the current imaging configuration and scene. This feature was exploited for the purpose of lens and pupil plane alignment. Pupil planes are manufactured using aluminium sheets of thickness  $0.35\text{mm}$ . Circular pupil planes of diameter  $25\text{mm}$  (same diameter as the EO lenses) are machined using a LPKF ProtoMat<sup>®</sup> C30/S.<sup>2</sup> Pupils of diameter  $1\text{mm}$  are drilled into the plane during the manufacturing process. Alignment of the image sensor with the optical lens and pupil plane is achieved by modifying the MFT lens

---

<sup>2</sup><http://www.lpkf.com>

mount. A C-mount lens connector is centrally integrated to the MFT mount. This enables the use of the full range of C-mount spacers and connecting tubes which accompany the EO lenses. Alignment of the pupil plane is guaranteed provided its diameter matches the lens and connector tube diameters. The experimental setup in terms of the camera body, lens elements and connectors is shown in Figures 4.26(a), 4.26(b), and 4.26(c).

Defining object points which are on the optical axis of the imaging system was one of the main difficulties with the experimental setup. Alignment of object points with the optical axis was acquired by using a laser. As shown in Figure 4.26(d), the laser was mounted on a custom made bracket. Four degrees of freedom are achieved using an electro mechanical rotation stage, a mechanical rotation stage and two mechanical translation stages. The two degrees which were dropped correspond to rotation about the optical axis and translation along the optical axis, of which neither affect the axial alignment. Alignment of the laser with the optical axis was a two stage process. Initially, the camera was calibrated using a standard planar approach to identify the intersection of the axis with the image plane i.e. the principal point. The second stage involved manually aligning the laser (via rotation and translation stages) with the marked principal point in the image. It was also necessary to ensure that the laser was intersecting the centre of the lens. This was achieved by visually aligning the laser spot with a custom made pupil plane with a pinhole located at its centre. The resolution of the electro mechanical rotation stage was  $0.001^\circ$ . The resolution of the manual rotation stage was  $0.1^\circ$ , while the mechanical translation stages had a resolution of  $0.01\text{mm}$ . Due to the sensitivity of the optical axis alignment, consistency across all experiments was maintained by mounting the camera body and laser on optical tables.

Object depth range was limited by the size of the laboratory in which the experiments were conducted. This equated to a range of approximately  $4.3\text{m}$ . As a result of the laser and optical axis alignment, object depths were defined by intersecting the axis with an orthogonal diffuse plane. This is illustrated in Figure 4.26(e) where it can be seen that when the plane intersects the axis, the laser beam spot is visible on the diffuser at that depth. Object depth was measured via alignment of the diffuse plane with a measuring tape, which was fixed parallel to the optical axis. The entire experimental setup is shown in Figure 4.27.

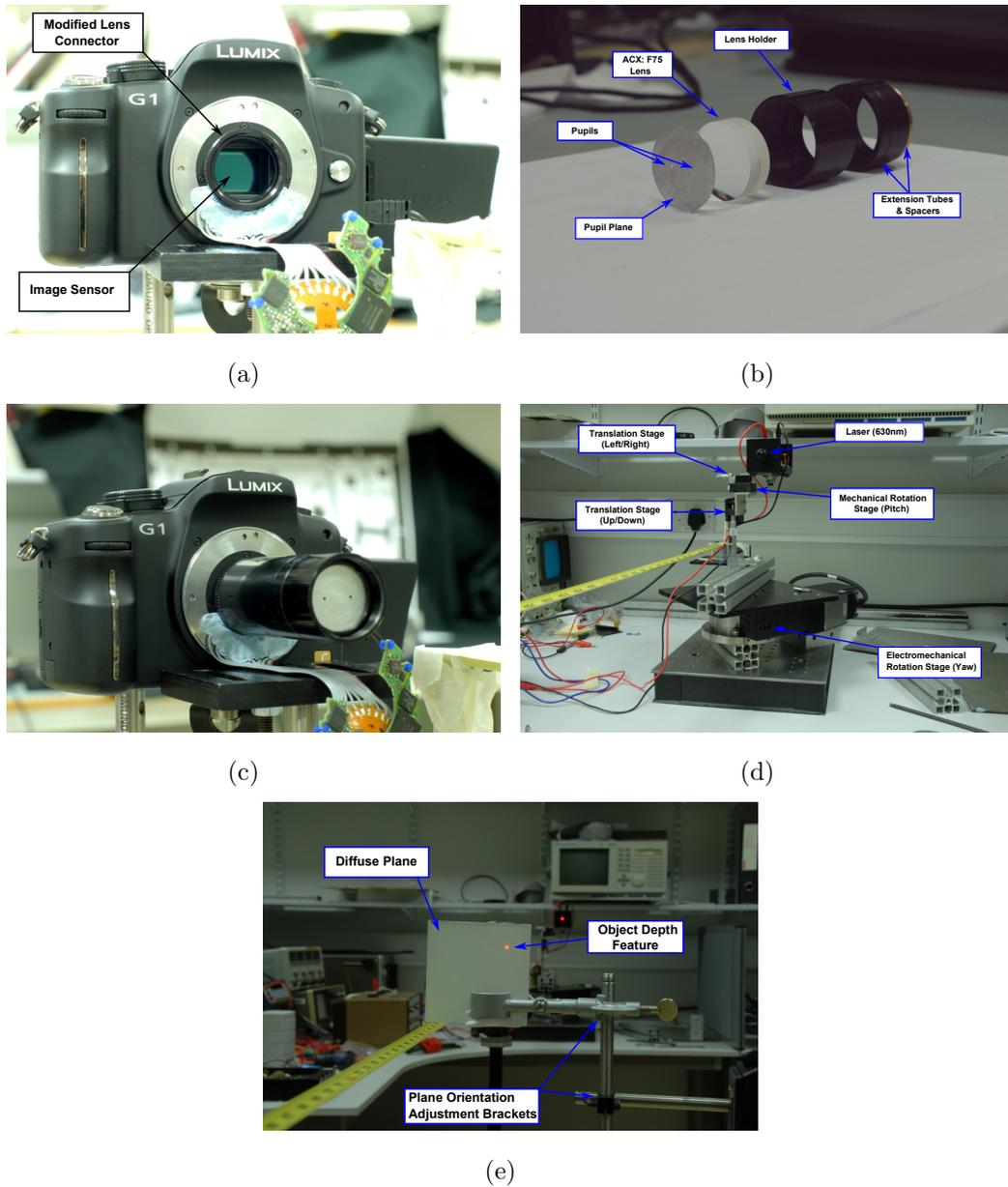


Figure 4.26: Experimental setup components. (a) Camera body with modified lens mount; (b) Custom lens arrangement with pupil plane; (c) Multi-pupil imaging system; (d) Laser setup; (e) Diffuse plane for enabling object depth measurements.

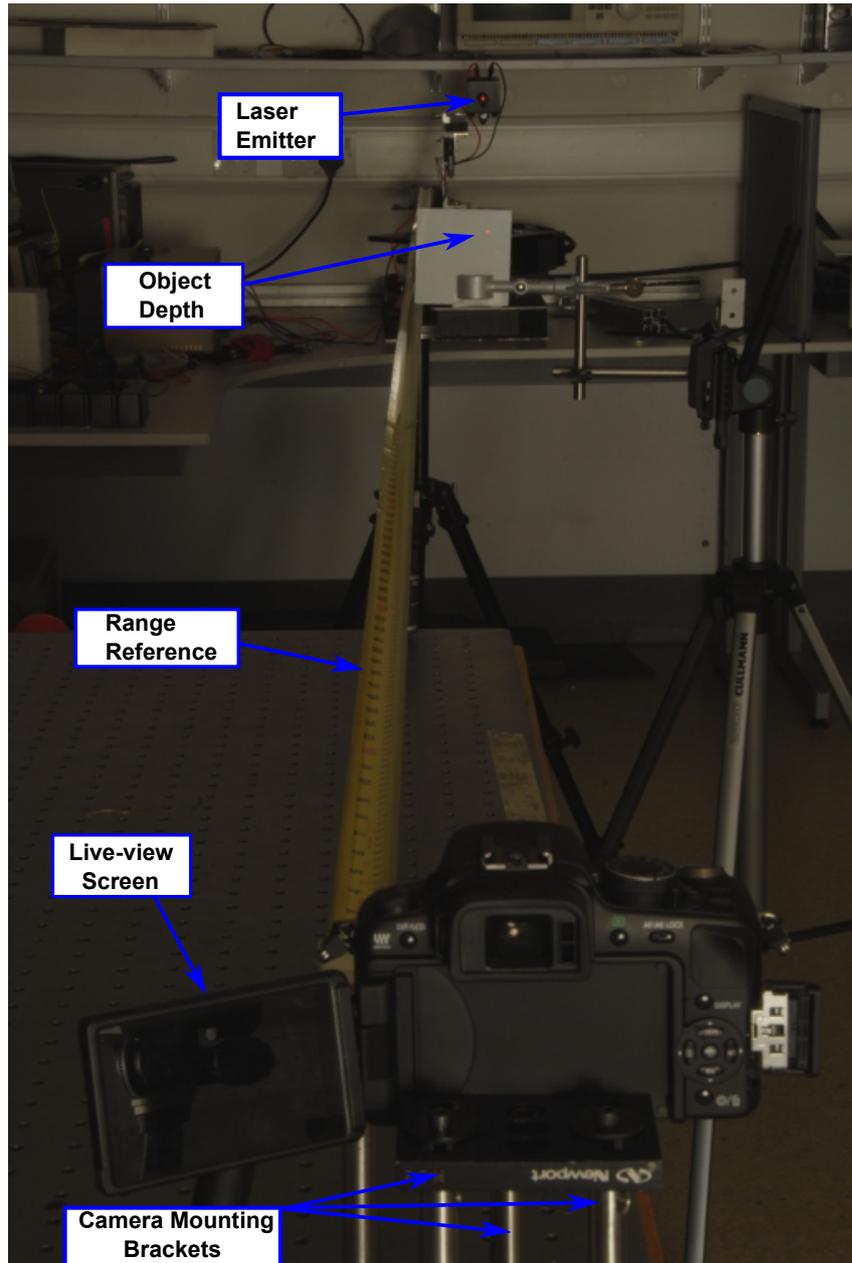


Figure 4.27: Experimental setup.

#### 4.4.2 Calibration Procedure

When metric measurements are required from an imaging system, the sensor depth is a fundamental parameter in completing this task. The sensor depth describes the optical configuration of the imaging system. Coupled with knowledge of the focal length and pixel pitch, it allows metric measurements to be made from resulting image pixel disparities. Two new methods have been proposed for calibrating the DPM and DPSM sensor depth. Simulated results have shown these methods to outperform the standard methods (Sec-

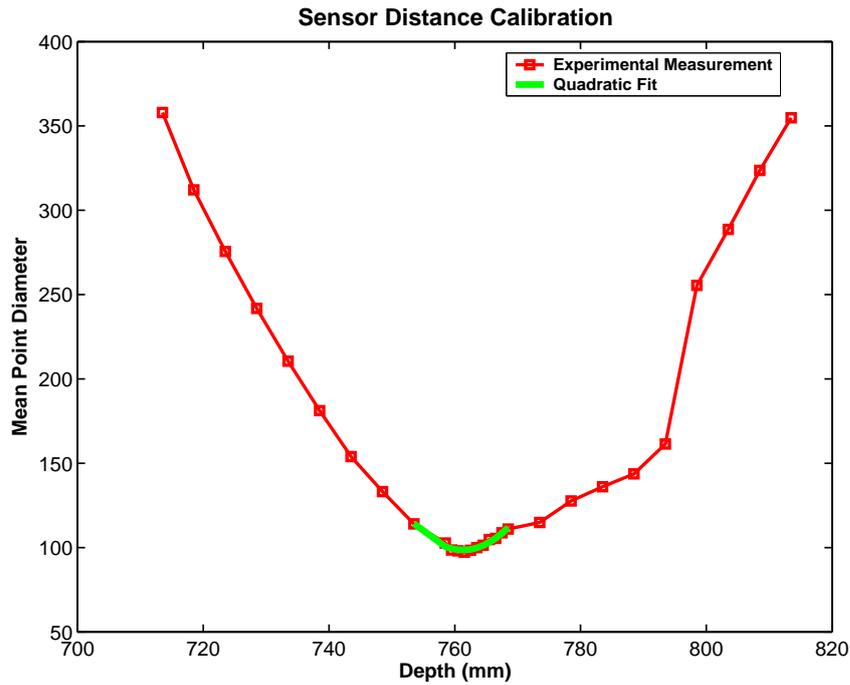


Figure 4.28: Ground truth sensor depth calibration.

tion 4.3.2). The aim of the experiments conducted in this section is to quantify the accuracy of the DPM and DPSM methods in calibrating the sensor depth in comparison to the pinhole approaches.

An initial experiment was conducted to obtain a ground truth for the sensor depth parameter in the experimental configuration. Object points were imaged every  $10\text{mm}$  over an interval of  $100\text{mm}$  within the region of the apparent focus. Subsequently a smaller region, around the focused depth, was imaged in  $1\text{mm}$  steps over a range of  $10\text{mm}$ . A Gaussian model was fitted to each imaged object point in which the average of the 2D width was taken as the spot size. The resulting minimum spot size over the entire interval was defined as the object depth that was in focus, which along with the known focal length of the lens allowed the sensor depth to be calculated. The results for this experiment are shown in Figure 4.28. A second order polynomial was fitted in the region corresponding to the minimum spot size to further refine the sensor depth estimate. An object depth of  $761.3\text{mm}$  was found to be the minimum spot size distance. This corresponds to a sensor depth of  $83.20\text{mm}$ .

### Standard Calibration Approaches

In a similar manner to the conducted simulated experiments, the standard calibration methods of Zhang (2000) and imaging a sphere are carried out with real data. In the case of the planar calibration, ten images of planar grids with varying pose are captured. Imaging the planar grids around the focus depth ensures features remain sharp and thus the feature extraction process is straightforward. The calibrated sensor depth using the method of Zhang (2000) was  $82.618mm$ .

Calibration using a spherical object is conducted by centering a spherical object of known diameter, which in this case is a black snooker ball, at the focus distance. In addition to being centred at the focus distance, the sphere must also be centred on the optical axis. Alignment is achieved by adjusting the height of the sphere until the laser intersects it at its centre. Subsequently, the scene is back lit using a light box, which increases the contrast between the black snooker ball and the apparent white background. This aides with the feature extraction process which in this case involves fitting an ellipse to the edge detected, snooker ball silhouette. An example calibration image captured in this procedure is presented in Figure 4.30. Since the calibration object being imaged is spherical, an adjustment of the diameter is required due to the nature of the experimental setup. Figure 4.29 illustrates where the adjustment is needed. Since the point on the sphere being imaged is not at the exact centre of the sphere, an adjustment of the depth is required. The adjusted object depth  $u'$  can be found by examining the two right angle triangles with sides  $(r, s, u_0)$  and  $(h, s, u')$ . In reference to Figure 4.29,  $s = \sqrt{u_0^2 - r^2}$ ,  $h = \frac{sr}{u_0}$  and  $u' = \sqrt{s^2 - h^2}$ . Now the adjusted object depth ( $u'$ ) and height ( $h$ ) are used for the sensor depth calculation,  $v = \frac{du'}{2h}$ . This experiment was conducted six times for varying levels of camera exposure. The resulting average calibrated sensor depth value for these experiments was  $82.55mm$ .

### Multi-pupil Calibration

The multi-pupil calibration methods proposed in Section 4.3 require an object point on the optical axis at various depths to be imaged. In the case of the DPM, two object depths are used in which the translation between object points is known. This scenario can be captured in a single image, however,

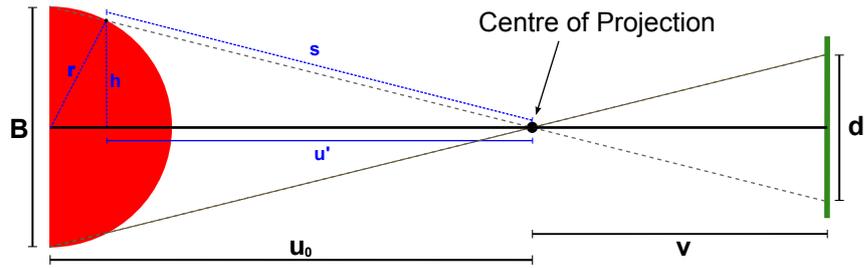


Figure 4.29: Adjusting the spherical calibration depth parameter.

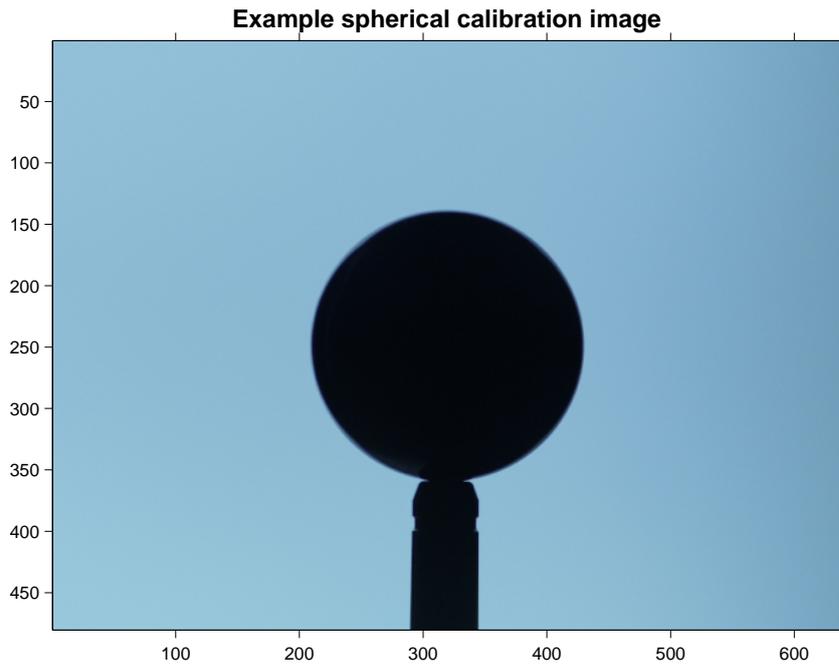


Figure 4.30: Example spherical calibration image (scaled to VGA).

the experimental setup outlined in this section only allows a single depth to be imaged each time. Therefore two images are acquired with a translational component of  $1m$  between each object point (as suggested in Section 4.3.2). Since each image only contains a single object depth, colour filters are not required for feature extraction.

An example image produced by the DPM is given in Figure 4.31. Images are captured in a dark room environment. Centroid localisation is implemented by fitting 2D Gaussian functions to each pupil point respectively and finding their centre. Centre of mass localisation was implemented but was found to be less accurate particularly as the image disparity decreased and the pupil points became less well defined/resolved. The DPM calibration was performed by taking an initial object depth approximately  $1.25m$  in front of the pupil plane. The second object depth was  $1m$  further on the object side. As noted,

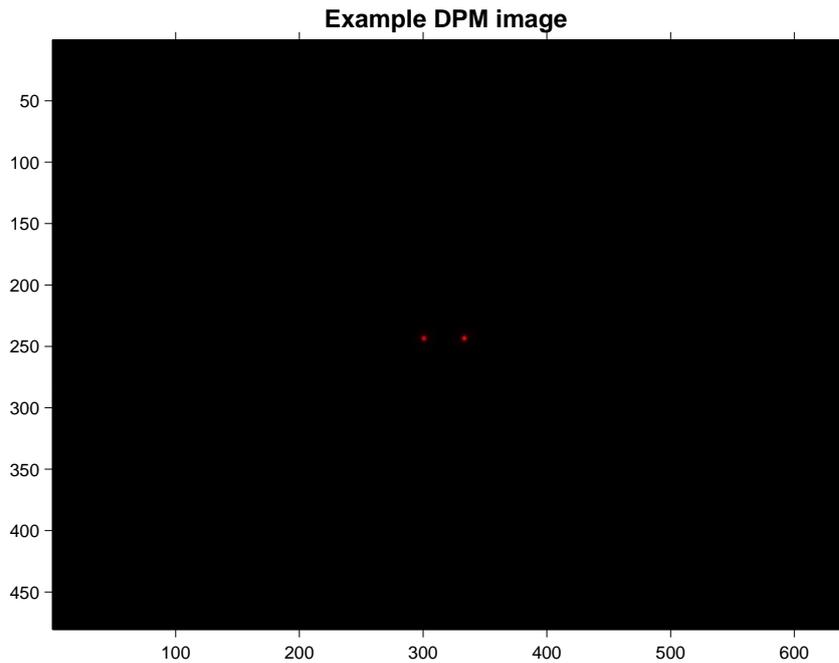


Figure 4.31: Example DPM image (scaled to VGA).

the actual object depths are not required, just an accurate measure of the translation between them. The resulting calibrated DPM sensor depth was  $83.19\text{mm}$  for a pupil gap separation distance of  $10\text{mm}$ . An identical calibration procedure was performed for the DPM with a pupil gap separation distance of  $20\text{mm}$ . The calibrated sensor depth in this configuration was  $83.18\text{mm}$ .

A DPSM imaging system was implemented with an axial translation of the pupil plane ( $z_p$ ) by approximately  $20.91\text{mm}$  (measured with digital calipers). Two DPSM configurations were tested corresponding to pupil gap separation distances of  $10\text{mm}$  and  $20\text{mm}$ . DPSM calibration differs from the DPM in that four distinct object depths are required. Since two parameters are being calibrated, sensor depth and pupil plane axial translation, additional object depths are required to enable sufficient constraints to solve the calibration equations. In this case, a minimum of four distinct object depths yields three independent equations which enables the camera parameters to be calculated. Images produced by the DPSM are similar to that presented in Figure 4.31. The only difference is that the image pixel disparities are marginally increased due to the pupil plane offset. As expected, and as is the case for the DPM, an increase in the pupil gap separation distance increased the image pixel disparity profile significantly. Object points for the calibration procedure were captured at depths of  $87\text{mm}$ ,  $1,000\text{mm}$ ,  $1,490\text{mm}$ , and  $2,290\text{mm}$ . A sensor depth of

Table 4.7: Comparison of calibration techniques. Percentage error given in brackets

Configuration	Parameter	Measured	Zhang	Sphere	DPM	DPSM
$P_g = 10$	$v$ (mm)	83.20	82.62 (0.69)	82.55 (0.77)	83.19 (0.01)	83.33 (0.16)
	$z_p$ (mm)	20.91	-	-	-	20.30 (2.93)
$P_g = 20$	$v$ (mm)	83.20	82.62 (0.69)	82.55 (0.77)	83.18 (0.01)	83.29 (0.10)
	$z_p$ (mm)	20.91	-	-	-	22.71 (8.6)

83.33mm and pupil plane axial translation of 20.30mm were calculated for a pupil separation distance of 10mm. In a 20mm pupil separation configuration, the sensor depth was found to be 83.29mm with a pupil plane axial shift of 22.71mm. The results of all calibrations performed along with the ground truth data are presented in Table 4.7

### Discussion

A single mid range focal length lens was chosen for the experiments (75mm) in which the imaging configuration was set to a near-mid focus. Two pupil gap separation distances were also tested: 10mm and 20mm. Simulated results for both the DPM and DPSM model have shown this focal length and focus range to yield the optimum conditions to implement both imaging models. Both pupil gap distances were implemented to examine the results of different image disparity baselines.

There are a number of conclusions to be drawn from the calibration results. The accuracy at which the sensor depth is estimated has a profound effect on the overall depth estimates of the multi-pupil imaging models. Simulated results presented in Sections. 4.1.2 and 4.2.2 outlined the implication of poor accuracy in the sensor depth estimation process due to the highly sensitive nature of this parameter. Table 4.7 outlines the results for each calibration experiment performed. It is clear that the multi-pupil methods are superior to the standard methods for accurately estimating the sensor depth. Standard planar calibration yields an error of 0.695% in comparison to the spherical method which resulted in an error of 0.773%. Even though the error is small in magnitude, it equates to modifying the focus of the entire imaging system which, in turn, changes the disparity profile for the subsequent metric measurements. The results for the DPM calibrations contain error of an order of

magnitude less than the standard techniques. Errors of 0.013% and 0.017% are recorded for pupil gap distance of  $10mm$  and  $20mm$  respectively.

In the case of the DPSM calibration, an additional model parameter is required from the calibration process. Therefore any perturbations within the calibration environment will propagate through affecting both parameter estimates. A total of four distinct object depths are required to perform this calibration, thus yielding three calibration target offsets (Eqn. 4.17). As noted in Section. 4.3.3, as long as the translation between the grids was roughly a quarter of the object distance, accurate results would be obtained in the presence of noise. During the initial experimentation, object distances similar to that used for the DPM were chosen. Due to the additional calibration targets required for DPSM calibration, some of the targets were located at the extremities of the depth range ( $\sim 4, 200mm$ ). This area of the disparity profile (corresponding to the large depths) is the least sensitive to changes in object depth, therefore any perturbations in determining the disparity at this depth will propagate through the calibration process to the estimation of the model parameters (due to the assumption of known translation distances between calibration objects). Therefore, an initial object depth situated in the most sensitive region of the disparity profile was chosen, in the  $< 300mm$  region. This ensured that the resulting calibration target disparities would not induce large error in the overall procedure.

Simulated experiments in relation to the sensitivity of the pupil plane axial shift parameter,  $z_p$ , were presented in Figure 4.14(c). It was shown that regardless of the magnitude of  $z_p$ , the induced error in depth estimation remained insignificant in comparison to induced error due to sensor depth inaccuracy. However the dependence of sensor depth on  $z_p$  within the DPSM calibration framework gives rise to the importance of obtaining an accurate estimate for the axial shift parameter. The results for  $z_p$  estimation were  $20.30mm$  and  $22.71mm$  for  $P_g = 10mm$  and  $P_g = 20mm$  respectively. This represents an error of 2.9% and 8.6% of the measured value for  $z_p$  respectively. Corresponding sensor depth estimation results in error of 0.16% and 0.1% of the measured value. In comparison to the DPM estimates, the error is approximately an order of magnitude larger, but remains significantly less than the standard calibration results.

The difference in performance between the standard calibration techniques

and the multi-pupil approach can be partially accounted for by the fact that there are aberrations present in the imaging lens. With the multi-pupil techniques, the calibration targets remain on the optical axis of the camera, thus only spherical aberration has an effect on the calibration targets. Since spherical aberration remains constant on axis, it is partially accounted for by the multi-pupil imaging models. An increase in  $P_g$  results in an increase in spherical aberration. Thus larger error is seen particularly with the DPSM model, as shifting the pupil plane has the effect of directing the incoming object rays towards the periphery of the lens. This is where spherical aberration is maximised in an imaging lens. As the object points move off axis, as is the case for planar calibration and calibration using a spherical object, additional aberrations will distort the true corresponding image point locations. Thus calibration results will become less reliable. Another factor to consider is ellipse fitting with the spherical calibration. Image blur renders this task quite susceptible to noise in the ellipse fit which results in inaccurate sensor depth estimates. Planar camera calibration will also suffer with feature detection problems in the presence of lens aberrations. This also results in poor sensor depth estimates. With regard to practicality, there are issues with adopting the standard calibration approaches to a DPSM imaging configuration. Direct modification of the imaging configuration is required by removing and replacing the pupil plane during the calibration procedure. This process greatly increases the chance of inducing error into the calibration process through misplacement/misalignment of the pupil plane. Additionally, the likelihood of modifying the sensor depth whilst performing manual alterations to the DPSM is increased.

### 4.4.3 Depth Estimation

Depth estimation experiments were performed for each multi-pupil camera model in both pupil gap separation distance configurations. The same imaging systems which were examined in the calibration experiments were examined for the depth estimation. Therefore, the depth estimation experiments also quantitatively evaluate the calibration results obtained in Section 4.4.2. Each experiment consisted of imaging an axial object point at various depths. A total of 21 object depths were imaged across a range of 4,300mm (22 depths were measured for the DPSM in a  $P_g = 10mm$  configuration). Each captured

Table 4.8: Depth estimation error using the calibration results of standard and multi-pupil approaches. (S) corresponds to standard approaches applied with measured  $z_p$ .

Configuration	Measurement	Zhang	Sphere	Zhang (S)	Sphere (S)	DPM	DPSM
$P_g = 10$	<i>Mean (% error)</i>	26.42	30.99	27.04	31.80	0.85	1.34
	<i>SD (%error)</i>	17.40	21.04	18.24	22.11	0.73	1.07
$P_g = 20$	<i>Mean (%error)</i>	36.11	41.76	36.36	42.35	0.73	0.99
	<i>SD (%error)</i>	26.42	31.58	27.61	33.11	0.62	1.11

image was subsequently processed by fitting Gaussian models to the imaged object point intensities. This allowed the corresponding image disparities to be calculated and subsequently fed into the multi-pupil imaging models for object depth calculation. Sensor depth calibration results obtained using the standard approaches were implemented in the multi-pupil imaging models for comparison. In the case of the DPSM, where the pupil plane axial shift parameter ( $z_p$ ) requires estimation, the measured value of  $z_p$  (21.91mm) was used for calculating object depth using the standard calibration approaches.

Results for object depth estimation of the DPM and DPSM in configurations with a pupil gap of 10mm are presented in Figure 4.32(a)-(d). Figures 4.32(a) and 4.32(c) show the estimated object depths calculated using the calibration results obtained from the standard approaches and the multi-pupil approaches. The corresponding depth estimation errors are presented in Figures 4.32(b) and 4.32(d). It is clear from both sets of results that the multi-pupil calibration models are considerably more accurate in depth estimation. The standard planar approach of Zhang (2000) performs slightly better than calibration using a spherical object. Table. 4.8 presents the mean and SD of the depth estimation errors as percentages. DPM and DPSM mean error (SD) are 0.85% (0.73%) and 1.34% (1.07%) respectively. These figures are an order of magnitude less than the observed error in depth estimation when implementing the standard calibration approaches. Planar calibration produces a mean error of 26.42% (17.4%) in a DPM configuration and 27.04% (17.24%) in a DPSM system. A poorer performance is recorded by employing the calibration results of the spherical approach. Mean errors of 30.99% (21.04%) and 31.80% (22.11%) are recorded in DPM and DPSM configurations using these calibration results.

The depth estimation results for a multi-pupil imaging system with a pupil separation distance of 20mm are presented in Figures 4.33(a)-(d). Once more,

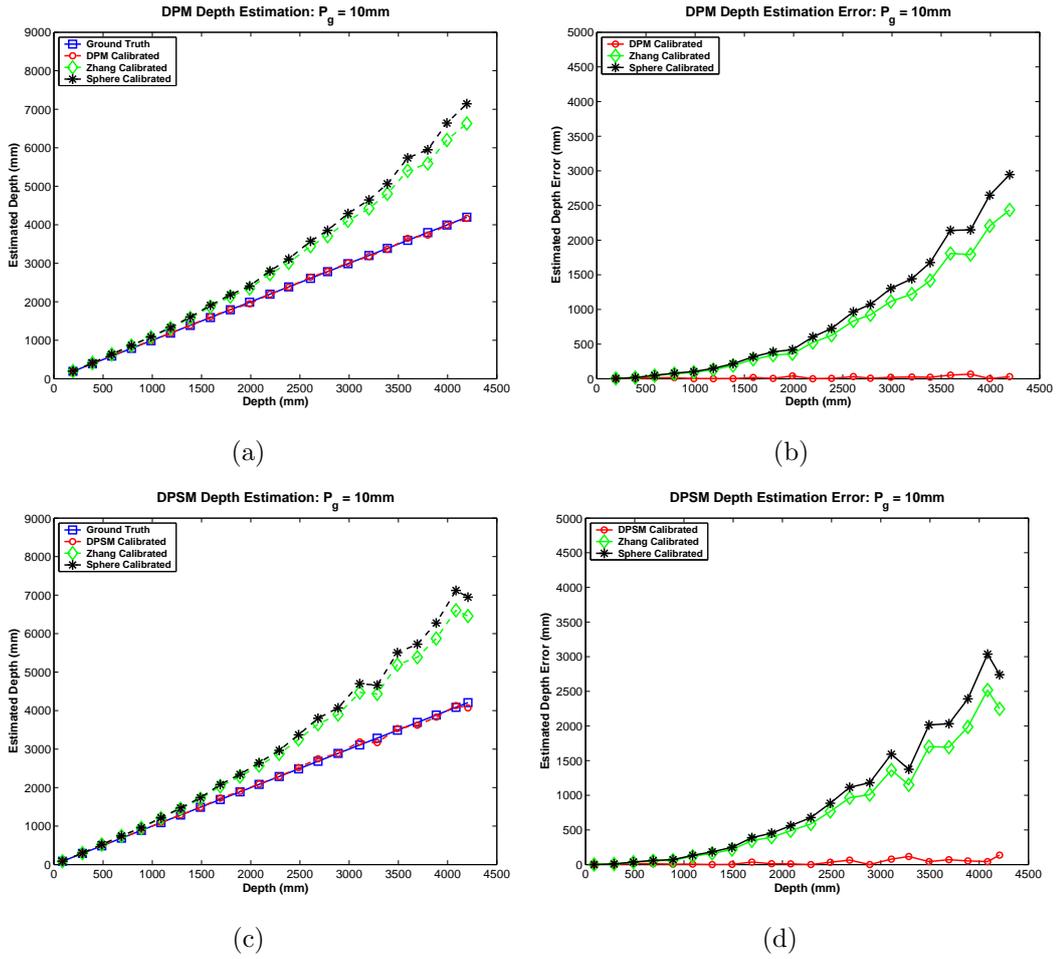


Figure 4.32: Depth estimation results for  $P_g = 10\text{mm}$  (a) DPM Depth estimation; (b) DPM Depth estimation error; (c) DPSM Depth estimation; (d) DPSM Depth estimation error.

the DPM and DPSM calibrated systems are estimating depth at a level of accuracy far superior to that of the standard calibration approaches. An increase in the mean and SD of the error is observed with the standard approaches in comparison to the 10mm pupil gap configuration. Planar calibration attributes 36.11% (26.42%) and 36.3% (27.61%) error with respect to a DPM and DPSM system. A spherical calibration approach results in error increases to 41.76% (31.58%) and 42.35% (33.11%). Therefore the depth estimation results presented in this section re-enforce the multi pupil approaches in calibrating a multi-pupil imaging system.

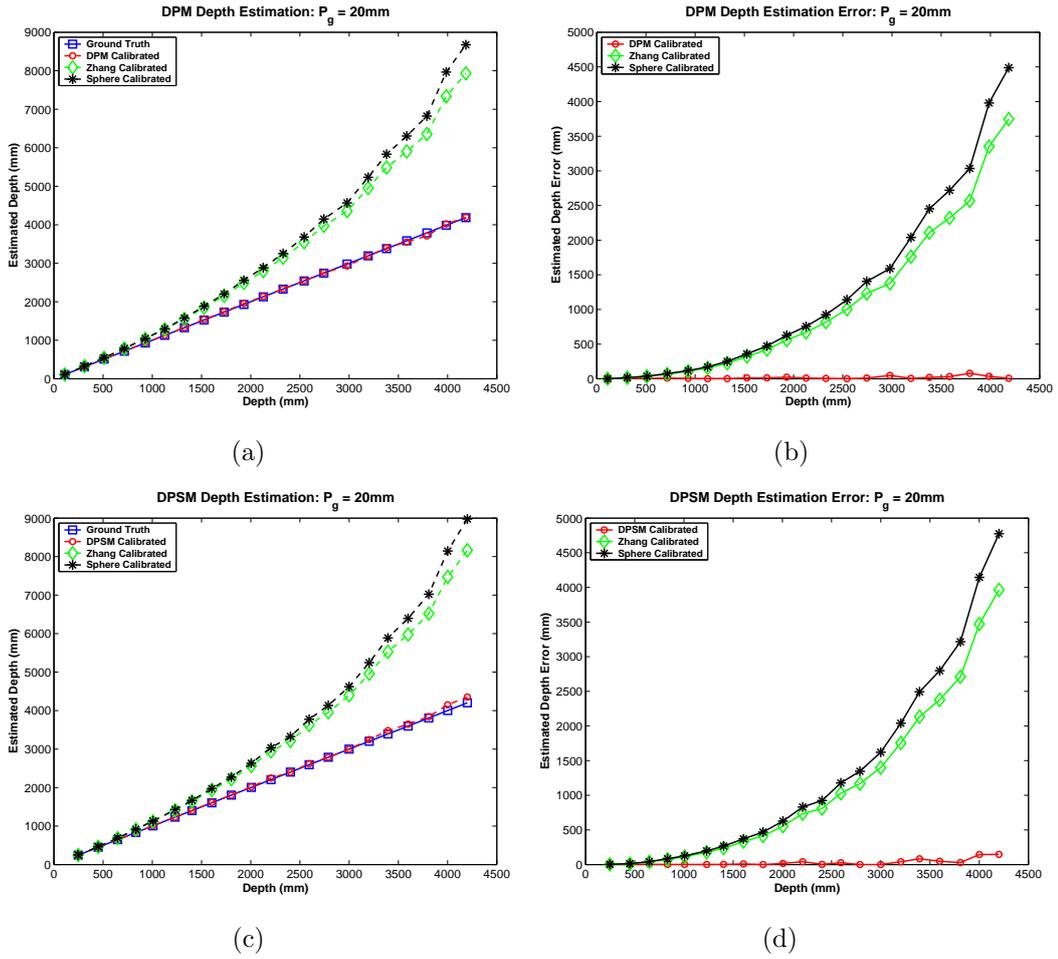


Figure 4.33: Depth estimation results for  $P_g = 20\text{mm}$  (a) DPM Depth estimation; (b) DPM Depth estimation error; (c) DPSM Depth estimation; (d) DPSM Depth estimation error.

## Discussion

There are a number of factors to consider when analysing the depth estimation results. As discussed in Section 4.4.2, the sensor depth is the most sensitive parameter of the multi-pupil imaging models. Poor estimates of the sensor depth, in the order of  $> 0.6\%$ , have a detrimental effect on depth estimation. Object depth errors of  $70\%$  are recorded with the planar calibrated sensor depth, while errors larger than  $100\%$  are seen with the spherical calibration. However, it is noted that sensor depth errors of this magnitude have a smaller effect on near-object depth estimation. Both standard approaches exhibit depth errors less than  $\sim 10\%$  within a range of  $\sim 1,000\text{mm}$ . For configurations with a pupil gap distance of  $20\text{mm}$  this range is slightly reduced to  $\sim 750\text{mm}$ . This robust performance is a result of the multi-pupil imaging system disparity profile

which is highly sensitive to changes in depth at this range. DPM and DPSM worst case depth estimates are 2.02% and 3.62% respectively which correspond to depth errors of  $78mm$  for the DPM and  $152mm$  for the DPSM. Errors of this magnitude are well within the bounds of acceptability, particularly across a depth range of  $4,300mm$ . The mean and SD error in depth estimation presented in Table 4.8, for the proposed calibration approaches, exhibits a marginal increase in performance for both imaging systems in configurations with a pupil gap of  $20mm$ . In the case of the DPM, this occurs despite an increase in the error between the measured and calibrated sensor depth (additional  $\sim 3\mu m$ ). With the DPSM, there is an increase in the accuracy of the sensor depth estimate, but, a decrease in the pupil plane axial translation estimation accuracy. Again, the increase in accuracy is marginal, however it is an important property of the multi-pupil imaging system. A closer examination of this behavior leads to the conclusion that the aberrations within the imaging lens have a significant role in defining the accuracy of the multi-pupil imaging models. The main difference between the multi-pupil imaging models with  $P_g$  set to  $10mm$  and  $20mm$ , besides the increased image disparity baseline, is an increase in spherical aberration. As discussed in Section 4.4.2, spherical aberration is present in images of on-axis object points. Rays emanating from object points passing through the pupils on the pupil plane strike the lens at a radial distance from its centre. The greater the radial distance of the ray striking the lens, the greater the spherical aberration observed on the image plane. This is realised on the image plane as a marginal increase in the magnitude of image disparities. Therefore, increasing the pupil gap distance effectively increases the magnitude of spherical aberration present in the imaging system. This is the reason why multi-pupil calibration approaches are more accurate than standard approaches. As standard approaches become even less accurate with an increase in  $P_g$ , the DPM and DPSM marginally adjust their calculation of the sensor depth to account for the spherical aberration. Naturally, there are additional monochromatic aberrations present in the imaging lens. However these are only realised for object points which do not lie on the optical axis. This explains the inaccuracy of the standard calibration approaches as both methods utilise off axis object points for the calculation of sensor depth. Depth estimation results employing the standard calibration approaches reaffirm their unsuitability for application to the multi-pupil imaging framework. The DPSM results, in particular, demonstrate how the multi-pupil model adapts its sensor depth calibration to account for the

spherical aberration. Since there are two parameters being minimised in the DPSM calibration equations (4.17), there is more flexibility in the minimisation to locate the local minimum which accounts for spherical aberration in the imaging system. This effect is highlighted by the difference in magnitude of the estimated sensor depths for a  $P_g$  of  $10mm$  and  $20mm$  in comparison to the measured and DPM calibrated values in Table 4.7. Despite the apparent increase in sensor depth error, the depth estimation results remain accurate to a level comparable to the DPM system (Table 4.8).

Analysis on the sensitivities of both multi-pupil imaging models was conducted in Sections 4.1.2 and 4.2.1. The depth estimation in this section has confirmed those simulated results which characterised the model parameters. The accuracy to which the focal length of the imaging lens is known ( $.01mm$ ) does not affect the overall depth estimation accuracy. Pupil plane properties such as the pupil gap separation distance, which was shown to be a sensitive parameter in both multi-pupil models, is manufactured with sufficient precision ( $7.8\mu m$ ) to accurately estimate depth. The most likely source of error to be induced in the system is in the measurement of image disparity. Image disparity can be measured to a sub-pixel level of accuracy. It is the definition of the centre of image intensities relating to each pupil which requires consideration. Initially, a simple centre of mass calculation was made for localisation, however it was found during experimentation that this step lacked the required accuracy, particularly around the focus depth where both image “spots” become difficult to separate. Difficulties were also observed at larger object depths ( $> 3,500mm$ ) where the image disparity profile becomes less sensitive to object depth changes. In order to increase the accuracy of the spot centre localisation, Gaussian functions were fitted. This step was sufficient to increase the resolution of the spot localisation in the difficult areas of the image disparity profile. If the object depth range was extended beyond  $4,300mm$ , it is expected that the accuracy of the depth estimation would decrease based on the resolution to which the image disparities could be estimated. This problem could be addressed by using higher quality imaging lenses which resolve image detail to higher degree of accuracy. This is a valid solution provided that the pixel pitch of the image sensor is of equal or higher resolution than the optical resolution. An alternative approach would be to increase the focal length in the multi-pupil imaging system. The scalability of multi-pupil imaging systems was identified in the system characterisation (Section 4.1 and 4.2). It

was shown that an increase in focal length of the imaging system increased the magnitude of the image disparity profile. Increased magnification of image disparity increases the baseline at larger object depth ranges. Thus localisation becomes less of an issue, however, a decrease in the Field of View (FOV) is observed due to the image magnification effect of a long focal length. This could be problematic in determining object depths on the near side of system focus. It is these object depths which will correspond to the largest image disparity, and thus may not necessarily fall on the image sensor. This could have practical implications for both multi-pupil calibration procedures.

Accuracy of the multi-pupil depth estimation results also validates the optical alignment process of the experimental procedure. Any error induced due to non-alignment of the optical axis with the sensor are of a magnitude small enough that it does not affect the overall depth estimation process. The rotational alignment of the pupil plane with the image sensor is unaffected due to the rotational symmetry of the imaging lens. The primary concern was that the pupil plane was parallel to all other surfaces in the imaging system. This constraint was guaranteed with the lens tubes and spacers used during experimentation. The depth estimation experiments were carried out using a  $75mm$  focal length lens. Simulations indicated mid range focal length as being ideal to obtain accurate depth estimates for a near to far object depth range. It was suggested in simulated results that the pupil gap distance should be maximised within the lens diameter. Therefore two pupil gap distances were examined during the depth estimation process. It was found that the increased pupil gap introduces spherical aberration into the imaging system. However the multi-pupil calibration algorithms naturally account for this aberration as it is fixed across all object depths. In order to achieve accuracy in the sensor depth, it is recommended that the pupil gap distance not be maximised within the lens diameter. Therefore  $P_g$ 's in the order of 0.5 times the lens diameter should be used. This will ensure that the effect of the spherical component is maintained at acceptable levels within the imaging system.

## 4.5 Discussion

This chapter proposes a new imaging model capable of recovering metric depth information from a single image. Two camera models are developed under the

proposed multi-pupil imaging criterion. Both models build on modified properties of the pupil plane within the imaging system. The standard pinhole camera model suffers a loss of depth information when projecting 3D information onto a 2D image plane. With multi-pupil imaging, 3D object points are captured by each pupil and imaged onto the 2D image sensor. Thus, the depth information is encoded into the resulting captured multi-pupil image. This encoding takes the form of a local pixel disparity between the detected object point imaged by each pupil of the pupil plane, within the captured image. Images captured using the multi-pupil camera model contain depth dependent image pixel disparities. With metric knowledge of pupil plane properties, and basic information of the imaging system configuration such as focal length and image sensor size, image pixel disparities can be used to recover metric depth information for the detected object points in the multi-pupil image.

The first camera model, DPM, is presented in Section 4.1.1. It consists of two pupils separated on the pupil plane by a known distance. An ideal lens is assumed and the pupil plane is located at the lens centre. Simulated experiments were performed to verify the DPM model. Zemax was also used to validate the DPM. A full characterisation of the model parameters was completed in which it was found that the sensor depth is the most sensitive model parameter within the system for depth estimation. A further examination of the relationship between image disparity and object depth reveals that this profile is defined by the sensor depth configuration of the imaging system. Therefore, depending on the focus configuration of the imaging system, the profile of the image pixel disparity could be advantageous in certain circumstances. Subsequent simulated depth estimation experimentation, with the DPM in Section 4.1.2 found such conditions to arise when the DPM was in a near-focus configuration. This equates to an imaging system with a large sensor depth. It was shown that the increased performance in the presence of noise was due to a broadening of the image pixel disparities as a result of a large sensor depth.

Due to the requirements of the DPM camera model, in particular the placement of the pupil plane exactly at the imaging lens surface, a second multi-pupil model is proposed. This is realised in the form of a shifted multi-pupil imaging system. The DPSM, presented in Section 4.2.1, relaxes the constraint of locating the pupil plane at the imaging lens and thus enables more flexibility in the design of the imaging system. Placement of the pupil plane can be freely located on the object side of the lens. This extension of the multi-pupil

imaging criterion requires an additional model parameter to describe the pupil plane axial shift. A complete characterisation of the DPSM camera model is presented in Section 4.2.2 along with a Zemax verification. The image disparity profile of the DPSM varies depending on the magnitude of the pupil plane axial shift parameter,  $z_p$ , as well as the sensor depth distance. As  $z_p$  increases, the disparity profile broadens, particularly at the larger object depths, which increases performance for imaging configurations in a far-focus setting. It is shown that with  $z_p$  set to zero, the DPSM becomes equivalent to a DPM system. Thus the sensitivity analysis of the shared parameters remain consistent. The additional parameter,  $z_p$  does not exhibit significant instability within the DPSM system, although it is shown to be sensitive to error in the pupil gap separation distance in a near-focus configuration. However, this property of the pupil plane is manufactured with sufficient precision to avoid any perturbations in the depth estimation. Simulated depth estimation experiments confirm the robustness of the DPSM model in estimating object depth. Simulations reveal that a mid-far focused configuration performs marginally better than other tested configurations.

Camera calibration is a fundamental task in the recovery of metric information from images. It comprises of estimating the camera model parameters which describe the 3D to 2D projection from the world to the image plane. To this end, Section 4.3 proposes two new calibration algorithms which recover the DPM and DPSM model parameters. Standard pinhole methods are applied in order to calculate the sensor depth for comparison with the proposed methods, however standard methods are not sufficient to calibrate the DPSM pupil plane shift parameter. The proposed pupil calibration methods estimate the sensor depth (and pupil plane shift in the case of DPSM) based on minimising the error of a geometric cost function which describes a known metric translation between two object depths. In the case of calibrating a DPM system, only a single translation is required. However, due to the additional parameter in the DPSM system, three object depth offsets are required to recover both camera parameters. Simulated experimentation is conducted to identify the most suitable calibration setup configuration with regard to object depth offsets. Ideal calibration configurations are identified as object depths offsets which are approximately one quarter the minimum object distance. This ensures that the image pixel disparity is more easily distinguishable between the object depths. Simulated experiments also show that the multi-pupil calibra-

tion methods recover the sensor depth parameter more accurately than the standard approaches.

Section 4.4 presents the real experimentation conducted in this chapter. In order to examine the standard and pupil calibration algorithms, a ground truth sensor depth value is required. An initial experiment was undertaken to recover this value which consisted of imaging a point light source at various depths within the imaging systems' depth of field. Subsequently, standard pinhole calibration procedures along with the proposed pupil calibration methods were implemented. A ground truth value for the pupil plane axial shift of the DPSM was measured using digital calipers. The calibration results confirm that both pupil approaches are superior to the standard approaches in recovering the camera model parameters. Depth estimation experiments are presented in Section 4.3.3. DPM and DPSM imaging configurations with pupil gap distances of  $10mm$  and  $20mm$  are examined. Camera parameters estimated via the calibration experiments are used for estimating depth. Mean depth estimation errors of less than 1.4% are recorded across all experiments performed for both the DPM and DPSM calibrated configurations. Standard calibration approaches were found to produce large error at the depth estimation stage. This was due to poor sensor depth estimation. Further analysis of the experimental results led to the conclusion that aberrations present within the imaging lens were having an effect on the results. In the case of standard calibration approaches, non-axial object points will be distorted by multiple aberrations which modify their true image point location. On the other hand, multi-pupil calibration approaches deal with axial object points, which are only affected by a fixed amount of spherical aberration, regardless of object depth. Additionally, the magnitude of spherical aberration present within the system is defined by the pupil gap separation distance. In configurations with  $P_g$  set at  $20mm$ , there is considerably more spherical aberration present. However, the pupil calibration algorithm partially accounts for these effects. This can be seen in the slight variation of the pupil calibrated sensor depth value in both imaging configurations.

In conclusion, this chapter has developed new imaging models in the form of multi-pupil imaging systems. Novel calibration algorithms have been proposed which recover the camera model parameters with sufficient accuracy to acquire metric information from images. Both camera models have been tested in recovering object depth over a range of 4,  $300mm$  and have been shown to

achieve a high level of accuracy.

# Chapter 5

## Aberration Modelling, Calibration and Removal

Lens designers from all aspects of the imaging community, whether it be microscopic, telescopic, or photographic strive to design lenses which produce images of high fidelity. The main difficulty in this process is that no optical system can faithfully reproduce an image of a scene within the realm of Gaussian imaging. The departure of optical systems from ideal Gaussian behaviour is characterised by aberrations. The primary goal of this chapter is to demonstrate how multi-pupil imaging can be applied to modelling, calibrating, and removing monochromatic aberrations from optical systems.

Distortion aberration removal is a mature area with much work done in the field of camera calibration to account for lens distortion influences on camera parameter estimates (Tsai, 1987, Zhang, 1998). These methods employ grid like feature points to calibrate the distortion. Chromatic aberrations have been accounted for in a similar manner using grid point calibration objects (Mallon and Whelan, 2007*a*, Kang, 2007). However, the remaining monochromatic aberrations are dependent on pupil plane intersection as well as field location. An alternative approach to estimating all aberrations is to precisely measure the PSF, which varies significantly, across the entire image (Schuler et al., 2011). Subsequent deconvolution with the estimated PSF recovers the aberration free image. Besides the practical issues with such a calibration approach, error in the estimation of the PSF introduces ringing artifacts into the recovered image. As a result of pupil plane dependence, these aberrations

can be accounted for through a multi-pupil approach, which yields additional information about the pupil plane in the imaging process. In addition to allowing the estimation of pupil plane dependent monochromatic aberrations, the multi-pupil approach accomplishes this using point feature correspondence.

Section 5.1 introduces monochromatic aberrations and examines the implications of each on the quality of an imaging system. Section 5.2 formulates a multi-pupil approach to calibrate all monochromatic aberrations. Finally, Section 5.3 examines the accuracy of the aberration calibration removal through depth estimation and object reconstruction experimentation.

### 5.1 Optical Aberrations

There are two categories of optical aberrations, monochromatic and chromatic. Monochromatic aberrations are generally caused by the geometric structure of the lens and occur in colour systems as well as monochromatic systems. There are five monochromatic aberrations which are spherical aberration, coma, astigmatism, field curvature, and distortion. Typically, these aberrations deteriorate image quality and deform image structure. Chromatic aberrations arise from the fact that refraction is a function of wavelength. Thus, for different wavelengths of light, the angle at which a ray passes through an optical system will vary. Chromatic aberrations are usually visible as “colour fringing” in the image. Compensation of aberrations is thus the primary concern of the lens maker when forming an optical system. However, even for high end photographic lenses, a significant amount of aberrations are present as no single optimisation will adequately compensate for all aberrations within the system. This chapter is concerned with the five monochromatic aberrations and their compensation within a multi-pupil framework.

#### 5.1.1 Aberration Formation

The fundamental theory of aberrations within an optical system was developed by considering the wave nature of light. Thus, the propagation of light through an optical system is seen as a wavefront or sphere, which for ideal imaging is assumed to be Gaussian. A wavefront afflicted by aberration becomes dis-

torted, and as a result, deviates from the ideal spherical shape. This allows the aberration to be expressed in terms of the difference between the spherical wavefront and aberrated wavefront. Therefore the general equation describing aberration formation is in wave format. However, the relationship between a wavefront and geometric rays is well known. Geometric rays approximate the orthogonal trajectory of a wavefront emanating from an object point, which in turn allows an expression for aberrations in terms of geometrical rays (Welford, 1986).

Image formation within a general optical system in Euclidean space is presented in Figure 5.1. A ray emanating from the object point  $\mathbf{P}_o$  passes through the pupil plane, which has a local coordinate system of  $(\xi, \eta)$ , and under ideal imaging assumptions intersects the image plane at the point  $\mathbf{p} = (x, y)$ . In reality, the ray is aberrated as it passes through the optical system and intersects the image plane at  $\tilde{\mathbf{p}} = (\tilde{x}, \tilde{y})$ , the observed image point. An expression for the wave aberration in polynomial form is derived from the Seidel perturbation eikonal as (Born and Wolf, 1980),

$$W = -\frac{1}{4}B_1\rho^4 + B_2\rho^2\kappa^2 - B_3\kappa^4 - \frac{1}{2}B_4r^2\rho^2 + B_5r^2\kappa^2, \quad (5.1)$$

where  $r^2 = x^2 + y^2$ ,  $\rho^2 = \xi^2 + \eta^2$ , and  $\kappa^2 = x\xi + y\eta$ . The primary Seidel aberrations are represented by each coefficient in Eqn. 5.1: spherical aberration ( $B_1$ ), coma ( $B_2$ ), astigmatism ( $B_3$ ), field curvature ( $B_4$ ), and distortion ( $B_5$ ). The  $x$  and  $y$  coordinates are the ideal image coordinates in Euclidean space while the pupil coordinates  $(\xi, \eta)$  are normalised locally to the maximum pupil diameter.

The relationship between the wavefront and geometrical ray aberration allows an expression for the transverse ray aberration to be calculated. The transverse ray aberration is the displacement of the observed image point ( $\tilde{\mathbf{p}}$ ) from its Gaussian counterpart ( $\mathbf{p}$ ). This expression for transverse ray aberration is found by taking the derivative of the wave aberration polynomial with respect to  $\xi$  for  $x$  directional displacement and  $\eta$  for  $y$  displacement thus,

$$\Delta x = x - \tilde{x} = \frac{\delta W}{\delta \xi}, \quad \Delta y = y - \tilde{y} = \frac{\delta W}{\delta \eta} \quad (5.2)$$

which, in turn, yields polynomial expressions for the transverse ray aberrations.

$$\begin{aligned} \Delta x &= B_1\xi(\xi^2 + \eta^2) - B_2(3x\xi + x\eta^2 + 2y\xi\eta) + 2B_3(x^2\xi + xy\eta) \\ &\quad + B_4\xi(x^2 + y^2) - B_5x(x^2 + y^2) \end{aligned}$$

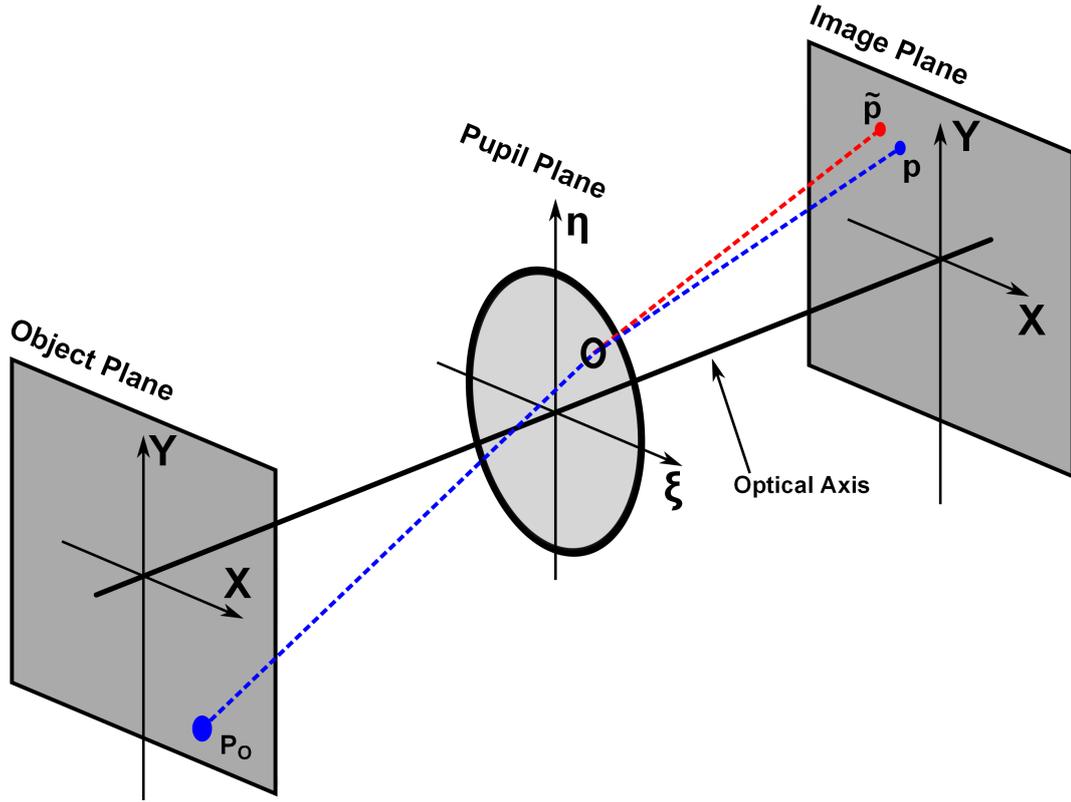


Figure 5.1: Image formation in a general optical system.

$$\begin{aligned} \Delta y = & B_1\eta (\xi^2 + \eta^2) - B_2 (3y\eta + y\xi^2 + 2x\xi\eta) + 2B_3 (xy\xi + y^2\eta) \\ & + B_4\eta (x^2 + y^2) - B_5y (x^2 + y^2) \end{aligned} \quad (5.3)$$

This expression represents a third order approximation of the transverse ray aberrations. Each aberration has a different effect on the resulting observed image point. Third order approximation is sufficient to capture the predominant effects within an imaging system, however, distortion aberration of photographic lenses typically contains a significant fifth order component ( $C_5$ ). This component can simply be added to Eqn. 5.3.

### Spherical Aberration

Spherical aberration can be described as a variation of focus with pupil plane diameter and is the only aberration which is present for on axis points. It arises due to the fact that lens surfaces are spherical in nature. The  $B_1$  term in Eqn. 5.3 is dependent solely on pupil plane position. Thus, with reference to Figure 5.1, when considering an on axis object point, the image formed for a ray which passes through the pupil plane at its periphery (large  $\xi, \eta$ ) will focus

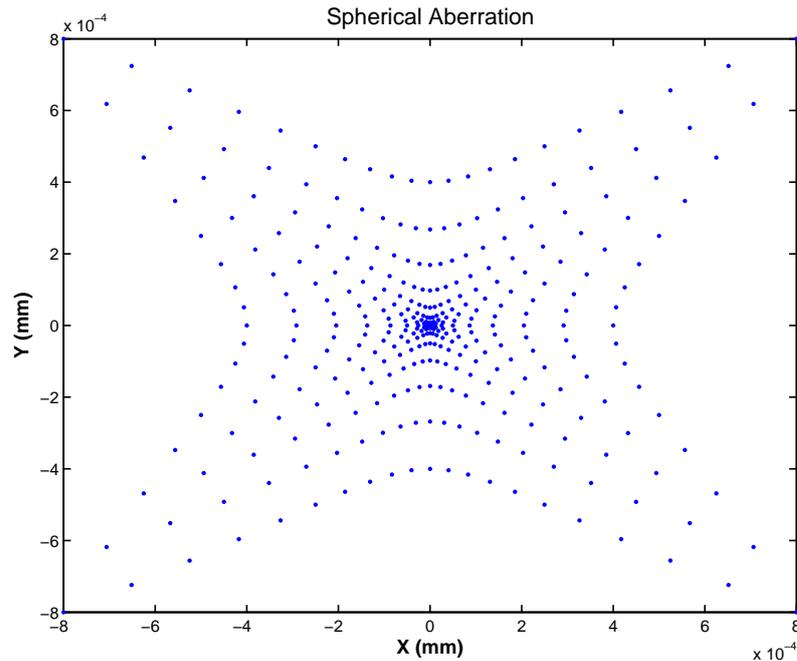


Figure 5.2: Spot diagram for spherical aberration.

on front of the image plane. In contrast, rays which pass through the pupil plane closer to its centre (small  $\xi, \eta$ ) will form a focused image on the image plane. The overall effect on the image plane, as the object point rays pass through the pupil plane with increasing  $\xi$  and  $\eta$ , is that of a formed image with a bright centre spot surrounded by subsequent halos of light. These halos correspond to the blurred image being formed by the rays which are spherically aberrated. This example is referred to as under-corrected spherical aberration. Depending on the spherical properties of the lens surface, an over-corrected aberration can also occur where the rays striking the margins of the pupil plane will focus behind the image plane. Spherical aberration is uniform across the field, thus the longitudinal focus difference between rays passing through the periphery and centre of the pupil plane does not depend on the obliquity of the incident rays.

An example spot diagram for an on axis point is presented in Figure 5.2. Each point in Figure 5.2 represents the intersection of a ray, emanating from a single on axis object point, passing through the pupil plane. The pupil plane has been sampled with a square grid. The ideal point passes through the centre of the pupil plane and intersects the image plane at (0, 0). Rays which pass through the pupil plane further from its centre intersect the image plane at greater distances from the Gaussian point.

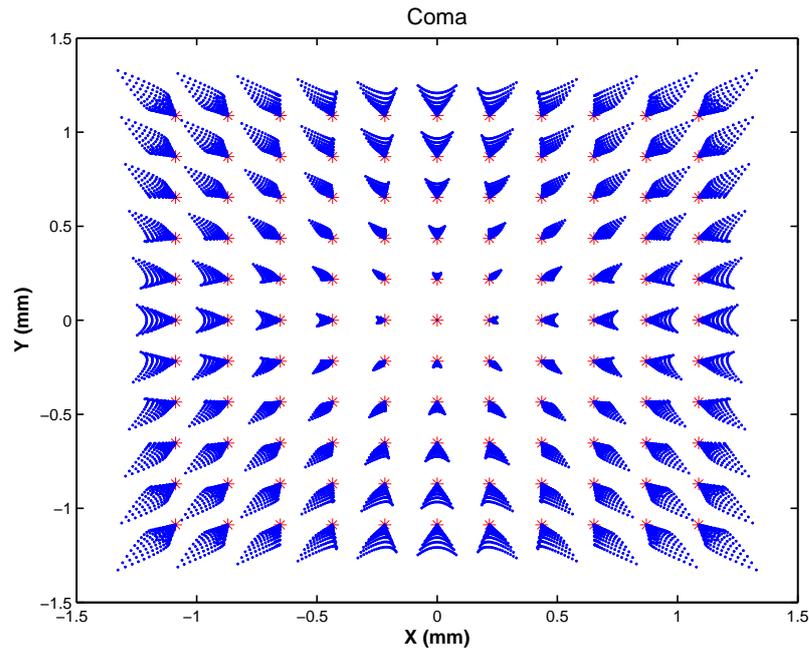


Figure 5.3: Coma spot diagrams for a grid of object points.

## Coma

Coma is represented by the  $B_2$  term in Eqn. 5.3. It is dependent on field position as well as pupil plane intersection. Coma can be defined as a variation of magnification with pupil plane. Therefore, as object height increases, the bundle of rays passing through the periphery of the pupil plane will be imaged at a different height than those passing through its centre. The magnitude of comatic aberration varies for the tangential and saggital components at the pupil plane. For an incident bundle of rays across the entire pupil plane, the tangential component of the comatic aberration is larger than the saggital component (Eqn. 5.3). However, approximately half of the entire incident light is concentrated within the ideal point to saggital area on the image plane. This difference in the saggital and tangential components gives the spot diagram its distinctive comet shape. Figure 5.3 presents a spot diagram of comatic aberration for a grid of object points passing through the pupil plane. The red asterisk points represent the ideal image corresponding to each object point. The blue points represent the intersection of the rays passing through the pupil plane (sampled with a square grid). The difference in saggital and tangential coma is most visible for the object points farthest from the optical axis, for example, the top left spot diagram in Figure 5.3.

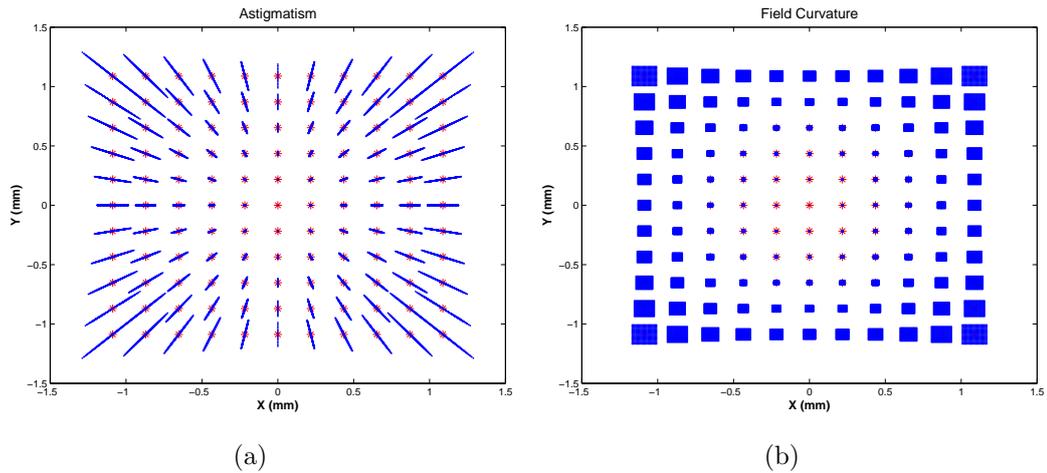


Figure 5.4: (a) Astigmatism spot diagrams for a grid of object points; (b) Field Curvature spot diagrams for a grid of object points.

### Astigmatism and Field Curvature

Astigmatism is represented by the  $B_3$  term in Eqn. 5.3. This aberration is related to the tangential and saggital planes of object space. Astigmatism occurs when the tangential and saggital images do not intersect in the same plane. In other words, the images of object points in these planes do not focus at the same point. Thus, the image of a point, when a ray bundle of either tangential or saggital nature is projected through the pupil plane, results in a line. An example of astigmatism is given in Figure 5.4(a). Each grid point is imaged as a line when projected through the (grid sampled) pupil plane.

Field curvature is represented by the  $B_4$  term in Eqn. 5.3. This aberration arises as a result of refraction in a lens and the non ideal spherical shape of lens elements in general. Thus, as object points further from the optical axis are imaged, they focus on a curved surface rather than the planar image sensor. Figure 5.4(b) presents the image of a grid of object points (red asterisk points). The blue points represent the images of the rays emanating from each object point and passing through a point on the pupil plane  $(\xi, \eta)$ . It is evident, as the object point distance from the optical axis increases, the rays sampled through the pupil plane are more spread out (follow a radial line from the centre of the image to the top left point). This has the effect of blurring the image for object points further from the optical axis.

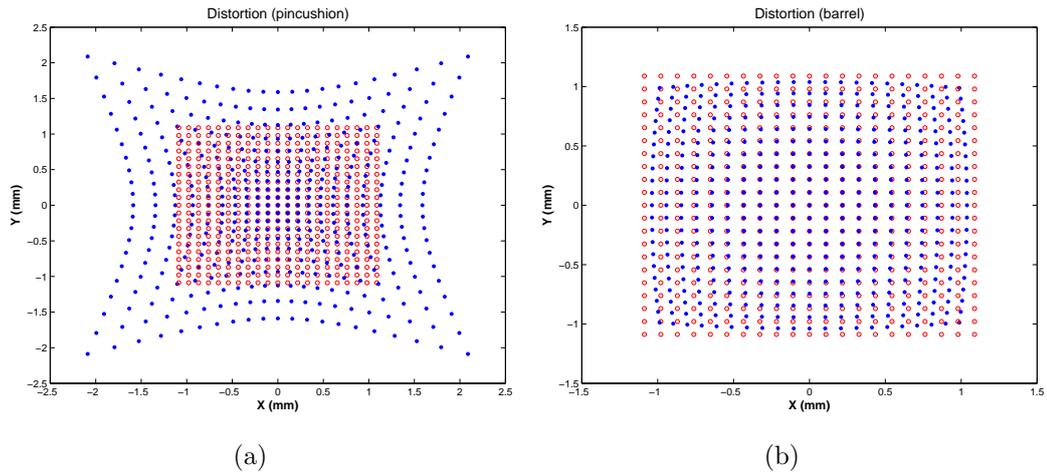


Figure 5.5: (a) Negative distortion; (b) Positive distortion.

## Distortion

Distortion is one of the more recognisable aberrations as it is typically present in most photographic quality lenses. It is represented in Eqn. 5.3 by the  $B_5$  term. The most defining characteristic of distortion aberration is that straight lines in object space are not projected into the image as straight. The magnitude of distortion is dependant on the radial distance from the optical axis to the object point. Thus object points further from the axis will be more distorted in the image. There are two main types of distortion modelled in Eqn. 5.3, they are pincushion or positive distortion, and barrel or negative distortion. Examples of both are given in Figures 5.5(a) and (b) where the ideal image points are represented by the red circles and the distorted by blue points. It is clear that the coordinates at which the object rays pass through the pupil plane have no effect on the distortion.

## 5.2 Calibration

The aberrations which have received most attention within the computer vision community are distortion and chromatic aberrations. This is primarily due to the fact that these aberrations can be well accounted for by means of calibration algorithms using point/line correspondences (Stein, 1996, Tsai, 1987, Zhang, 1998, Mallon and Whelan, 2007a, Kang, 2007). The main issue with calibration of spherical, coma, astigmatism, and field curvature is the de-

pendence of each one on the location of object ray interception with the pupil plane. Thus, feature points such as chessboard grid corners/edges and point light sources do not allow the calibration of these aberrations. It is difficult to ascertain which part of an extracted feature point corresponds to a particular area of the pupil plane. An alternative approach to removing aberration artifacts from images is to accurately estimate the PSF across the entire image. This is a difficult task and requires precise calibration in a controlled environment. Schuler et al. (2011) recently proposed such an approach to calibrate all aberrations. The main problem with methods which model the PSF is that, in order to remove the aberrations, a deconvolution operation is required. Thus, any error in the PSF estimation is amplified in the deconvolution process and results in ringing artifacts across the image. Additionally, the presence of all monochromatic aberrations results in a PSF which varies greatly across the image (as seen in Figures 5.2, 5.3, 5.4, and 5.5). Therefore multiple PSFs are required for the deconvolution process which further increases the complexity of the problem.

A key link between multi-pupil imaging and the overall effect of aberrations within an imaging system can be explored for the purpose of aberration removal. A freely located pupil is equivalent to sampling a single centred ray of the pupil plane at its location (see Section 2.3.2). Thus, when employing multiple pupils within the pupil plane, one can obtain an accurate sample on the image plane of the aberrated image points due to pupil offset. As a result, a calibration approach can be derived for correcting spherical, coma, astigmatism, and field curvature utilising grid feature points. This eliminates the requirement of accurately modelling and estimating a varying PSF across the image and allows a calibration based on grid point feature extraction of all monochromatic aberrations.

### **5.2.1 Multi-Pupil Formation**

The role of monochromatic aberrations within an imaging system with respect to multi-pupil imaging can be defined by examining Eqn. 5.3. Transverse ray aberrations, with the exception of distortion, are dependent on the location at which object rays pass through the pupil plane ( $\xi, \eta$ ). Thus, multi-pupil imaging offers an advantage over traditional methods in that the location of pupils within the pupil plane are well defined. Consequently, expressions for

the aberrations can be defined in terms of the observed and ideal image points.

A pupil plane containing three pupils is designed as shown in Figure 5.6. Since distortion ( $B_5$ ) is the only aberration not dependent on pupil location, a centrally located pupil is defined for the purpose of its calibration. The central pupil has local coordinates on the pupil plane of ( $\xi = 0, \eta = 0$ ), which on examination of Eqn. 5.3 results in an expression for transverse ray aberration of,

$$\begin{aligned}\Delta x_c &= -B_5 x (x^2 + y^2) \\ \Delta y_c &= -B_5 y (x^2 + y^2)\end{aligned}\tag{5.4}$$

where the  $c$  subscript defines the aberrations for the centrally located pinhole. Given a grid of control points in object space, the image formed via the central pupil (the observed points  $\tilde{\mathbf{p}}$ ) is given by

$$\begin{aligned}\tilde{x}_c &= x - B_5 x (x^2 + y^2) \\ \tilde{y}_c &= y - B_5 y (x^2 + y^2)\end{aligned}\tag{5.5}$$

Two additional pupils are formed on the pupil plane at an equal distance of 5mm either side of the central pupil. This allows an expression for the images formed through both pupils to be defined. The pupil plane contains a local coordinate system which is normalised to the maximum diameter (as noted in Section 5.1.1). However, in the case of multi pupils, the maximum diameter is defined by the pupil which is farthest from the pupil plane centre. Therefore, the local pupil coordinates are ( $\xi = -1, \eta = 0$ ) and ( $\xi = 1, \eta = 0$ ). Additionally, as depicted in Figure 5.6, the pupil located at  $\xi = -1$  is coloured green and the pupil at  $\xi = 1$  is red. This represents a colour filter which is placed at each pupil in order to separate the respective images formed by each pupil. Since the red and green pupil are not centrally located on the pupil plane, they form images which contain all monochromatic aberrations.

$$\begin{aligned}\tilde{x}_g &= x_g - B_1 + 3B_2 x_g - 2B_3 x_g^2 - B_4 (x_g^2 + y_g^2) - B_5 x_g (x_g^2 + y_g^2) \\ \tilde{y}_g &= y_g - B_2 y_g - 2B_3 x_g y_g - B_5 y_g (x_g^2 + y_g^2)\end{aligned}\tag{5.6}$$

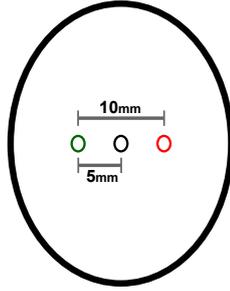


Figure 5.6: Multi-pupil formation.

$$\begin{aligned}
 \tilde{x}_r &= x_r + B_1 - 3B_2x_r + 2B_3x_r^2 + B_4(x_r^2 + y_r^2) - B_5x_r(x_r^2 + y_r^2) \\
 \tilde{y}_r &= y_r - B_2y_r + 2B_3x_ry_r - B_5y_r(x_r^2 + y_r^2)
 \end{aligned}
 \tag{5.7}$$

where the subscripts  $r$  and  $g$  represent the images formed through the pupil of that colour. The distortion term in Eqns. 5.6 and 5.7 can be eliminated by estimating  $B_5$  via the central pupil image. This leaves spherical, coma, astigmatism, and field curvature as the remaining aberrations to calibrate. This is the general approach taken with the multi-pupil calibration of monochromatic aberrations.

### 5.2.2 Multi-Pupil Calibration

Calibration of the aberrations is achieved within a multi-pupil framework by capturing a dense feature point set across the image. Assuming such a calibration object is used, and that each pupil image can be segmented from a single captured image, numerical values for the monochromatic aberrations can be calculated using Eqns. 5.5, 5.6, and 5.7. An inherent property of the aberrations is wavelength dependency. As rays pass through the multi-pupils and lenses, the angle at which they refract is a function of wavelength. Therefore, estimation of the aberration coefficients must be completed separately for each of the offset pupils.

The initial step in the calibration procedure is to separate the captured calibration image for each pupil respectively. An example calibration image is shown in Figure 5.7. There are a number of calibration parameters which are required a priori. The metric structure of the calibration plane must be known along with the object distance as well as lens focal length, sensor pixel pitch, and the metric structure of the pupil plane. It is also assumed that

the sensor depth is fixed. This knowledge of the magnification factor accounts for any image displacement due to an object depth which is not conjugate to the image plane. Taking the centre pupil image, which is effectively a pinhole image, it has been established that the sole aberration present in this image is distortion. Thus, modelling the distortion to fifth order accuracy, an error function can be formed,

$$E_i^c = \begin{bmatrix} x_{ci} - \tilde{x}_{ci} - B_5 x_{ci} (x_{ci}^2 + y_{ci}^2) - C_5 x_{ci} (x_{ci}^2 + y_{ci}^2)^2 \\ y_{ci} - \tilde{y}_{ci} - B_5 y_{ci} (x_{ci}^2 + y_{ci}^2) - C_5 y_{ci} (x_{ci}^2 + y_{ci}^2)^2 \end{bmatrix} \quad (5.8)$$

where the subscript  $c$  relates to the pupil location (centre) and subscript  $i$  indicates that it is the  $i^{th}$  observed calibration target point. The ideal image points can be calculated with the known structure of the calibration grid and its object distance. Magnification ( $m_z$ ) can be calculated as  $-f/(D-f)$  where  $f$  is focal length and  $D$  is object depth, which enables the calculation of the ideal points as  $m_z \mathbf{P}_o$ . However, there are practical issues with the orientation of the physical calibration grid in object space. Slight in plane orientation is observed (notice left side of grid in Figure 5.7 is marginally higher than the right). Therefore an adjustment of the calculated points is required. The use of planar homographies has been outlined in Chapters 2 and 3. They encode information about the planar orientation of the calibration target in space (see Eqn. 3.1). A planar homography can be calculated between the observed image pixel coordinates and the ideal points in the canonical position on the image sensor (in Euclidean space). The orientation ( $R$  component) of the actual imaged grid can subsequently be approximated by manipulating properties of the homography matrix. The orthonormal properties of the rotation element of the homography allow an approximate rotation matrix to be formed as  $R_{approx} = [h_1 \ h_2 \ h_1 \times h_2]$ . A more accurate rotation matrix is calculated using  $R_{approx}$  and forming the new rotation matrix  $R$  using SVD. The rotation matrix is subsequently decomposed into its roll, pitch, and yaw components in which the roll angle is taken and applied to the ideal image points. This ensures that the ideal points  $(x_{ci}, y_{ci})$  correspond to the true experimental data.

With the observed points  $(\tilde{x}_{ci}, \tilde{y}_{ci})$  and ideal points both represented on the image sensor in Euclidean space<sup>1</sup>, the estimation of distortion ( $B_5$  and  $C_5$ ) is

---

<sup>1</sup>Observed points are converted into sensor coordinates in Euclidean space using the calculated magnification and principal point. The optical axis is assumed to intersect the image at its centre

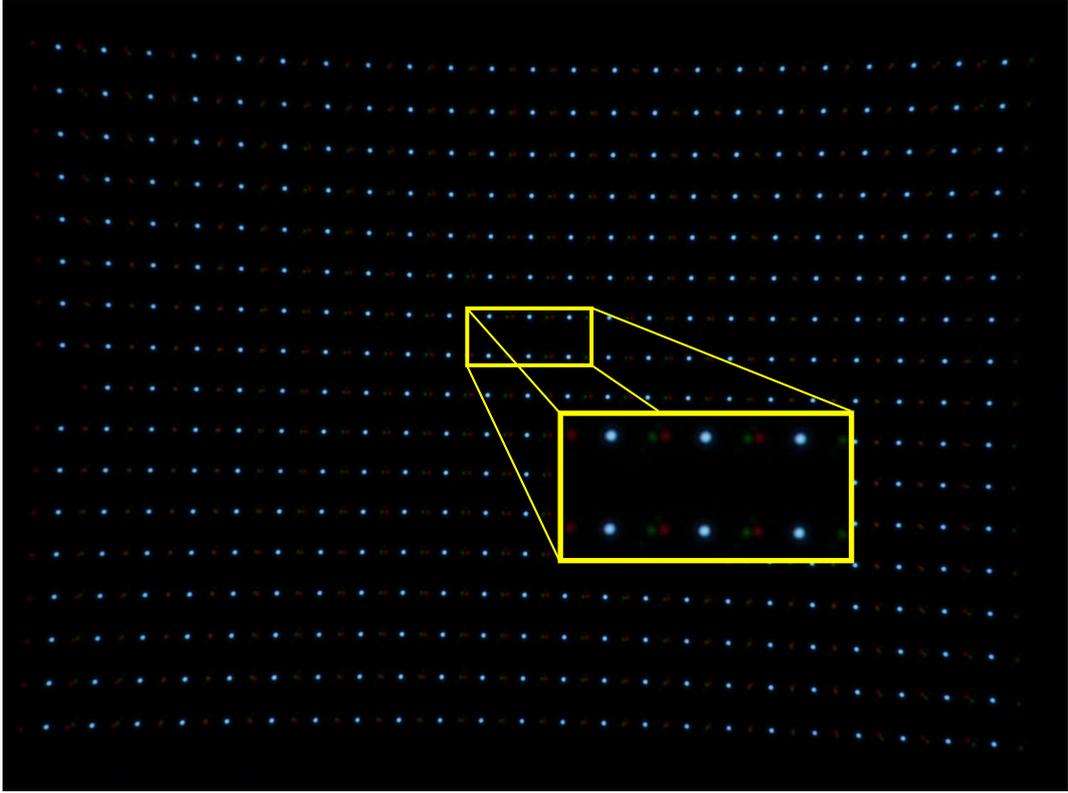


Figure 5.7: Calibration image for multi-pupil aberration calibration.

reduced to minimising the error function  $E^c$  over all  $n$  points.

$$\min \sum_{i=1}^n (E_i^c)^2 \quad (5.9)$$

Since distortion is constant across the pupil plane, the estimated distortion coefficients of the centre pupil can be used to eliminate the distortion component from the respective red and green pupil images as,

$$\begin{aligned} \hat{x}_{pupil} &= \frac{\tilde{x}_{pupil}}{1 - B_5 x_{pupil} (x_{pupil}^2 + y_{pupil}^2) - C_5 x_{pupil} (x_{pupil}^2 + y_{pupil}^2)^2} \\ \hat{y}_{pupil} &= \frac{\tilde{y}_{pupil}}{1 - B_5 y_{pupil} (x_{pupil}^2 + y_{pupil}^2) - C_5 y_{pupil} (x_{pupil}^2 + y_{pupil}^2)^2} \end{aligned} \quad (5.10)$$

where the *pupil* subscript indicates either the red or green pupil image. Using the undistorted pupil images  $\hat{p}_r$  and  $\hat{p}_g$ , the four remaining aberrations can be estimated by forming the appropriate functions ( $E_i^{pupil}$ ) and minimising the error similarly to Eqn. 5.9 using standard non-linear minimisation techniques.

$$E_i^g = \begin{bmatrix} x_{gi} - \hat{x}_{gi} - B_{1g} + 3B_{2g}x_{gi} - 2B_{3g}x_{gi}^2 - B_{4g}(x_{gi}^2 + y_{gi}^2) \\ y_{gi} - \hat{y}_{gi} - B_{2g}y_{gi} - 2B_{3g}x_{gi}y_{gi} \end{bmatrix} \quad (5.11)$$

$$E_i^r = \begin{bmatrix} x_{ri} - \hat{x}_{ri} + B_{1r} - 3B_{2r}x_{ri} + 2B_{3r}x_{ri}^2 + B_{4r}(x_{ri}^2 + y_{ri}^2) \\ y_{ri} - \hat{y}_{ri} - B_{2r}y_{ri} + 2B_{3r}x_{ri}y_{ri} \end{bmatrix} \quad (5.12)$$

Abberation removal is subsequently performed given the calibrated values of  $B_1$ ,  $B_2$ ,  $B_3$ , and  $B_4$  for both red and green pupil images. This yields the corrected image points  $\bar{p}$ ,

$$\begin{aligned} \bar{x}_g &= \frac{\hat{x}_g + B_{1g} + B_{4g}\tau}{1 + 3B_{2g} - 2B_{3g}x_g - B_{5g}\tau - C_{5g}\tau^2} \\ \bar{y}_g &= \frac{\hat{y}_g}{1 - B_{2g} - 2B_{3g}x_g - B_{5g}\tau - C_{5g}\tau^2} \end{aligned} \quad (5.13)$$

and,

$$\begin{aligned} \bar{x}_r &= \frac{\hat{x}_r - B_{1r} - B_{4r}\tau}{1 - 3B_{2r} + 2B_{3r}x_r - B_{5r}\tau - C_{5r}\tau^2} \\ \bar{y}_r &= \frac{\hat{y}_r}{1 - B_{2r} + 2B_{3r}x_r - B_{5r}\tau - C_{5r}\tau^2} \end{aligned} \quad (5.14)$$

where  $\tau = (x_{pupil}^2 + y_{pupil}^2)$ .

In summary, the calibration procedure can be outlined as,

1. Separate images according to pupil (colour plane)
2. Use centre pupil to calibrate distortion (Eqn. 5.9)
3. Undistort red and green pupil images (Eqn. 5.10)
4. Estimate spherical, coma, astigmatism, and field curvature for red and green pupils by minimising Eqns. 5.11 and 5.12. The LM algorithm is used to perform the minimisation with parameters initialised at zero.

## 5.3 Experiments

Experiments are conducted to validate the multi-pupil approach to modelling, calibrating and removing aberrations. Quantitative results are obtained by correcting aberrated images and subsequently estimating the depth of each observed point using the theory developed in Chapter 4.

### 5.3.1 Experimental Setup

The experiments carried out in this section are similar in setup to those outlined in Chapter 4. The Panasonic DMC-G1 was used in conjunction with the EO 75mm focal length achromatic lens. The pupil plane is designed as depicted in Figure 5.6. EO colour filters (red and green KODAK WRATTEN filters) are attached on front of the off centred pupils. The experimental setup at the pupil plane is shown in Figure 5.8(a). In order to calibrate all monochromatic aberrations in the imaging system, a dense set of feature points is required in object space. Additionally, the metric structure of the calibration target must be accurately known. The calibration target used in all experiments is shown in Figure 5.8(c). It consists of an LCD monitor (19 inch) in which the LC component (and outer frame) has been removed, leaving a screen which acts as a light box due to the high (and even) illumination needed for LCD screens. Feature points are subsequently created using an opaque thin plastic, which contains an adhesive side, and attaches to a planar surface. The LPKF ProtoMat<sup>®</sup> C30/S prototyping machine is used to manufacture pinholes of diameter  $0.3\text{mm}$ . A rectangular grid of  $23 \times 17$  pinholes is constructed with equal separation in both directions of  $12.5\text{mm}$ . The pinhole grid is subsequently attached to a glass surface and mounted directly on front of the modified monitor. In order to enable accurate alignment of the grid with the optical axis, the entire unit is mounted on a tripod with three degrees of freedom (Figure 5.8(c)).

Orthogonal alignment of the planar calibration target with the optical axis is required to ensure that each object point is at a single object depth. Therefore, an alignment process is repeated for each object depth whereby a laser aligned with the optical axis of the imaging system (see Chapter 4 Section 4.4.1) is used to orthogonally align the planar grid. This is achieved by attaching a planar reflective material to the rear side of the planar target. Pitch, yaw, and roll adjustments are subsequently applied to align the reflected laser spot with the emitting point. Figure 5.8(b) presents an example of the laser alignment process. Calibration images are captured as shown in Figure 5.8(d) where, after each alignment, object depth is referenced with respect to the range tape.

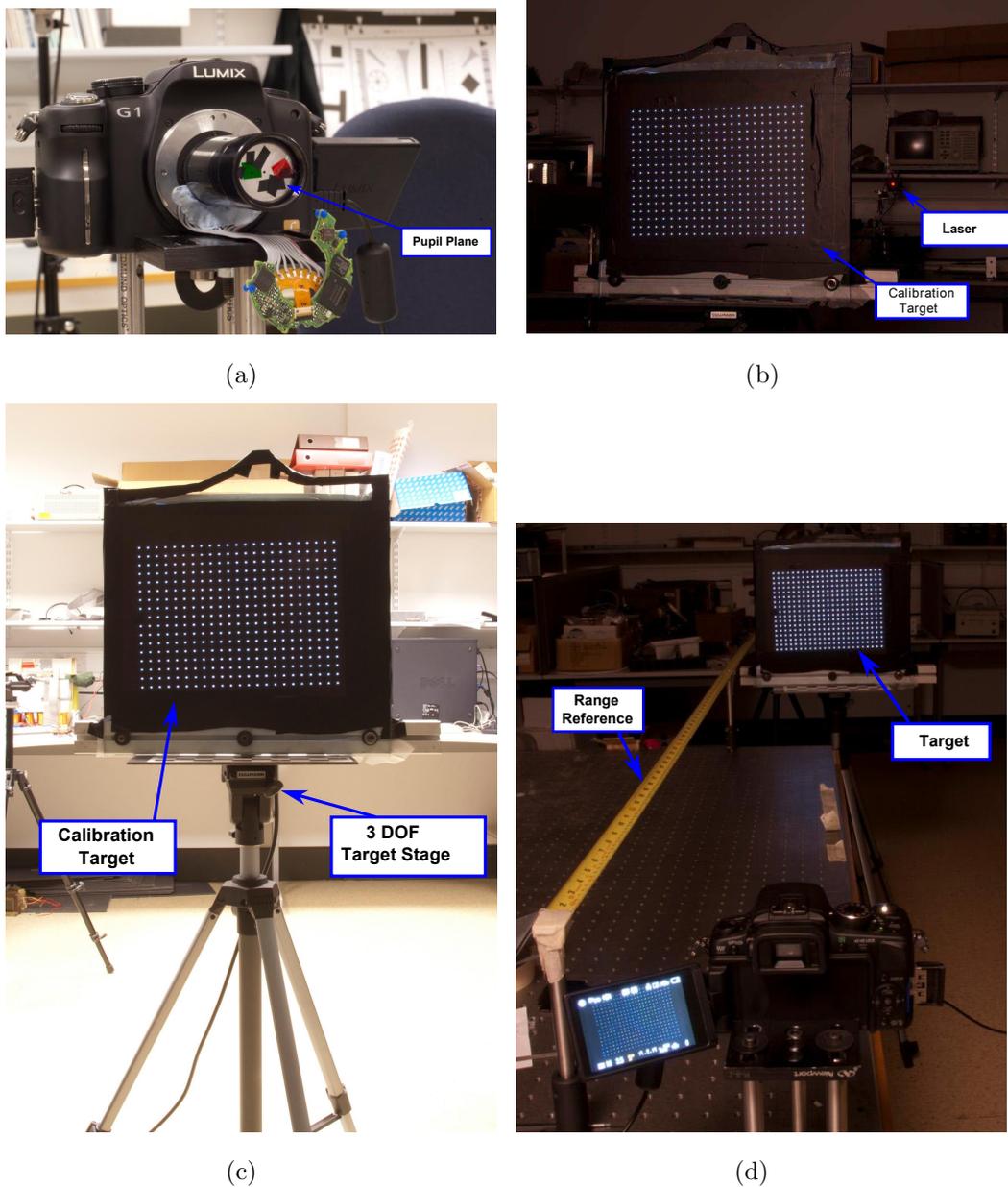


Figure 5.8: (a) Camera and pupil plane formation; (b) Alignment of Target; (c) Target and stand; (d) Calibration setup.

### 5.3.2 Depth Estimation

Images of the calibration target were acquired at five separate depths between approximately 870mm and 2,000mm. A precise range of object depths was chosen because it is necessary to fill the image sensor with feature points for the calibration process. This allows the aberration affects across the image to be captured. The procedure outlined in Section 5.2.2 is followed. Thus, images are initially separated into pupil images and feature points are extracted and

ordered in a traditional manner. Distortion parameters are estimated via the centred pupil (Eqn. 5.9) and the remaining monochromatic aberrations are estimated from the red and green pupil images (Eqns. 5.12 and 5.11).

The accuracy with which the aberration coefficients are estimated is evaluated by reconstructing the planar target in object space. Object depth is calculated in a multi-pupil framework fashion (see Chapter 4). Hence, with the known structure of the pupil plane, focal length, pixel pitch, and sensor depth (calibrated value), object depth can be calculated using the red/green pupil disparity (Eqn. 4.4). World  $x$  and  $y$  locations are reconstructed using the centred pupil image (after undistortion) and the known magnification factor (from the calibration phase). Image disparity, due to the red/green pupils is calculated after aberration removal (Eqns. 5.14 and 5.13).

Depth estimation results are presented in Figures 5.9(a)-(e). The 3D points are reconstructed from the original aberrated images (blue points), ground truth (red points), and estimated points from the aberration corrected images (green points). Subsequently, a best fit plane is calculated for each set of 3D points using a least-squares fit (plane colour matches points). It is clear that there are significant aberrations present as indicated by the blue points. As expected the centre of the image contains the least amount of aberration (only spherical), however, as the object point increases its distance from the optical axis, the depth varies by up to a metre in some cases. The aberration corrected images produce depth estimates which are in better agreement with the ground truth values. Figure 5.11 shows the Root Mean Square Error (RMSE) of depth calculated for the corrected and un-corrected images at each depth. The maximum RMSE error for the corrected images is 41.93mm (at 1,727.1mm) in comparison with 625.48mm (at 1,727.1mm) for the un-corrected set. The minimum error is observed at the nearest depth (878mm) where the RMSE is 7.46mm for the corrected and 142.51mm for the uncorrected.

Figures 5.10(a) and (b) present all depth estimates for the ground truth and corrected planes. The primary contributor to the error in depth estimation (of the corrected images) is the correction of points at the periphery of the object grid. This can be partly explained by the process in which ideal grid point alignment is achieved as outlined in Section 5.2.2. The orthogonal alignment of the calibration grid with the optical axis is likely to induce slight pitch/yaw variation. Using the rotation matrix estimated from the planar ho-

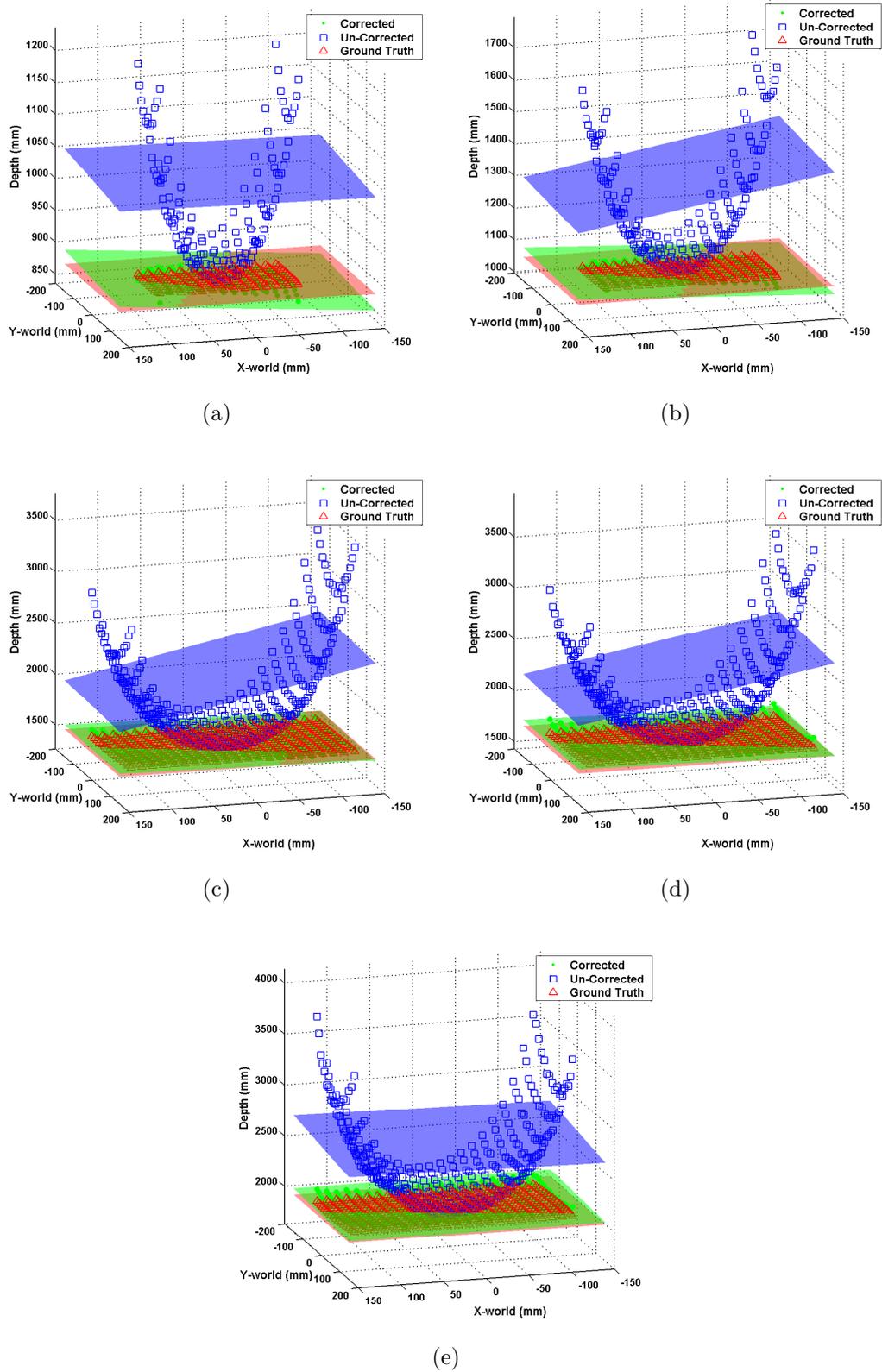


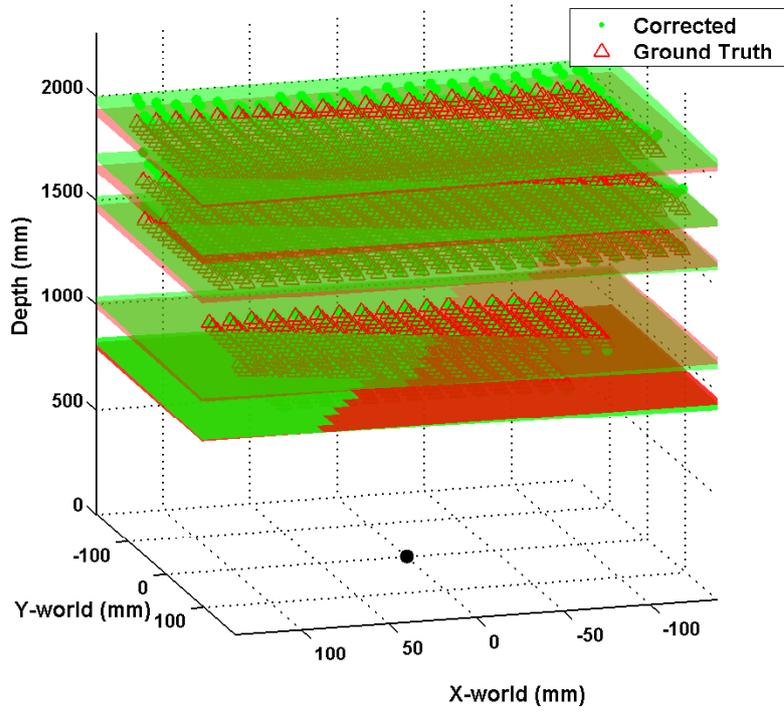
Figure 5.9: Depth estimation and best fit plane (a)  $878mm$ ; (b)  $1070.2mm$ ; (c)  $1535.2mm$ ; (d)  $1727.1mm$ ; (e)  $1995.2mm$ .

mography, this variation was found to be minimal ( $<0.01^\circ$ ) however, it is likely to marginally increase/decrease the actual depths of the peripheral grid points in the calibration process. Secondly, translational adjustments are not accounted for in the orientation alignment process. Although the laser is aligned with the optical axis, its primary function is to provide a means to orthogonally align the calibration target with the optical axis. Thus, it can not be used to align the centre of the calibration target with the optical axis (centre control point with image centre). Therefore, ideal image points are aligned with the observed image points by translating the ideal grid to the observed image point nearest to the optical centre. This translational adjustment will affect the distortion calibration since distortion increases with radial distance from the optical centre. Thus, an increase in error is expected at the periphery of the grids, which is the primary source of error in the corrected depth estimation in Figure 5.11.

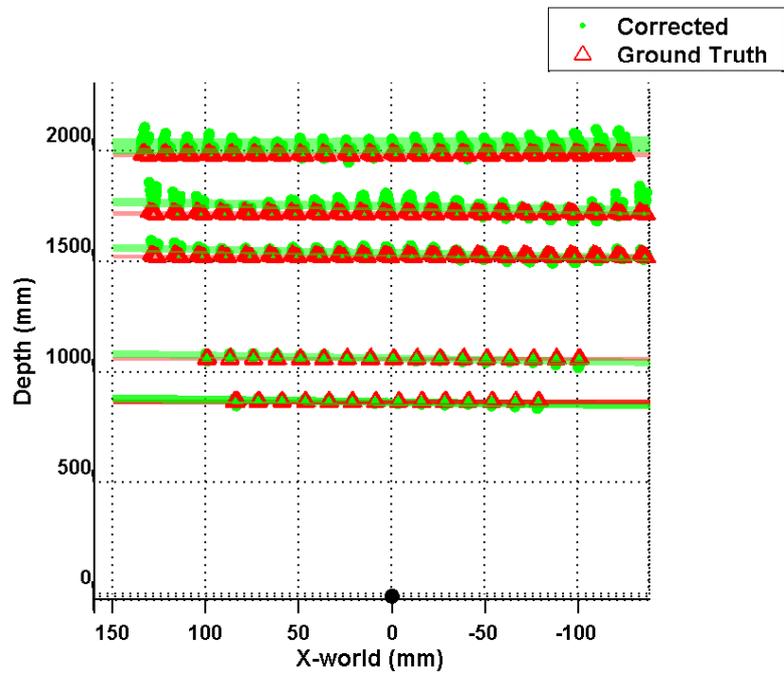
## 5.4 Discussion

This chapter highlights the benefits of applying a multi-pupil approach to calibrating monochromatic aberrations present in imaging systems. Multi-pupil imaging holds a key advantage over traditional aberration removal methods in that it provides a means to calibrate all monochromatic aberration using point correspondences. This is achieved with the additional aberration information that is available due to the structure of the multiple pupils in the pupil plane. Aberrations which are dependent on the pupil plane (spherical, coma, astigmatism, and field curvature) can be accurately estimated using a third order model. A pupil plane containing three pupils is examined in this chapter. Object reconstruction is performed after aberration correction using a centred pupil for world  $(x, y)$  locations with object depth estimated employing the remaining pupils within the multi-pupil framework.

The formation of monochromatic aberrations is presented in Section 5.1.1. An expression for transverse ray aberration is derived from the general wave equation of aberrations in an optical system. The dependence of certain aberrations on the location at which a ray from the object passes through the pupil plane is outlined. Subsequently, the affect each aberration has on image formation is examined. Simulated experiments show how a grid of object points are imaged



(a)



(b)

Figure 5.10: (a) Estimated depths (green); (b) Side view of (a).

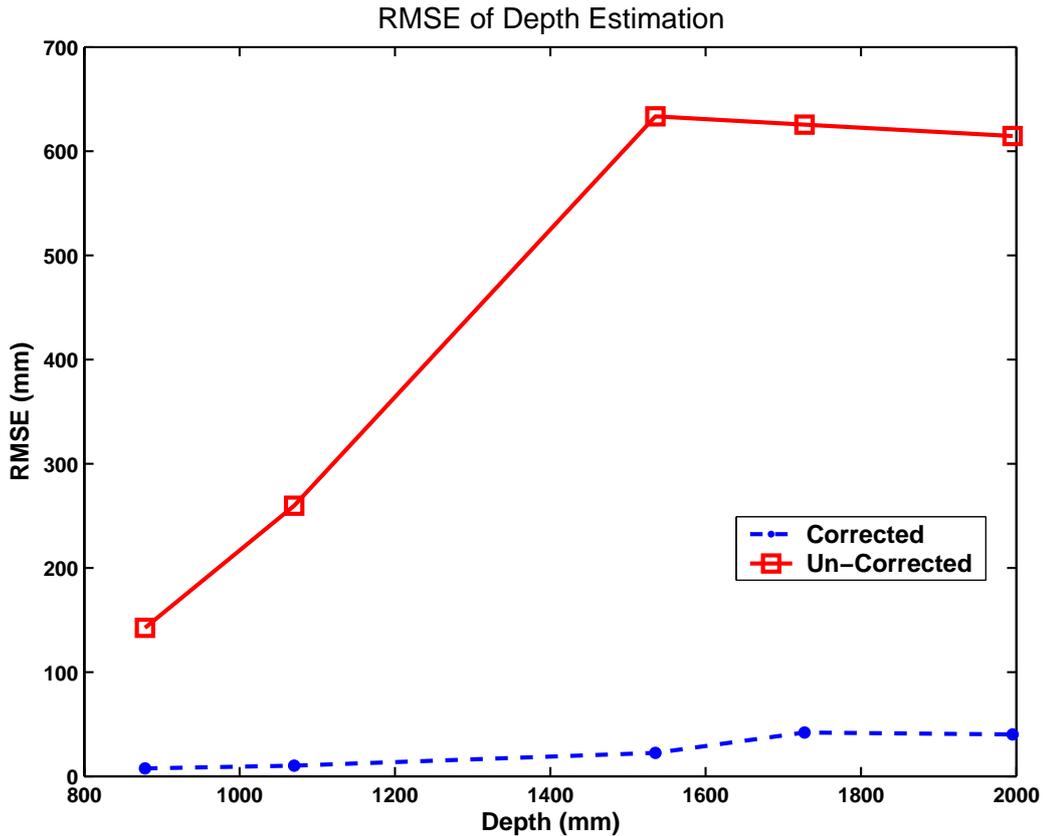


Figure 5.11: RMSE of Depth for corrected vs un-corrected images.

when subjected to each aberration individually.

Calibration and removal of optical aberrations within a multi-pupil framework is introduced in Section 5.2. A pupil plane consisting of three pupils is designed. Two of the pupils are located symmetrically about the third which is placed at the centre of the pupil plane. Colour filters are attached to the symmetrical pupils which enables separation of the images corresponding to each of the three pupils. Using the third order model for aberrations introduced in Section 5.1.1, an image formation model for each pupil is formed (Eqns. 5.5, 5.6 and 5.7). A calibration procedure is devised in Section 5.2.2 where a planar grid of control points is imaged at a known object depth. The centre pupil is used solely for the calibration of distortion. This reduces the complexity of the model equation for each of the off-centre pupils. Spherical aberration, coma, astigmatism, and field curvature are subsequently estimated by minimising the error in Eqns. 5.11 and 5.12.

Experiments are conducted in Section 5.3 to assess the quality of the aberration calibration and removal. The accuracy of this process is quantified by recon-

structing the object plane in world space. The calibration object was placed at five object depths in the range 878mm to approximately 2,000mm. After calibration and aberration correction of each image, object depth is estimated using the off-centred pupils while spatial  $(x, y)$  information is reconstructed using the centred pupil. Depth estimation results using the multi-pupil approach confirm the accuracy of the aberration removal which is shown to have 7.46mm RMSE for the near object depth and 40.11mm RMSE for the farthest depth.

The work in this chapter has highlighted the benefits of adopting a multi-pupil imaging approach for aberration removal. Firstly, all monochromatic aberrations can be calibrated using point correspondences, and secondly, object reconstruction can be achieved with only a single image.

# Chapter 6

## Conclusions and Future Work

Extracting metric information from images remains of significant interest to the computer vision community, and as a result, new approaches to obtaining this information are constantly being developed. Improvements in image sensor technology are enabling the development of new imaging techniques which capture data in a non-conventional manner and subsequently take advantage of the increased processing power available to render a new image or new image data. In addition to improvements in speed, the physical size of image sensors has reduced significantly which has advantages and disadvantages. The reduced footprint has allowed access to imaging as a viable solution for tasks which require metric information that previously availed of alternative techniques. However, with reduction in size, optical quality of the imaging system has decreased. Therefore, researchers are left with the task of developing techniques for a range of new imaging modalities whilst having to deal with imaging systems afflicted by optical inadequacies. This thesis addresses these issues with the development of a new imaging framework which allows the recovery of metric information and corrects for optical aberrations present within the imaging system.

In order to facilitate the development of the new imaging framework, limitations of current camera models and their calibration is investigated. It is evident from the literature that there is a trade-off between flexibility and accuracy when the task of modelling and calibrating a camera arises. Consequently, the simple pinhole model is consistently used for common computer vision tasks. Similarly, the popularity of planar camera calibration methods

within the computer vision community demonstrates the desire of practitioners to calibrate a camera with sufficient accuracy and little effort. The initial work in this thesis examines properties of the pinhole and lens models as well as the calibration of these models. Subsequently, an approach is proposed which improves the accuracy of planar camera calibration whilst simultaneously reducing the input requirements of the practitioner.

Computational cameras, or non-conventional imaging models, have evolved with improved image sensor technology. Capturing images in a non-conventional manner has redefined the possibilities of what type of information can be acquired from images. One of the drawbacks of implementing a pinhole or standard lens model is that the acquisition of metric depth information typically requires multiple images or sensors. Thus, the multi-pupil imaging framework developed in this thesis takes advantage of the non-conventional formation of image data to encode additional information for the retrieval of metric depth from a single image. Furthermore, this new framework contains inherent properties which enable it to deal with issues of optical quality within imaging systems. Consequently, new calibration approaches are developed, and demonstrated in this thesis, which enable the recovery of camera model parameters as well as correcting for optical aberrations present in the system. The results presented in this thesis have significant implications for the development of camera models in the fast changing environment of image sensor technology. The multi-pupil imaging framework offers a flexible solution to metric information retrieval, particularly in the area of compact imaging solutions which naturally contain optical systems of lesser quality.

The contributions and outcomes of the thesis are summarised in Section 6.1 and a list of publications resulting from the thesis work is presented in Section 6.2. Possible directions for further work are outlined in Section 6.3.

### 6.1 Thesis Contributions

This thesis deals with both traditional and non-conventional camera models as well as their calibration. For each problem a solution is proposed, evaluated through real and simulated experimentation, and then validated by the results from these experiments. Integral to these solutions are other less significant

contributions. The contributions are summarised in the following sections.

### 6.1.1 Efficient Camera Calibration

Planar camera calibration has been established as the dominant and preferred type of calibration for cameras modelled as pinhole. Compared to the traditional techniques, which require high precision calibration objects and setup, input requirements are relaxed. Imaging a plane in a minimum of three unknown orientations provides full recovery of all intrinsic and extrinsic camera parameters. Chapter 2 presents the foundations on which planar camera calibration is formed. A geometric invariant in the form of the IAC is the tool used to perform the calibration. The camera intrinsics are recovered by estimating the IAC using planar homographies between the calibration plane and its images. Five points are required to uniquely define a conic, therefore, three images provide sufficient constraints. A minimal case of two images (providing four points) can be used with a zero skew constraint. An alternative geometric interpretation in the form of the centre circle constraint is presented in Chapter 2. Although based on the constraints of the IAC, it provides a more intuitive framework in which to analyse the calibration method.

The sensitivity of planar camera calibration is investigated and outlined in the initial part of Chapter 3. Degenerate configurations correspond to planar target orientations which do not contribute independently to estimating the IAC, thus, the calibration fails. Additionally, planar target orientations which are close to degenerate contribute to poor intrinsic estimates. A key link between these configurations and the estimation of the IAC is presented. It is shown that a real geometric entity, in the form of the image CL can be utilised for the identification of such configurations. Conducted experiments demonstrate that, in the two-plane case, properties of the image CL can be used to detect the manifold of all degenerate configurations.

Consideration of the planar target orientation within the network of calibration images is an area which has received little attention. As a result, calibration practitioners often capture many input images which are, in effect, redundant in the camera parameter estimation stage. The latter part of Chapter 3 introduces a novel framework for the formation of input images for efficient planar camera calibration. Image networks which contain ideal geometry for estimat-

ing a unique IAC are proposed. The method focuses on increasing camera parameter estimation accuracy whilst reducing the input requirement load. This is achieved by enforcing geometric constraints on image CLs. GINs, presented in Section 3.2.1, require an initial input calibration image supplied by the user. Based on the geometric configuration of this seed image CL, ideal synthetic calibration images are formed for user replication and subsequent calibration. An alternative approach based on an image selection strategy is proposed in Section 3.2.2. This method, SIN, searches through a large IN and selects the images which contain the most favourable geometry for the camera parameter estimation process.

Extensive experimentation was conducted to highlight the advantages of using GINs and SINs over conventional image network approaches. SIN experiments show an overall increase in accuracy of camera parameter estimates for INs containing less images than the standard approach. Distortion correction is also performed which demonstrates the superior intrinsic parameter estimation using SINs. Section 3.3.2 details experiments conducted utilising GINs. The accuracy to which non-expert practitioners can replicate the synthetic images is examined and shown to be sufficient to obtain accurate camera parameters. Increased accuracy in the calibration of the intrinsic parameters is observed for image networks containing less than 5 images. Additionally, the stability of the intrinsic parameter estimations are evaluated. Increased stability is observed using GINs over the conventional technique. A key advantage of the GIN method is that, as well as specifying optimal planar target orientation, it also aids the practitioner in capturing this configuration.

### 6.1.2 Multi-Pupil Imaging

Certain scene information is lost in the imaging process using the pinhole model, specifically object depth information. There is scope within the context of computational/non-conventional imaging to encode this information in a suitable manner for subsequent retrieval. However, this requires the formation of a new imaging model. Multi-pupil imaging is such a model and the foundations upon which it is built are described in Chapter 2 and further developed in Chapter 4.

The equivalence of the pinhole model and lens models is outlined in Chapter

2. It is seen that with the introduction of a lens to the imaging system, optical phenomena which provide depth cues in captured images are available which are not modelled via a pinhole. Specifically, depth dependent blur can be modelled with a lens model. Consequently, it is the pupil plane within an imaging system which defines the quantity of blur introduced to an imaging system. Experimentation, using Zemax, was conducted in which both the location of the pupil plane within an imaging system, and the location of a pupil within the pupil plane, were examined in terms of the effect each had on the system PSF. Two conclusions were drawn from these experiments. Firstly, a pupil plane located in object space is desirable due to the sensitivities of placement in image space. Furthermore, image space placement results in an entrance iris which is not proportional to the magnitude of blur observed on the PSF. Secondly, it is shown that the pupil plane can be sub-sampled by use of smaller pupils (fractional in diameter compared to the conventional pupil) which results in accurate image observations of pupil ray data.

The multi-pupil imaging framework is proposed, utilising both pupil plane results, in Chapter 4. Multi-pupil imaging is achieved by modifying the pupil plane of an imaging system. The DPM, introduced in Section 4.1, consists of a pupil plane placed at the lens, and contains two pupils. A relationship for object depth is developed based on the structure of the pupil plane and imaging system properties. This allows depth information to be recovered from a single image of a scene. Extensive model characterisation is carried out using simulated data, in which the sensitivities of all model parameters to depth estimation in the presence of Gaussian noise are evaluated. This allows the identification of optimal imaging configurations, in a multi-pupil framework, for depth estimation. The DPSM presented in Section 4.2 relaxes the constraints on pupil plane location. Thus, the pupil plane can be shifted axially in object space. A depth estimation equation is derived from the model parameters, and similarly to the DPM, comprehensive model characterisation is conducted to validate the model.

The importance of camera calibration has been highlighted throughout this thesis. In this regard, novel calibration approaches to determine the multi-pupil imaging parameters required for depth estimation are presented in Section 4.3. The calibration object/target consists of object points situated on the optical axis of the imaging system. In the case of the DPM, two object points with known translation distance (between both points) are required.

For DPSM calibration, four object points with known translation are required (due to the additional model parameter for pupil plane location). Multi-pupil calibration is benchmarked against standard approaches. Extensive experimentation and evaluation using real data is presented in Section 4.4. The multi-pupil calibration approach is shown to outperform standard approaches. Subsequent depth estimation is performed using the calibrated model parameters for each calibration approach. There are two conclusions to be drawn from the depth estimation results. Firstly, they validate the multi-pupil imaging approach for depth estimation. Secondly, they confirm the accuracy of the proposed multi-pupil calibration procedure. The primary contribution is the ability to accurately estimate object depth from a single multi pupil image.

### 6.1.3 Aberration Modelling, Calibration, and Removal

Despite the best efforts of lens designers and manufacturers, residual optical aberrations are present in all imaging systems. This has the effect of deviating object rays from their true path under ideal imaging conditions. As a result, the formed image is typically blurred in certain areas and generally deformed. With the growing use in the smartphone and tablet market, imaging systems are undergoing a transformation which requires high resolution cameras that are extremely compact. Naturally, there is a trade-off with the optical quality of the camera. Consequently, there is an increasing need for software correction solutions. Standard approaches to removing optical aberrations from images require precise modelling of the system PSF. Depending on the aberrations present within the imaging system, the PSF can vary greatly across the image. Thus, errors in the PSF estimation are amplified when the corrective deconvolution process is applied to the image (causing ringing artifacts). Other approaches make use of standard control points in object space. However, due to the nature of certain aberrations, which depend on pupil plane location, correction can not be achieved for all aberrations.

The outlined problems with aberration removal can be addressed with the implementation of multi-pupil imaging. Chapter 5 demonstrates how multi-pupil imaging can be applied to modelling, calibrating, and removing optical aberrations from imaging systems. With the multi-pupil approach, aberrations which are pupil plane dependent can be accurately modelled due to the unique structure of the modified pupil plane. This enables the forma-

tion of an accurate multi-pupil plane model of all monochromatic aberrations. Section 5.2.2 outlines a novel aberration calibration approach which utilises object control points to estimate all monochromatic aberrations. This allows a feature-correspondence-like calibration, an approach which was not previously possible. Experimentation is conducted using a grid of control points captured over a range of object depths. The accuracy of the calibration and subsequent aberration removal is quantified by performing object depth and structure estimation. It is concluded that the application of the multi-pupil framework to aberration modelling and calibration is a viable approach to aberration removal.

## 6.2 Publications Arising

All publications are full length papers that have been peer reviewed.

**Efficient Planar Camera Calibration via Automatic Image Selection**, Byrne, B. P., Mallon, J. and Whelan, P. F., ‘Proceedings of the 4th International Conference on Computer Vision Theory and Applications, Lisbon, Portugal’, Vol. 1, pp. 90–94, 2009.

**Optimal Image Networks for Planar Camera Calibration**, Byrne, B. P., Mallon, J. and Whelan, P. F., ‘Proceedings of the IEEE Irish Machine Vision and Image Processing Conference, Dublin, Ireland’, 2011. (Best Paper Award)

## 6.3 Directions for Future Research

Each of the main topics investigated in this thesis has potential for further examination in order to achieve additional improvements. Some ideas for these directions are outlined below.

### 6.3.1 Camera Calibration

Optimal image networks have been outlined in Chapter 3 and shown to obtain camera parameter estimates with greater accuracy and stability than those calibrated with conventional imaging strategies. Although this image network strategy is defined for constant internal parameters, it would be a natural progression to consider varying internal parameters. Thus, changes in sensor depth and principal point such as those seen with zoom lenses could be calibrated with a single data set, rather than calibrating each zoom setting separately. One of the main benefits of optimal image network strategies is the simplification of the calibration image capturing process. Given that synthetic poses (with ideal target orientation) are provided to the user, the replication process lends itself well to alternative imaging media such as smartphones and tablets. Coupled with adjustments for varying intrinsic parameters, an optimal image network application could be developed for such imaging media with a view to providing accurate camera parameter estimates for mobile computer vision applications.

### 6.3.2 Multi-Pupil Imaging

Experimentation and evaluation of the multi-pupil imaging models was conducted using pupils with a diameter of 1mm. Therefore, a longer than normal exposure was required to form the multi-pupil images. An interesting route of exploration would be to replace the pupil plane with a programmable LCD or similar. With a programmable pupil plane, two possible approaches could be taken to increase the intensities captured at the image plane. Firstly, a time coded pupil plane could be deployed within a single exposure. If one of the time coded pupils had the form of a conventional pupil (at the maximum diameter) it would provide additional scene radiance to the image sensor. Using a multi-pupil within the exposure period would thus maintain the depth dependent disparity and allow the multi-pupil imaging model to be applied for depth estimation. However, a significant challenge with this approach would be the segmentation of conventional pupil data from the multi-pupil data as there would be an overlap of frequency information captured from both pupils. The second possible route using a time coded pupil plane is to capture two separate images. An image would be captured for each pupil within the pupil

plane. Although this approach requires two images, no movement of the camera is necessary and the issues of segmentation are avoided. Therefore, it still maintains advantages over traditional stereo techniques.

The method of aberration calibration and removal outlined in Chapter 5 uses colour filters (red and green) to aid in the identification of image features which correspond to projection through certain pupils of the pupil plane. Therefore, the calibration procedure requires estimation of the aberrations for each wavelength. Consequently, the aberration removal process must also be applied to each pupil image separately. This is a direct result of chromatic aberration present in imaging systems. Therefore, an interesting extension to the methods described in Chapter 5 would be to include additional pupils within the pupil plane for the purpose of calibrating chromatic aberration. This would enable a single estimation of the aberration parameters, and similarly, a single operation to remove all pupil plane dependent aberrations from the image.

Finally, it is envisaged that the multi-pupil imaging framework could be applied in many fields which require depth estimation and metric measurement. Since the imaging model generates depth dependent disparity based on the focusing properties of an imaging system, it is highly scalable. This would allow the implementation of multi-pupil imaging systems in areas which use physically small sensors, such as medical endoscopy, and physically large configurations, such as industrial robotics used for line automation.

## Appendices

# Appendix A

## Additional OIN Results

An experiment was conducted where a single calibration image was captured and its CL angle calculated. Subsequently, a number of images were manually selected (from a large database of calibration image captured with the same camera and same settings) with CLs corresponding to the ideal angle, which in the case of a two image network is 90 degrees, and also with CLs corresponding to non-ideal angles. Tables A. 1 and A. 2 present the calibration results obtained with image networks containing favourable geometry (ideal and close to ideal angle between image CLs) and the networks with non-optimum angles between the image CLs. These parameter estimates with ideal and close to ideal CL orientation produce stable calibration results compared to those achieved with the non-optimum orientations. These indicative tests demonstrate the benefits of considering the image CL as a guide to selecting calibration images and avoiding poor image network geometry.

Table A.1: Results for calibration with image networks with non-optimum angle between CLs. Parameter estimations and SD given in pixels

Angle (deg)	$u_0$	$v_0$	$f_u$	$f_v$
5	808.2	617.4	1796.5	1794.1
10	806.9	607.7	1789.4	1787.9
15	812.2	638.6	1808.5	1798.0
20	804.85	609.8	1791.4	1790.3
25	792.1	620.9	1801.2	1803.7
<b>SD</b>	7.6159	12.2738	7.7210	6.2809

## Appendix A – Additional OIN Results

Table A.2: Results for calibration with image networks with optimum angle and close to optimum angle between CLs. Parameter estimations and SD given in pixels

Angle (deg)	$u_0$	$v_0$	$f_u$	$f_v$
80	802.4	617.6	1800.2	1797.8
85	801.2	619.3	1798.8	1795
90	804.1	620.9	1803.5	1801.7
95	802.6	619.6	1805.5	1802.7
100	803.5	620.1	1803.5	1801.4
<b>SD</b>	1.1104	1.2227	2.7285	3.2244

# Bibliography

- Abdel-Aziz, Y. I. and Karara, H. M. (1971), Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry, *in* 'Proceedings of the ASP Symposium on Close-Range Photogrammetry', Virginia, USA, 1971, pp. 1–18.
- Adelson, E. and Wang, J. (1992), 'Single lens stereo with a plenoptic camera', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), 99–106.
- Bando, Y., Chen, B.-Y. and Nishita, T. (2008), 'Extracting depth and matte using a color-filtered aperture', *ACM Transactions on Graphics* **27**, 134:1–134:9.
- Beardsley, P. and Murray, D. (1992), Camera calibration using vanishing points, *in* 'Proceedings of the 3rd British Machine Vision Conference', Leeds, UK, 1992, pp. 416–425.
- Born, M. and Wolf, E. (1980), *Principles of Optics (Sixth Edition)*, Pergamon Press.
- Bouguet, J. Y. (2008), 'Camera calibration toolbox for Matlab', [Accessed: October 2008].  
**URL:** [http://www.vision.caltech.edu/bouguetj/calib\\_doc/](http://www.vision.caltech.edu/bouguetj/calib_doc/).
- Brown, D. C. (1971), 'Close-range camera calibration', *Photogrammetric Engineering* **37**(8), 855–866.
- Brown, M., Burschka, D. and Hager, G. (2003), 'Advances in computational stereo', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(8), 993 – 1008.
- Caprile, B. and Torre, V. (1990), 'Using vanishing points for camera calibration', *International Journal of Computer Vision* **4**(2), 127–139.
- Chaudhuri, S. and Rajagopalan, A. N. (1999), *Depth from defocus - a real aperture imaging approach*, 1 edn, Springer.
- Cipolla, R., Drummond, T. and Robertson, D. (1999), Camera calibration from vanishing points in images of architectural scenes, *in* 'Proceedings of the 10th British Machine Vision Conference', Nottingham, UK, 1999, pp. 382–391.

## Bibliography

- Cossairt, O., Zhou, C. and Nayar, S. K. (2010), ‘Diffusion Coding Photography for Extended Depth of Field’, *ACM Transactions on Graphics (also Proc. of ACM SIGGRAPH)* **29**(4), Article 31.
- Dou, Q. and Favaro, P. (2008), Off-axis aperture camera: 3D shape reconstruction and image restoration, *in* ‘Proceedings of the 21st IEEE Conference on Computer Vision and Pattern Recognition’, Anchorage, USA, pp. 1–7.
- Dowski, E. J. and Cathey, W. (1994), ‘Single-lens single-image incoherent passive-ranging systems’, *Applied Optics* **33**, 6762–6773.
- Dowski, E. J. and Cathey, W. (1995), ‘Extended depth of field through wavefront coding’, *Applied Optics* **34**, 1859–1866.
- Ens, J. and Lawrence, P. (1993), ‘An investigation of methods for determining depth from focus’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **15**(2), 97–108.
- Farid, H. and Simoncelli, E. P. (1996), A differential optical range camera, *in* ‘Proceedings of the Annual Meeting of the Optical Society of America’, Rochester, USA, 1996, pp. 90–99.
- Farid, H. and Simoncelli, E. P. (1998), ‘Range estimation by optical differentiation’, *Journal of the Optical Society of America* **15**(7), 1777–1786.
- Faugeras, O. (1993), *Three Dimensional Computer Vision: A geometric Viewpoint*, MIT Press.
- Faugeras, O. D., Luong, Q.-T. and Maybank, S. J. (1992), Camera self-calibration: Theory and experiments, *in* ‘Proceedings of the 2nd European Conference on Computer Vision’, Ligure, Italy, 1992, pp. 321–334.
- Fife, K. G. (2009), Devices for integrated multi-aperture imaging devices, PhD thesis, Department of Electrical Engineering, Stanford University, Stanford CA.
- Gao, C., Ahuja, N. and Hua, H. (2007), Active aperture control and sensor modulation for flexible imaging, *in* ‘Proceedings of the 20th IEEE Conference on Computer Vision and Pattern Recognition’, Minneapolis, USA, 2007, pp. 1–8.
- Gokturk, S., Yalcin, H. and Bamji, C. (2004), A time-of-flight depth sensor - system description, issues and solutions, *in* ‘Proceedings of the Computer Vision and Pattern Recognition Workshop’, Vol. 3, Washington, USA, 2004, pp. 35–43.
- Golub, G. H. and Van Loan, C. F. (1996), *Matrix Computations (3rd Edition)*, John Hopkins University Press.
- Grossmann, P. (1987), ‘Depth from focus’, *Pattern Recognition Letters* **5**(1), 63–69.

## Bibliography

- Guillemaut, J.-Y., Aguado, A. and Illingworth, J. (2005), ‘Using points at infinity for parameter decoupling in camera calibration’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(2), 265–270.
- Guillemaut, J.-Y. and Illingworth, J. (2008), ‘The normalised image of the absolute conic and its application for zooming camera calibration’, *Pattern Recognition* **41**(12), 3624–3635.
- Gurdjos, P., Crouzil, A. and Payrissat, R. (2002), Another way of looking at plane-based calibration: The centre circle constraint, in ‘Proceedings of the 10th European Conference on Computer Vision’, Copenhagen, Denmark, 2002, pp. 252–266.
- Gurdjos, P. and Payrissat, R. (2001), Plane-based calibration of a camera with varying focal length: the centre line constraint, in ‘Proceedings of the 12th British Machine Vision Conference’, Manchester, UK, 2001, pp. 623–632.
- Hammarstedt, P., Sturm, P. and Heyden, A. (2005), Degenerate cases and closed-form solutions for camera calibration with one-dimensional objects, in ‘Proceedings of the 10th IEEE International Conference on Computer Vision’, Vol. 1, Beijing, China, 2005, pp. 317 – 324.
- Hartley, R. and Zisserman, A. (2003), *Multiple view geometry in computer vision*, second edn, Cambridge University Press, Cambridge, UK.
- Hiura, S. and Matsuyama, T. (1998), Depth measurement by the multi-focus camera, in ‘Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, Santa Barbara, CA, USA, 1998, pp. 953–959.
- Jones, D. G. and Lamb, D. G. (1993), Analysing the visual echo: passive 3-d imaging with a multiple aperture camera., Technical Report CIM 93-3, Department of Electrical Engineering, McGill University.
- Kang, S. B. (2007), Automatic removal of chromatic aberration from a single image, in ‘Proceedings of the 20th IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, Minneapolis, USA, 2007, pp. 1–8.
- Kim, J.-S., Gurdjos, P. and Kweon, I.-S. (2005), ‘Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**(4), 637–642.
- Kim, J.-S. and Kanade, T. (2011), ‘Multiaperture telecentric lens for 3D reconstruction’, *Optical Letters* **36**(7), 1050–1052.
- Kingslake, R. (1989), *A History of the Photographic Lens*, Academic Press.
- Knuth, D. E. (1998), *The Art of Computer Programming*, Vol. 3, Addison-Wesley.

## Bibliography

- Koh, K., Kuk, J. G., Jin, B., Choi, W. and Cho, N. I. (2011), ‘Autofocus method using dual aperture and color filters’, *SPIE Journal of Electronic Imaging* **20**(03), 033002.
- Lai, S.-H., Fu, C.-W. and Chang, S. (1992), ‘A generalized depth estimation algorithm with a single image’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(4), 405–411.
- Lee, E., Kang, W., Kim, S. and Paik, J. (2010), ‘Color shift model-based image enhancement for digital multifocusing based on a multiple color-filter aperture camera’, *IEEE Transactions on Consumer Electronics* **56**(2), 317–323.
- Levin, A., Fergus, R., Durand, F. and Freeman, W. T. (2007), ‘Image and depth from a conventional camera with a coded aperture’, *ACM Transactions on Graphics* **26**.
- Liang, C.-K., Lin, T.-H., Wong, B.-Y., Liu, C. and Chen, H. (2008), ‘Programmable aperture photography: Multiplexed light field acquisition’, *ACM Transactions on Graphics* **27**(3), 1–10.
- Lowe, D. G. (2004), ‘Distinctive image features from scale-invariant keypoints’, *International Journal of Computer Vision* **60**(2), 91–110.
- Lumsdaine, A. and Georgiev, T. (2009), The focused plenoptic camera, in ‘Proceedings of the 1st IEEE International Conference on Computational Photography’, San Francisco, USA, 2009, pp. 1–8.
- Mallon, J. and Whelan, P. F. (2005), ‘Projective rectification from the fundamental matrix’, *Image and Vision Computing* **23**(7), 643–650.
- Mallon, J. and Whelan, P. F. (2007a), ‘Calibration and removal of lateral chromatic aberration in images’, *Pattern Recognition Letters* pp. 125–135.
- Mallon, J. and Whelan, P. F. (2007b), ‘Which pattern? biasing aspects of planar calibration patterns and detection methods’, *Pattern Recognition Letters* **28**(8), 921–930.
- Miyagawa, I., Arai, H. and Koike, H. (2010), ‘Simple camera calibration from a single image using five points on two orthogonal 1-D objects’, *IEEE Transactions on Image Processing* **19**(6), 1528–1538.
- Nagahara, H., Zhou, C., Watanabe, T., Ishiguro, H. and Nayar, S. K. (2010), Programmable aperture camera using lcos, in ‘Proceedings of the 11th European Conference on Computer Vision: Part VI’, Springer-Verlag, Berlin, Heidelberg, 2010, pp. 337–350.
- Ng, R., Levoy, M., Brdif, M., Duval, G., Horowitz, M. and Hanrahan, P. (2005), Light field photography with a hand-held plenoptic camera, Technical report, Stanford University.

## Bibliography

- Pentland, A., Darrell, T., Turk, M. and Huang, W. (1989), A simple, real-time range camera, *in* ‘Proceedings of the 2nd IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, San Diego, USA, 1989, pp. 256–261.
- Pentland, A. P. (1987), ‘A new sense for depth of field’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **9**(4), 523–531.
- Poncelet, J. V. (1862), ‘Trait des propriétés projectives des figures’, *Imprimerie de Mallet-Bachelier, Paris*.
- Rupp, S. and Elter, M. (2007), Robust camera calibration - a generic, optimization-based approach, *in* ‘Proceedings of the 2nd International Conference on Computer Vision Theory and Applications’, Barcelona, Spain, 2007, pp. 61–68.
- Salvi, J., Armangue, X. and Batlle, J. (2002), ‘A comparative review of camera calibrating methods with accuracy evaluation’, *Pattern Recognition* **35**(7), 1617–1635.
- Sangjin Kim, Eunsung Lee, V. M. and Paik, J. (2010), ‘Real-time image restoration for digital multifocusing in a multiple color-filter aperture camera’, *SPIE Optical Engineering* **49**, 1–3.
- Schuler, C., Hirsch, M., Harmeling, S. and Schölkopf, B. (2011), Non-stationary correction of optical aberrations, *in* ‘Proceedings of the 13th IEEE International Conference on Computer Vision’, Barcelona, Spain, 2011, pp. 659–666.
- Slama, C. C., Theurer, C. and Henriksen, S. W., eds (1980), *Manual of Photogrammetry*, American Society of Photogrammetry.
- Smith, W. J. (2000), *Modern Optical Engineering*, 3 edn, McGraw-Hill New York.
- Stein, G. P. (1996), Lens distortion calibration using point correspondences, *in* ‘Proceedings of the 9th IEEE Conference on Computer Vision and Pattern Recognition’, San Francisco, USA, 1996, pp. 602–608.
- Stemmer (2011), ‘Stemmer imaging: The imaging and vision handbook’.  
**URL:** <http://www.stemmer-imaging.co.uk/en/handbook/>
- Sturm, P. F. and Maybank, S. J. (1999), On plane-based camera calibration: A general algorithm, singularities, applications, *in* ‘Proceedings of the 12th IEEE Conference on Computer Vision and Pattern Recognition’, Vol. 1, Colorado, USA, 1999, pp. 432–437.
- Subbarao, M. and Surya, G. (1994), ‘Depth from defocus: A spatial domain approach’, *International Journal of Computer Vision* **13**, 271–294.

## Bibliography

- Subbarao, M. and Wei, T.-C. (1992), Depth from defocus and rapid autofocus-ing: a practical approach, *in* ‘Proceedings of the 5th IEEE Computer Society Conference on Computer Vision and Pattern Recognition’, Champaign, IL, USA, 1992, pp. 773–776.
- Sutherland, I. E. (1964), Sketch pad a man-machine graphical communication system, *in* ‘Proceedings of the SHARE design automation workshop’, Vol. 6 of *DAC '64*, ACM, New York, NY, USA, 1964, pp. 329–346.
- Triggs, B. (1998), Autocalibration from Planar Scenes, *in* ‘Proceedings of the 5th European Conference on Computer Vision’, Vol. 1, Springer-Verlag, Freiburg, Germany, 1998, pp. 89–105.
- Tsai, R. Y. (1987), ‘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* **3**(4), 323–344.
- Walter, E. and Pronzato, L. (1997), *Identification of Parametric Models*, Springer.
- Wang, J. and Liu, Y. (2006), Characteristic line of planar homography matrix and its applications in camera calibration, *in* ‘Proceedings of the 18th International Conference on Pattern Recognition’, Hong Kong, China, 2006, pp. 147–150.
- Welford, W. J. (1986), *Aberrations of Optical Systems*, Adam Hilger.
- Weng, J., Cohen, P. and Herniou, M. (1992), ‘Camera calibration with distortion models and accuracy evaluation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(10), 965–980.
- Yang, C., Sun, F. and Hu, Z. (2000), Planar conic based camera calibration, *in* ‘Proceedings of the 15th International Conference on Pattern Recognition’, Vol. 1, Barcelona, Spain, 2000, pp. 555–558 vol.1.
- Zhang, Z. (1998), A flexible new technique for camera calibration, Technical report, Microsoft Research.
- Zhang, Z. (2000), ‘A flexible new technique for camera calibration’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(11), 1330–1334.
- Zhang, Z. (2004), ‘Camera calibration with one-dimensional objects’, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(7), 892–899.
- Zhou, C., Cossairt, O. and Nayar, S. (2010), Depth from Diffusion, *in* ‘Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition’, San Francisco, USA, 2010, pp. 1110–1117.
- Zhou, C., Lin, S. and Nayar, S. K. (2009), Coded Aperture Pairs for Depth from Defocus, *in* ‘Proceedings of the 12th IEEE International Conference on Computer Vision’, Kyoto, Japan, 2009, pp. 325–332.

## Bibliography

Zhou, C. and Nayar, S. K. (2009), What are Good Apertures for Defocus Deblurring?, *in* 'Proceedings of the 1st IEEE International Conference on Computational Photography', San Francisco, USA, 2009, pp. 1–8.