What makes a good dataset for light-field camera calibration?

Ruth Kravis
u6052450

Supervised by
A/Prof. Jochen Trumpf
Sean O'Brien

June 2019

ENGN3712 - Engineering Research and Development Project
The Australian National University

Australian
National
University

Abstract

This report presents an investigation of how certain dataset properties affect the quality of light-field camera calibration. Relatively little attention has been devoted to investigating what kind of datasets produce good calibration results. The rules-of-thumb that currently exist to guide people calibrating light field cameras have not been systematically verified, and the full range of possible factors has not been considered. By determining what makes a good calibration dataset, thorough recommendations can be made to the computer vision community, which will likely improve the quality of calibration people can achieve with light field cameras.

Through the detailed analysis of real datasets, we show that evaluating the calibration quality using standard reprojection and backprojection errors does not always reflect key differences in the calibration parameter estimates. This finding casts doubt on which error measures should be used for evaluating calibration quality. We show that the intrinsic parameter K_2, which is unique to light field cameras, is highly sensitive to the inclusion of images in which the calibration target is offset from the centre of the image. We test this sensitivity in a depth estimation task and demonstrate that including images where the calibration target is offset produces better depth estimates.

Contents

1 Introduction **1**
- 1.1 Overview . . . 1
- 1.2 Report Structure . . . 2
- 1.3 Summary of Contributions . . . 3

2 Background **4**
- 2.1 Theory . . . 4
  - 2.1.1 The 4-D Light Field . . . 4
  - 2.1.2 Standard and Light Field Cameras . . . 5
  - 2.1.3 Light Field Images . . . 6
  - 2.1.4 Camera Calibration . . . 7
- 2.2 Previous Work . . . 8
  - 2.2.1 Calibration Methods . . . 8
  - 2.2.2 Calibration Performance Measures . . . 10
  - 2.2.3 Calibration Datasets . . . 12
- 2.3 Problem Description . . . 13
  - 2.3.1 Lytro Illum Camera . . . 13
  - 2.3.2 Calibration Dataset Properties . . . 14

3 Experimental Setup **16**
- 3.1 Linear Stage . . . 16
- 3.2 Target with Fixed Depths . . . 17
- 3.3 Image Processing and Decoding . . . 17
- 3.4 Experimental Errors . . . 18

4 Preliminary Investigations **19**
- 4.1 Consistency of Parameter Estimates . . . 19
  - 4.1.1 Motivation and Hypotheses . . . 19
  - 4.1.2 Experimental Design . . . 19
  - 4.1.3 Results and Discussion . . . 21
    - 4.1.3.1 Performance Metrics . . . 21
    - 4.1.3.2 Parameter Estimates . . . 23
    - 4.1.3.3 Precision of Parameter Estimates . . . 27
- 4.2 Varying Apparent Size Using Different Grid Sizes . . . 29
  - 4.2.1 Motivation and Hypotheses . . . 29
  - 4.2.2 Experimental Design . . . 29
  - 4.2.3 Results and Discussion . . . 30
- 4.3 Angled Images . . . 34
  - 4.3.1 Motivation and Hypotheses . . . 34
  - 4.3.2 Experimental Design . . . 34
  - 4.3.3 Results and Discussion . . . 35

5 Primary Investigation **38**
5.1 Hypotheses 38
5.2 Experimental Design 39
5.3 Results and Discussion 41
5.3.1 Performance Metrics 41
5.3.2 Depth Estimation 42
5.4 Summary 45

6 Conclusions and Future Work **47**

A Appendix A: Additional Figures **49**

B Appendix B: Additional Tables **53**

Bibliography **55**

Chapter 1

Introduction

1.1 Overview

Light field cameras have the ability to sample the *light field*, a 4-D function recording the intensity of light in free space over a set of positions and directions. This gives light-field cameras an advantage over standard cameras, which are only able to sample over a set of positions. This additional directional information captured in a single light field image is comparable to the information captured by taking multiple images from slightly different positions using a standard camera. This additional information can be used to refocus a light-field image at different depths, or to to compute a depth map of the scene.

Calibration of both light field and standard cameras is a fundamental task required for many computer vision applications. Calibration is the process of relating 3-D points in a scene in front of the camera to their corresponding points on the image plane. The relationship between these points cannot be determined by camera geometry alone (or by a combination of the geometry and the settings of the camera) because of real lens effects and imperfections in the construction of the camera (e.g. not achieving coplanarity between the image sensor and the main lens). Therefore, the relationship has to be obtained experimentally. Good calibration is crucial for applications such as depth estimation and 3-D reconstruction.

Most calibration methods require the collection of a calibration dataset. A calibration dataset is a set of images of a target with known feature point locations. A feature point can be any point that is easy for a computer to detect. A common target choice is a checkerboard, where the internal checkerboard corners are used as feature points, since they can be easily detected in an image. The physical locations of the feature points with respect to one another needs to be known, in order to obtain ground truth on the 3-D side of the relationship. These 3-D points can be projected to their expected locations in the image given a set of calibration parameter estimates. An optimisation routine can then be used to determine what calibration parameters will accurately project these known 3-D points onto the image plane. the optimisation routine typically tries to minimise the difference between the projected points and the detected points.

Much attention has been paid to different methods of calibrating light field cameras. In comparison, there has been little attention paid to the collection of calibration datasets, and what properties the dataset should have. It is evident that the quality of the data fed into any calibration method will affect the calibration quality. There are some generic recommendations for calibrating standard cameras, which instruct users to take many im-

ages at every possible position and orientation. Similar instructions are given for light field cameras. To date, however, there has been no systematic verification of these recommendations, or exploration of how calibration dataset properties affect calibration outcomes. Knowing how to collect calibration datasets to achieve good calibration results would undoubtedly have a positive impact on members of the computer vision community who rely on accurate calibration for their work and research.

This research was undertaken with the aim of determining what properties a good calibration dataset possesses. The initial approach was to evaluate the calibration result using a set of error measures commonly used in the community, and examine the effect of dataset properties on these errors. However, it became clear that this typical method of measuring reprojection and backprojection errors did not give the same verdict on calibration quality as measuring the precision and accuracy of the calibration parameter estimates. Additionally, the errors and parameter estimates were shown to respond differently to the same factors. This was both a useful finding but also a challenge, as it meant that the error measures and the parameter estimates needed to be investigated independently for each experiment, and then compared to identify any correlation.

This report presents the experimental results from four targeted experimental series that were run, investigating how dataset size, pose set (particularly offset and angle), and apparent target size affected the calibration results. In the first experimental series, the dataset size is shown to strongly influence the precision of the calibration estimates, therefore larger dataset sizes are recommendation required for more precise parameter estimates. An effect on the performance metrics due to the apparent size of the calibration target in the images was also identified in this experiment. Lastly, the inclusion of offset images is shown to have a large effect on the estimate of K_2, a parameter unique to light field cameras. In the second experimental series we explore the apparent size effect further, and show that smaller apparent target size systematically decreases two important performance metrics. In the third experimental series, angle is investigated, and while concrete conclusions were difficult to draw, the experiment does show that severely angled images should be avoid to achieve acceptable performance metrics and stable parameter estimates. Lastly, through performing a depth estimation task that uses the parameters unique to light field cameras, we were able to verify that including offset images in the dataset produces better calibration parameters, although not necessarily better performance metrics.

1.2 Report Structure

This report is organised into five chapters, excluding this introduction. Chapter 2 (Background) is divided into three main sections: the first (Theory 2.1) gives an overview of light fields, light field imaging, calibration, and disparity estimation. The second section (Previous Work 2.2) covers the calibration method that will be used for the investigation. The metrics used to measure calibration performance will also be discussed, as well as the general recommendations that are currently given for the collection of calibration datasets for light field cameras. The last section of this chapter (Problem Description 2.3) provides details on the camera used, the dataset properties of interest, and why each property may be important.

Chapter 3 (Experimental Setup) is dedicate to an in-depth description of the constructed apparatus used to obtain experimental data. This chapter also gives additional detail on the image processing and the decoding methods used. There is a short discussion of experimental errors (Section 3.4) and their possible effects on the experiment outcomes.

Chapter 4 (Preliminary investigations) contains the methodology, design, and results for each of the preliminary experiments conducted. The first section (4.1) investigates the consistency of parameter estimates and serves as a general introduction to the kind of behaviour we are interested in. The second section (4.2) describes a follow-up experiment that was conducted to verify results relating to the performance metrics from the first experiment. The final section (4.3) describes the experiments conducted that investigated the effect of angled images on calibration results.

Chapter 5 (Primary Investigations) contains the results from the more targeted experiment that was conducted to verify the preliminary findings in a depth estimation application. Not all of the data collected for this experiment was used, so there is a short discussion at the end of this chapter regarding future analysis of that remaining data and suggested methods for doing so.

Chapter 6 gives a summary of the work completed and the key contributions. There is also a discussion of future work that could be undertaken both within the initial scope of this research but also in broader areas. Lastly, additional figures from the preliminary investigations are contained in Appendix A, and additional tables in Appendix B.

1.3 Summary of Contributions

- Recommendation for the use of large datasets to achieve precise parameter estimates for light field camera calibration
- Analysis of the relationship between commonly used reprojection and backprojection errors and the accuracy of camera parameter estimates, showing that the errors are not always good indicators of the parameter quality
- Identification of sensitivity in two error measures to the apparent size of the calibration target
- Identification of sensitivity of K_2, a parameter unique to light field cameras, to the inclusion of images where the calibration target is offset from the image centre.
- Showing that the inclusion of images where the calibration target is offset from the image centre improves performance in a depth estimation task
- Recommendations for which factors should be prioritised for future work and suggested methods for analysis

Chapter 2

Background

This chapter is composed of three sections which cover the relevant theory, literature, and the motivations for conducting this research. In the first section, the theory of light fields, light field imaging, and camera calibration will be covered. In the second, relevant literature within the area of interest will be reviewed. The final section will focus on presenting and formulating a problem description that targets the identified gap in the field. The material in the first and second sections draws primarily from [13; 12; 2; 8; 1; 6].

2.1 Theory

This section will outline what a light field is and how light field cameras can capture it. There will be a detailed discussion of the images produced by a light field camera, followed by a high-level discussion of camera calibration. This content will provide the background information necessary to understand the content in the next section.

2.1.1 The 4-D Light Field

At the most fundamental level, light allows image sensors (and by extension, cameras) to retrieve information about the way the world is. Rays of light reflect off objects in the world and travel from a particular point with an intensity and in a direction to an image sensor. The image captured depends on the position and orientation of these objects relative to the sensor, because this will determine which rays hit the image sensor. Therefore, the most useful representation of light will be one that can capture intensity as a function of both position and direction.

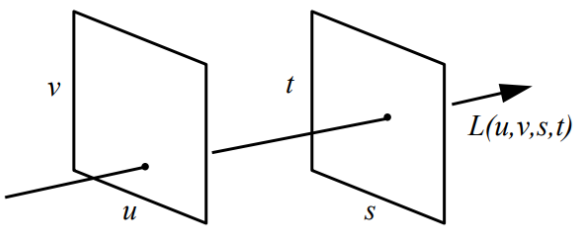


Figure 2.1: Two-plane parametrization of the light field [8]

One representation that achieves this is the *light field*. The light field is a vector function that describes the radiance of light rays as a function of position and direction [8]. This

representation can have up to seven dimensions, if time and wavelength are considered, although for static scenes, neither time nor wavelength are necessary. The remaining five parameters consist of three position and two direction parameters. However, in free space where rays are not blocked, one of the position parameters is redundant, since radiance will not change along the axis of propagation [8]. This leaves us with a 4-D representation of the light field which can be parametrized in several different ways. One parametrization that is frequently used in imaging applications is the two-plane parametrization, which describes the position and direction of a ray by its intersection with two parallel planes, u, v and s, t. This parametrization is shown in Fig 2.1. The value of the function $L(u, v, s, t)$ is an intensity, and the parameters (u, v, s, t) define a single light ray. While useful, this parametrization has certain limitations. In particular, it fails to capture rays travelling parallel to the two planes. The discussion of light field cameras in the next section will demonstrate why this limitation does not pose a problem for camera applications.

2.1.2 Standard and Light Field Cameras

Standard cameras form an image by integrating the intensity of all light rays arriving at a pixel from different directions. An image taken by a standard camera is therefore a 2-D sample of the 4-D light field, since no information from different perspectives is captured in the image. To construct a light field camera, this information about direction has to be captured. One way of doing this is by 'multiplexing' the angular (directional) domain within the spatial domain [16]. Practically, this is achieved by inserting a micro-lens array (MLA) behind a main focus lens and in front of the image sensor, such that multiple pixels sit behind each micro-lens and the main lens, MLA, and image sensor are coplanar [11]. A more detailed explanation of how the camera geometry is modelled is given to help understand how angular information can be stored within the spatial domain, and the trade-off that exists between angular and spatial resolution. This explanation will also clearly demonstrate how such an arrangement of lenses can sample the full 4-D light field.

The main focus lens of the light field camera is generally modelled as a thin lens with radial distortion. The light rays that travel to the main lens from a point P_i intersect the lens and are refracted. These rays subsequently converge to a single point behind the lens to form a 'virtual' image at the *image point* [13]. The rays then diverge until they intersect the MLA. At the MLA, each ray intersects a lenslet, and each lenslet is modelled as a pinhole camera. Therefore, each lenslet produces its own tiny circular *subimage* on the pixels behind it. The image plane ends up being populated with the subimages of each lenslet in the array [5]. Examples of what light field images look like are given in Fig 2.3 in Section 2.1.3.

This focal length of the lenslets in the MLA cannot be adjusted in the same way the focal length of the main lens can. When the MLA is positioned one focal length (where focal length refers to the focal length of the lenslet) from the image sensor, we refer to the camera as an 'unfocused' light field camera [17]. Some light field cameras (such as the Raytrix R42) contain an MLA with micro-lenses of multiple focal lengths [12]. This is desirable if both large aperture and larger depth of field is required [15]. This is possible because the micro-lens focal lengths can be chosen so that their depths-of-field nearly overlap, giving an enhanced overall depth of field.

Fig 2.2 shows that a single 3-D point P is seen by multiple lenslets. Within each subimage

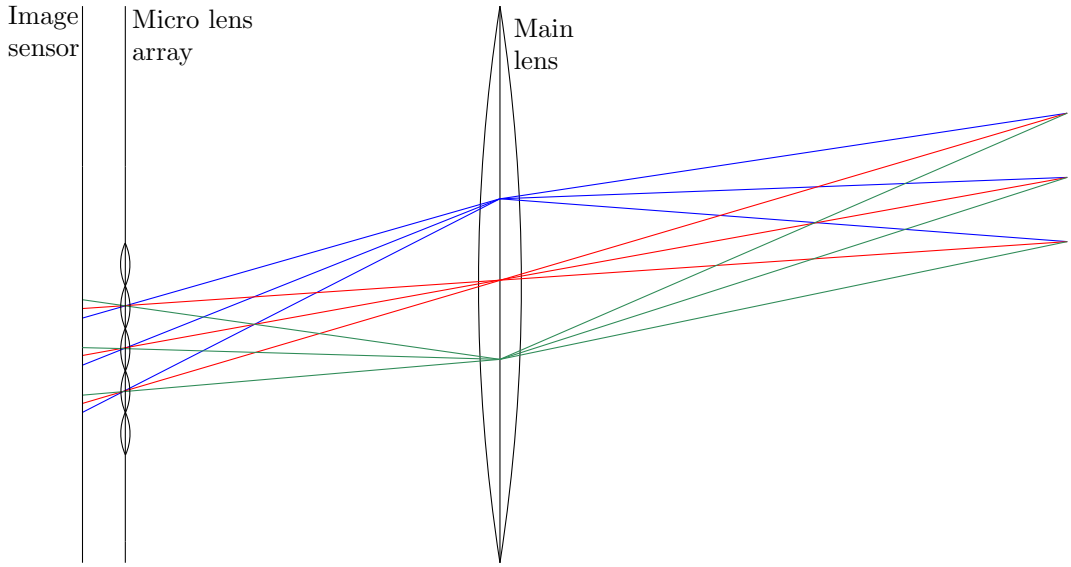


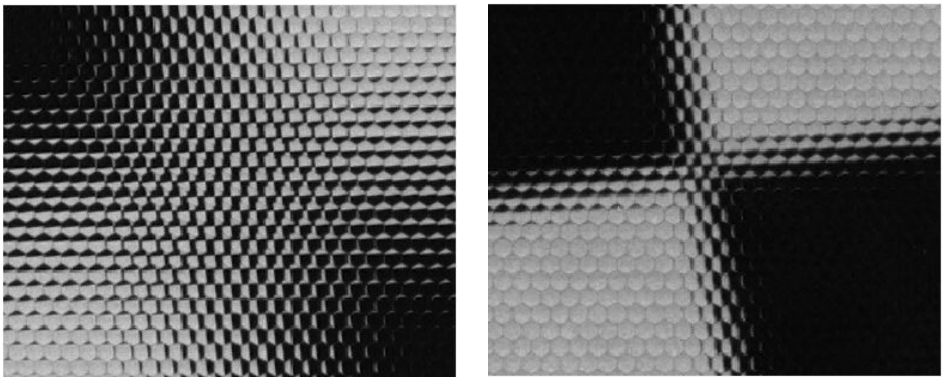
Figure 2.2: Diagram of a lenslet-based light field camera. Note how rays leaving each point from different directions are projected to different pixels on the image sensor. If all rays passing through a point on the main lens (e.g. all red rays) are selected, a single perspective image like one produced by a standard camera is produced.

produced by these lenslets is a single pixel that corresponds to P. The geometry of the lenslet-based unfocused light field camera means that rays leaving P in different directions intersect the main lens at different angles and get imaged to separate pixels, thereby preserving the directional information. As discussed in Section 2.1.1, 2-D images like those obtained from a standard camera can be extracted by selecting all rays passing through a point on the (u, v) plane [5]. This means selecting all rays passing through a single point on the aperture. In reality, each pixel is not a point, and since it has some size, the pixel still integrates a volume of light, but we can approximate this volume as a single ray [2].

2.1.3 Light Field Images

The image formed on the image sensor of a light field camera has a distinctive appearance. We refer to the image in this format as the *raw image*. Fig 2.3 shows two examples of different raw images. First, note how the lenslets are 'physically' visible in the raw image: they are packed in a hexagonal array to achieve a high packing factor. Second, we can also see that a single 3-D point (e.g. the checkerboard corner) appears in multiple subimages. Furthermore, in each of those subimages, the corresponding pixel is in a slightly different position relative to the lenslet centre. This is shown more clearly and discussed in more detail in Section 2.2.1. These raw images show how the concept of an object being 'in focus' applies to raw light field images. If an object is in focus (e.g. Fig Section 2.3b), it will appear in fewer subimages than if it is 'out of focus' (e.g. Fig Section 2.3a).

Raw images are not the only image format we can extract from a light field camera. There are also *sub-aperture images* [5; 11]. In Section 2.1.1 it was mentioned that producing a 2-D slice of the light field (like what is captured with a standard camera) requires selecting only those rays passing through a single point on the (u, v) plane, and considering all the corresponding points on the parallel (s, t) plane. For the light field camera, the (u, v) plane is



(a) 'Blurred' raw image (b) 'Focused' raw image

Figure 2.3: Cropped sections of typical raw images of a checkerboard corner. Taken with the Lytro Illum camera

the main lens plane. Sub-aperture images are simply these 2-D slices. They are equivalent to the image you would get by placing a pinhole camera at a single point on the main lens of the camera [13; 5]. The number of sub-aperture images you can generate is determined by the radius in pixels of the subimages. The resolution of these sub-aperture images is determined by the resolution of the raw image, and the radius in pixels of the subimages. A large subimage radius but small raw image resolution will result in many low-resolution sub-aperture images. If we describe a camera like this in terms of its angular and spatial resolution, we would say that it has a high angular resolution (because many sub-aperture images can be generated, and each sub-aperture image is a slightly different perspective of the scene). However, the spatial resolution would be lower, since each sub-aperture image will necessarily be lower-resolution. In contrast, a small subimage radius but large raw image resolution will result in fewer high-resolution sub-aperture images, and the trade-off between the angular and spatial resolution will be opposite (lower angular resolution but higher spatial resolution). The usefulness of sub-aperture images in calibration will be discussed in the following section.

2.1.4 Camera Calibration

Camera calibration is required whenever metric knowledge of a scene is required. Calibration involves estimating both intrinsic and extrinsic camera parameters by taking images of a target with known geometry and physical size (e.g. a checkerboard). Extrinsic camera parameters define a rigid-body transformation from 3-D points in a fixed world frame to points in the camera's body-fixed frame. Extrinsics are also referred to as the camera pose ('pose' refers to the combination of position and orientation). A single image will have one corresponding pose comprised of six extrinsic parameters: three position parameters (which form a translation vector), and three rotation parameters (which can be written as a skew vector or a rotation matrix). Intrinsic camera parameters define the transformation of 3-D points in the camera frame (obtained via the extrinsic parameters) to points on the image plane. For standard cameras, this is a 3-D to 2-D projective transformation. Together, extrinsic and intrinsic parameters define the camera model that maps 3-D points in a fixed world frame to points on the image plane.

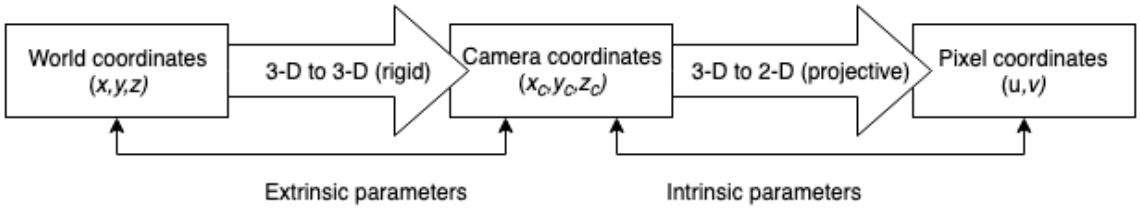


Figure 2.4: Camera calibration diagram for a standard camera. For a light field camera, 3-D points are projected to multiple pixel coordinates, and the projection is not 3-D to 2-D.

Calibration of light field (and other types of cameras) generally consists of three stages. In the first step, features in raw images are estimated. Theoretically, only a single image is needed for calibration. Checkerboards and arrays of small dots are commonly used as calibration targets, since there are robust methods for detecting their location using image processing techniques. The second step involves using the feature data to estimate initial values for the calibration parameters. Lastly, these initial parameters are refined by minimising an error function in an optimisation routine. Figure 2.4 shows this routine as a block diagram, highlighting how the extrinsic and intrinsic parameters relate points in the different coordinate systems.

2.2 Previous Work

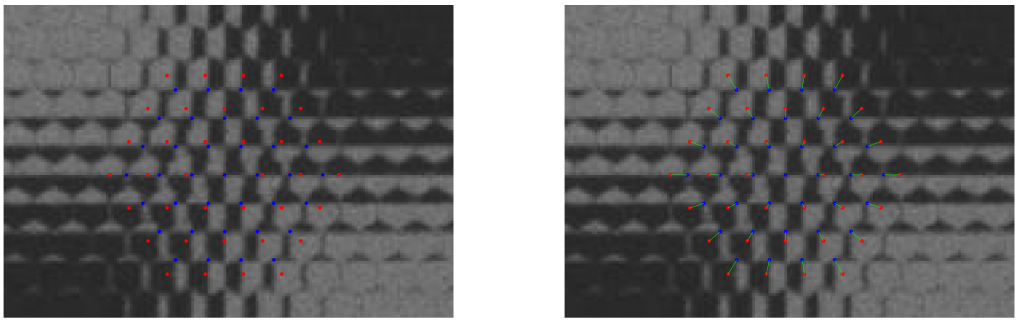
In this section, the calibration method used in this research will be discussed in detail, followed by a discussion of the performance measures commonly used to evaluate calibration quality. Lastly, a review of current recommendations regarding calibration dataset collection will be given.

2.2.1 Calibration Methods

There are a range of approaches to light field camera calibration in the literature: some use ray-based methods that relate pixels to corresponding rays, others match features within subimages or within sub-aperture images [2; 1; 12; 1]. The method used for this research was developed by O'Brien *et al.* and is described in detail in [13]. A summary of the key features of this method will be given. The Matlab implementation of this method is available online[1].

The features used in this calibration method are called plenoptic discs. These discs are in one-to-one correspondence with feature points in the scene, and can therefore be used as features in a calibration routine. The underlying intuition for plenoptic discs is that in a raw image, a point P_i (such as the corner of a checkerboard) appears in multiple lenslet subimages. Therefore, each point P_i will have a collection of *lenslet-pixel pairs*, which are simply the coordinates of each lenslet in which P_i appears, and the corresponding pixel location of P_i within that lenslet's subimage. The lenslet-pixel pairs for P_i contain two pieces of information. First, which lenslets can see a given point, and second, the pixel within that lenslet's subimage where the point appears. Fig 2.5 shows what this looks like

[1] Available at: https://github.com/sgpobrien/PlenCalToolbox



(a) Lenslet-pixel pairs (b) Lenslet-pixel pairs matched

Figure 2.5: Lenslet-pixel pairs for a checkerboard corner point P_i. Lenslet centres are in red, and the corresponding pixel for the lenslet is in blue. The green lines in (b) indicate which pixel belongs to which lenslet.

on the raw image. It is clear that this set of lenslet-pixel pairs forms a disc, which is centred at the lenslet which see the point in the centre of its subimage. The edge of this disc is defined by the coordinate of the lenslet that sees the point on the edge of its subimage, at the maximum distance, the subimage radius. Thinking in this manner, we can describe a *plenoptic disc* that corresponds to each point P_i. The plenoptic disc is defined by the triple (w_s, w_t, R) where (w_s, w_t) is the centre of disc and R is a signed radius. These disc parameters can be determined for each point by utilising the linear relationship between the lenslet coordinate and the position within the lenslet subimage that the point appears. This linear relationship is visually represented in Fig 2.5b.

Using the intrinsic geometry of the light field camera, the following projection equation can be generated:

$$\Pi(P) = \left(-f^u \frac{P^x}{P^z} + c^u, -f^v \frac{P^y}{P^z} + c^v, -\frac{rK_2}{P^z} - rK_1 \right). \tag{2.1}$$

In Eqn 2.1, P^x, P^y, and P^z are the coordinates of a single point P in the camera coordinate system, r is the subimage radius in pixels, and f^u, f^v, c^u, c^v, K_1 and K_2 are parameters estimated for in the calibration routine. Given estimates of the feature data (plenoptic discs), this equation can be used to estimate the intrinsic and extrinsic parameters of the camera.

The procedure for estimating this disc data is as follows. First, N sub-aperture images indexed $k = 1, ..., N$ are generated by selecting from every subimage a pixel with constant offset (u_k, v_k). These pixels are stitched together into a subimage with dimensions $\frac{U}{r}$ by $\frac{V}{r}$ where $U \times V$ is the resolution of the raw image. Since the lenslets are hexagonally arranged, convex interpolation is used to generate colours for the sub-aperture image pixels. Matlab's **detectCheckerboardPoints** function is used to detect corner points in the sub-aperture images, and to detect the dimensions of the checkerboard. Each detected corner in the sub-aperture image is associated with its corresponding lenslet-pixel pair. Because there are (typically) multiple lenslet-pixel pairs for a given feature point P_i, this provides an overdetermined linear system of equations for which a least-squares estimate of (w_s, w_t, R) can be computed. This completes the feature estimation step of the calibration

Parameters	Description
K_1	Relates to the distance between the MLA and main lens and the focal length.
K_2	Relates to the distance between the MLA and the main lens, and the MLA and the image sensor, should not depend on focal length.
(f^u, f^v)	Focal length of the pinhole camera model for the lenslets
(c^u, c^v)	Centre pixel of the raw image
k_1	First coefficient for the radial distortion model
k_2	Second coefficient for the radial distortion model

Table 2.1: Intrinsic parameters estimated for in the calibration method used. These parameters are estimated for the calibration dataset as a whole.

Parameters	Description
τ^x, τ^y, τ^z	x, y, z position of camera in the coordinate system of the target
$\alpha^x, \alpha^y, \alpha^z$	Skew values, corresponding to entries of rotation matrix through the matrix exponential

Table 2.2: Extrinsic parameters estimated for in the calibration method used. These parameters are estimated for every frame.

routine.

The calibration initialisation is performed by deriving a linear system of equations that estimates initial values for all calibration parameters we want to estimate. The exact procedure follows that given in [1]. The optimisation step uses Matlab's **lsqnonlin** function to perform non-linear optimisation using the Levenberg-Marquardt algorithm. The error that is minimised in this optimisation routine is called the plenoptic reprojection error and is given by:

$$(\Lambda, \Xi; \Phi) = \sum_{i,j} (\Pi_\Lambda(P_{i,j}) - (w^s_{i,j}, w^t_{i,j}, R_{i,j}))^2 \tag{2.2}$$

where $P_{i,j} = X_j^{-1} \mathbf{o} P_i$ describes the rigid transformation of coordinates from the target frame to the camera frame, and Π_Λ denotes the plenoptic projection with lens distortion modelled by a second-order approximation, and parameters given by the intrinsic parameter estimate Λ. The parameters estimated can be separated into a set of intrinsic parameters and a set of extrinsic parameters, which are summarised in Table 2.1 and 2.2 respectively. In the remainder of the report, the parameters (f^u, f^v) and (c^u, c^v) will also be denoted (f^x, f^y) and (c^x, c^y).

2.2.2 Calibration Performance Measures

There is no single standard performance measure for light field camera calibration. Most calibration methods will optimise an error that is unique to the calibration method. However, in addition to the error minimised in the optimisation step, there are three common measures used to evaluate calibration accuracy. These errors have a non-transitive relation to each other. A description of each error and how it is calculated will be given, which will make the most sense in the context of the plenoptic disc method described in the previous section.

The first performance measure is mean 3-D reconstruction error, which will be referred

to as mean backprojection error (MBE). This is a backprojection into 3-D space. To calculate this error, the feature data (plenoptic disc data) is transformed into camera frame coordinates. The distortion model is applied using the estimated distortion coefficients. Lastly, using the estimated extrinsics, the points are rigidly transformed from the camera frame to the fixed world frame of the checkerboard [13]. The distance between the feature point estimates and the actual feature points (ground truth) is measured, and the result is divided by the estimated depth of the feature point to obtain MBE as a percentage.

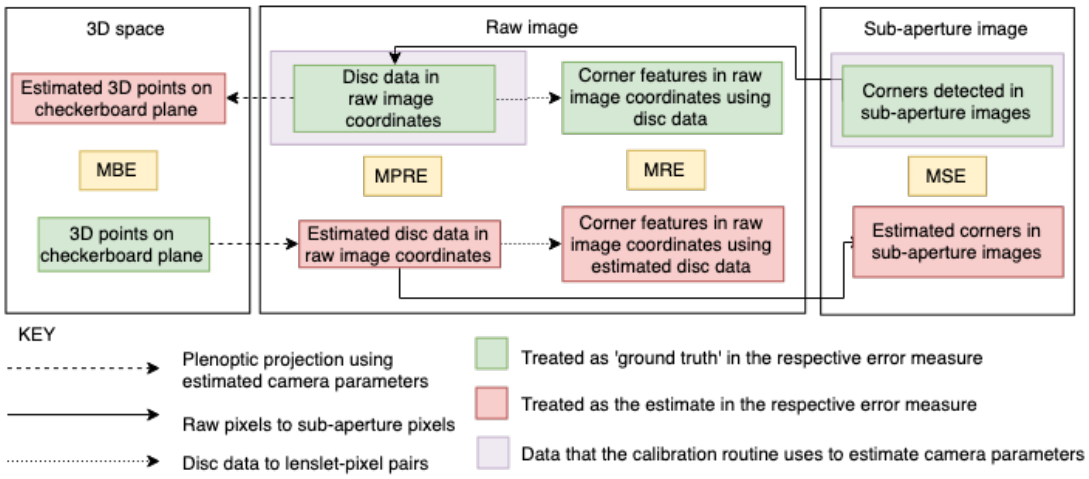


Figure 2.6: Error diagram indicating how each performance measures calculates the error in different coordinate systems.

The second performance measure is mean reprojection error (MRE). This is a reprojection onto the image plane rather than into 3-D space. It is calculated by comparing the estimated feature points (initially extracted from the raw data, functioning as ground truth) to the feature points produced by projecting the true 3-D feature points through the camera model to get the reprojected feature points on the raw image [13]. The average distance between these two sets of feature points is calculated to give MRE, measured in pixels.

The final performance measure is mean sub-aperture reprojection error (MSE). Like MRE, it is also a reprojection. It is calculated in two stages. First, we project the true 3-D feature points through the camera model to get the reprojected feature points on the raw image. Then these reprojected feature points are mapped onto sub-aperture images. The average distance is taken between these reprojected feature points and the feature points originally extracted from the sub-aperture images [13]. This error is also measured in pixels.

The error used in the calibration routine is called the plenoptic reprojection error. It is conceptually most similar to the mean reprojection error. The mean plenoptic reprojection error is calculated by projecting plenoptic discs onto the raw image using the estimated camera parameters, and taking the mean of Eqn 2.2. Fig 2.6 gives a visual summary of the performance measures, summarising how they relate and which coordinate system they operate in. From Fig 2.6 it is clear that we should not expect the errors to necessarily be correlated with each other.

2.2.3 Calibration Datasets

Calibration datasets are simply the set of images that calibration is performed on. These datasets usually consist of a number of images taken of a target, usually with different camera poses. An example of a calibration dataset is given Fig 2.7. The dataset collected will define the sample of 3-D points and corresponding features in the image that the calibration algorithm has to work with. Therefore, it is intuitive that different calibration datasets will produce different results, because they will contain different samples of 3-D points. However, there is very little literature on how to collect good calibration datasets, what good makes a good calibration dataset for light field cameras. Furthermore, when suggestions are made on what good datasets might look like, they tend to lack robust supporting evidence.

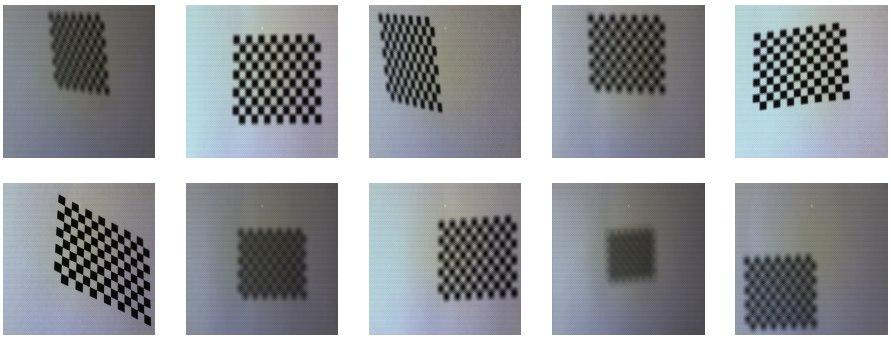


Figure 2.7: Example of a calibration dataset with large variations in camera pose

There are a few commonly followed techniques for collecting calibration datasets. As a rule-of-thumb, calibration datasets should contain a variety of poses and a reasonable number of images (6 - 10 at minimum) [2; 12]. The justification for variety is intuitive: more 3-D points are sampled in a set of two images taken from different poses than a set with two images taken from identical poses, and more information should benefit any calibration algorithm. The strongest justification for a minimum number of images is given in [12] where K_1, K_2, c_x, and c_y are shown to be estimated with decreasing standard deviation as the number of frames per set increases. Limitations with investigation given in [12] include (a) the small range of set sizes investigated: only k-combinations from 10 images were used, (b) not stating what kind of images were used, and (c) not giving any accompanying performance metrics for the accuracy of the converging values of the listed parameters.

In [3] it is noted that shadows on the raw images and difficult lighting conditions result in fewer corners being matched, which decreases the performance of calibration for camera-to-camera calibration. [4] also notes that uneven illumination can cause the feature detection to fail, which evidently has effects on the calibration quality, especially if this failure is not taken into account by removing the affected image(s).

Matlab's Computer Vision Toolbox also gives guidance on how to improve calibration accuracy. The suggestions are: (a) use more radial distortion/tangential distortion/skew coefficients, (b) take more calibration images, ensuring the images have different 3-D orientations and all parts of the field of view are covered (particularly edges and corners to ensure distortion coefficients are well estimated), and (c) exclude images that have high

reprojection errors and re-calibrate [10].

There are also techniques that improve the accuracy of the calibration that are not related to dataset collection but are related to dataset use within the algorithm. In [1] it is noted that calibration errors can be reduced by discarding sub-aperture images generated using pixels from the edges of subimages. However, discarding too many sub-aperture images will decrease the amount of data available and will reduce the accuracy of feature detection. This technique is also only applicable to methods that directly use sub-aperture images to estimate feature data.

In summary, little is known about how properties of calibration datasets affect the calibration outcomes beyond the rules-of-thumb presented. There is a clear gap in the field regarding what makes good calibration datasets. By closing this gap, useful advice can be given to members of the computer vision community on how to collect good calibration datasets.

2.3 Problem Description

This section will expand upon the dataset properties of interest and the focus area for this research. First, the key details of the two cameras used will be presented. Using two different light field cameras makes it possible to determine whether the results can be generalised or are simply effects that are tied to differences in design or quality between different light field cameras. Second, the dataset properties of interest will be outlined, with a discussion of how and why each property could influence the calibration outcome.

2.3.1 Lytro Illum Camera

The Lytro Illum was predominantly used for this research presented in this report. The Lytro Illum is marketed as a consumer light field camera, and is not of the same quality as cameras designed for high precision industrial and research applications, such as those manufactured by Raytrix. Key properties of the Lytro Illum are given in Table 2.3.

Property	Value
Aperture	f/2.0
Megarays	40
Megapixels	5
Resolution	5368 x 7728
Subimage radius (pixels)	7
Multifocus	No
Rolling shutter	No

Table 2.3: Details of Lytro Illum camera [9]

The key parameters that are relevant for calibration are the resolution and the subimage radius. The resolution of the Illum is comparable to a camera like the Raytrix R42. However, the Raytrix R42 has a subimage radius of 15 pixels compared to 7 for the Illum. This means that compared to the Raytrix, the Illum will produce higher resolution sub-aperture

images due to its low subimage radius, and therefore achieve quite good spatial resolution at the expense of lower angular resolution. Comparatively, the Raytrix achieves very good angular resolution, but has lower resolution sub-aperture images, highlighting the trade-off that was discussed in Section 2.1.3. It is also important to note that the Lytro suffers noticeable main lens distortion. To account for this, a second-order radial distortion model was used, where two distortion coefficients are estimated.

2.3.2 Calibration Dataset Properties

Several key properties of calibration datasets were alluded to in Section 2.2.3. A comprehensive summary of dataset properties is given below in Table. 2.4:

Name	Description
Dataset size	Number of images
Number of features	Number of features on target
Target symmetry	Symmetric or asymmetric
Pattern size	Millimetres (mm)
Pose set	Set of translation vectors and rotation matrices (or skew vectors) relating for each image
Target type	Checkerboards, grids of dots, and square grids are common examples.
Focal distance	Millimetres (mm) from main lens

Table 2.4: Calibration dataset properties

Under the assumption that each image in the dataset is different (i.e. we are not feeding in the same image multiple times), the dataset size is expected to influence the calibration outcome because additional images contain additional samples of 3-D points and corresponding feature points, which should improve the calibration parameter estimates that describe the relationship between all 3-D points and corresponding feature points. This factor will depend heavily on the variation between images. For number of features, if the target has few features, there are less samples of the 3-D points and corresponding feature points for the algorithm to use. Therefore, we think that more features should always be better. For feature symmetry, because the raw images produced by light field cameras (at least the Lytro Illum and Raytrix R42) are rectangular, the symmetry of the target may matter. For example, an asymmetrical target can occupy the entire raw image in a single frame, which gives a more complete set of 3-D points at that particular distance.

The size of the pattern is expected to influence the calibration outcomes in the following way. Features in small pattern sizes at large distances become difficult to detect, which will decrease the robustness of the feature detection and decrease the calibration quality. This factor will be closely related to the focal distance, since short focal distances require smaller pattern sizes [13]. The pose set is the most complex factor of those listed. While it may be clear how pose in a single image affects the disc data for that image, it is not clear how that disc data affects the calibration outcome, nor how multiple images interact after the disc data estimation stage. Not only does the pose per image matter, but there may be effects between poses. If all the poses are clustered or very similar, we will not get an even sample of 3-D points and corresponding feature points. This factor is likely to be

as important as number of features or dataset size, since pose set determines exactly what the distribution of sampled 3-D points looks like.

Lastly, the target type is expected to influence the calibration outcome. The extent of this influence may depend on the particular calibration algorithm, and the method used to detect feature points. For example, high-accuracy feature extraction from subimages is made difficult by the low resolution of the subimages. A calibration method that extracts features from subimages may benefit more from features that are more constrained (e.g. triangular tiling versus checkerboard). Lastly, the focal distance (which will be determined by the desired application) will shift the distances at which the target needs to be placed to produce the same disc data. Since the disc data is what the calibration algorithm actually operates on, it is conceivable that we should be trying to achieve the same disc data independently of focal distance, and therefore focal distance may interact with pose (and pattern size) to affect the calibration outcome.

Performing calibration can take a significant amount of time, particularly on large datasets. Given that it was not possible in the time-frame of this project to independently vary every factor in every possible way, not all factors could be investigated. The factors that were not considered in this research were pattern symmetry and target type. These two factors were excluded because to investigate them, completely new datasets would be required, whereas a factor such as number of images or pose set can be investigated using subsets of a single large dataset.

Chapter 3

Experimental Setup

This chapter describes the setup for the series of experiments undertaken. An experimental approach of collecting real data was favoured over an analytical approach that studies the propagation of errors through the calibration algorithm. Given that the intention is to make recommendations on calibration dataset collection, by collecting real data, these recommendations can be made in the appropriate language. For example, an analytical approach may reveal that the algorithm is sensitive to the value of a certain variable within the algorithm, that may or may not have a clear association with a property of the dataset. This evidently has limited use if the goal is to make recommendations on what a good dataset is and how to collect it. By experimenting with real data, the outcomes are already in practically applicable terms. This chapter is divided into two short sections. Section 3.1 describes the experimental setup used to capture data. Section 3.4 discusses possible experimental errors, the potential impact on data collection of these errors, and how they were mitigated.

3.1 Linear Stage

To conduct the experiments initially planned, a simple linear stage was built, which is shown in Fig 3.1. Its key components are the rail and the moving target piece. The rail is rigidly attached to a wooden board. At one end of the rail, a raised block was constructed for the camera to attach to. A right-angled bracket with screw fixing holes was attached to this platform to allow the camera to be attached via its tripod screw thread. The target piece consists of a sliding block of wood with two adjustable screws that allow the target to be rigidly fixed at any point along the rail. On one end of the sliding block is a rigid wooden board, mounted with two screws, used for scenarios where no angle is desired. On the other end is a wood board that can pivot about its vertical axis. The calibration targets are taped to either board, depending on the experiment and the desired dataset properties. There is space on both boards to place checkerboards at some constant offset in the x-y plane. This setup gave relatively precise control over depth. However, the camera's optical axis always intersects the same point on the target (if the camera is setup to point directly along the rail). The problems this may cause are discussed in Section 3.4.

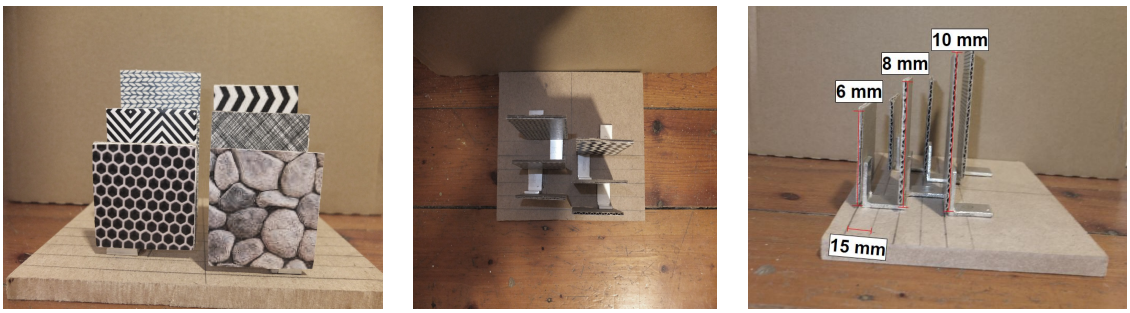
Tape was used to keep the targets in place on the boards. Extra care was taken when attaching the paper to the board to ensure the checkerboard lay flat. Overall, the rail setup could be used to reliably collect data over an 83 cm range, measured from the edge of the platform on which the camera is mounted.

(a) Camera view (b) Overhead view (c) Double-sided target

Figure 3.1: Linear stage setup, demonstrating typical positioning of camera and target

3.2 Target with Fixed Depths

For the validation of the experimental results, a target scene with known depths was constructed. This is a common way to verify the accuracy of calibration methods [7]. A photo of the target is shown in Fig 3.2. The setup mimics the scene constructed by Bok *et al.* in [1]. There are 6 target cards, each 15 mm apart in a staggered fashion. They are made from cardboard, and each is covered with a distinct texture. In Fig 3.2 a, the front two cards are 6 mm high, the middle two cards are 8 mm high, and the back two cards are 10 mm high. All cards are 6 mm wide.



(a) Front view (b) Overhead view (c) Side view

Figure 3.2: Target with known depths. Each target card has a distinct texture so that disparity can be estimated

.

Constructing a target where the depths of objects are precisely known allows calibration parameters to be tested in a realistic depth estimation application. The method of performing depth estimation used in this report uses a disparity map, which requires sufficiently textured scenes. The textures on each board were chosen for their high and distinct textures.

3.3 Image Processing and Decoding

The Light Field Toolbox for Matlab was used to extract raw images from the Lytro[1][2]. This toolbox is designed to convert .LFR files into a 5-D structure described in [2]. For

[1]Available from https://au.mathworks.com/matlabcentral/fileexchange/49683-light-field-toolbox-v0-4

the calibration method used in this research, the raw 2-D format described in Section 2.1.3 and displayed in Fig 2.3 was required. To extract this from the toolbox required making the **LFDecodeLensletImageSimple** return function to save the variable **LensletImage** as a .jpg file after de-vignetting and de-mosaicing had been applied. The devignetting is performed using the white images stored on the camera, and corrects the vignetting within the subimages. Demosaicing is performed using Matlab's **demosaic** function.

Decoding was also performed using the Lytro Power Tools (LPT). The LPT were released for the first two generations of the Lytro camera. They provide complete access and control to the camera firmware via Python and Android Debug Bridge (ADB). The LPT were used to extract the raw images of the target scene as an *external standardized light field* for use in the depth estimation task. In this format, the hexagonal lenslet array is reformatted to a square array. Data was fed into the LPT in the XRAW format from the Lytro Illum. This format contains the white images stored permanently in the camera that are required to decode the .LFR [9]. These square-lenslet images were exported in .png format. Since it was unclear how the reformatting to a square array might affect calibration, the calibration data was also reformatted for the depth estimation task.

3.4 Experimental Errors

There were several stages in the calibration pipeline where errors could occur. The linear stage constructed did not allow for the precision attainable with high-end, purpose made linear stages. In particular, it was not always possible to ensure that targets were precisely orthogonal to the camera's optical axis. The effect of this error was that the intended pose did not always correspond to the actual pose, and subsequently, effects could be mis-attributed.

Another potential error could have occured for datasets where the optical axes for each frame were colinear. It is possible that this setup results in multiple solutions. This would need to be proven theoretically, but should it be the case, it would affect any calibration method, so it is not an easily avoided error if we want to investigate factors without varying all aspects of pose except depth.

When using the Lytro, there was the risk of motion blur when capturing images due to the capture mechanism of a button on the body of the camera. To avoid introducing motion blur in the datasets, the timer feature on the Lytro was used for all image capture. This increased the likelihood of the camera being still at the time of capture, thus decreasing noise in the dataset.

Chapter 4

Preliminary Investigations

This chapter describes the preliminary investigations undertaken. The purpose of conducting the experiments described in this chapter was to test basic assumptions regarding calibration datasets. The aim of each experimental series was to test hypotheses and either come to conclusions that could be tested in a real application, or indicate the need for a more targeted experiment. The results and discussion part of each experiment will be broken up into a section focusing on the performance metrics and a section focusing on the parameter estimates.

4.1 Consistency of Parameter Estimates

4.1.1 Motivation and Hypotheses

To become familiar with the behaviour of the four errors and the parameter estimates, we first wanted to collect a dataset and examine the calibration outcomes for datasets made from every possible combination of images. While theoretically calibration can be performed on a single light-field image, there are most likely many benefits in calibrating on larger datasets - these benefits may manifest in better error measures or more precise parameter estimates. The main hypothesis to test is that larger datasets increase the precision of the parameter estimates. We are also interested in the calibration results for datasets with different pose sets. At this stage it is hard to hypothesise what the relationships between pose and the calibration outcomes will be, which is why the intention is to collect data with some limited pose variation and systematically run every possible combination of images. Then, the sensitivity of the calibration results to the inclusion of each frame can be examined.

4.1.2 Experimental Design

The Lytro Illum was used for this experiment. One dataset was collected with 9 images. The details of the calibration set are given in Table 4.7. In this dataset there is variation in depth, angle, and offset. Fig 4.1 shows the raw data for this experiment. Frames 8 and 9 are the offset and angled images. Because it was economical to run all subsets for a dataset this small, it was possible to examine different types of subsets. The types of subsets we were primarily interested in comparing included subsets with versus without angled frames, subsets with large depth range versus small depth range, subsets with versus without feature points close to the edge of the image. Since the angled images were also offset from the camera's optical axis, an offset effect versus an angled effect will not necessarily be distinguishable from this dataset.

Property	Value
Dataset size	9 frames
Number of features	289 internal corners
Feature symmetry	18 x 18 checkers
Pattern size	7 mm grid size
Pose set	Seven orthogonal images with no x or y offset from the optical axis, at $160, 220, 330, 440, 550, 660$ mm from the main lens aperture. Two frames angled inwards by 20° at distances of 548 mm and 223 mm respectively, with offsets from the optical axis of 77 and 67 mm respectively.
Target type	Checkerboard
Focal distance	Approximately 440 mm from the main lens aperture

Table 4.1: Dataset properties for the proposed experiment

After running the feature estimation stage of the calibration it was observed that for frames 8 and 9, the actual range in the disc radius was within the maximum and minimum for the entire dataset. We would expect angled frames to have a large disc radius range because the checkerboard should be sampling a range of different depths, which correspond to different radii. However, with only a 20° angle (and a large distance for frame 8), the effect was not as noticeable as anticipated. Therefore, we should think of the key difference between frames 8 and 9 and the rest of the frames in the dataset as being offset from the optical axis rather than being angled.

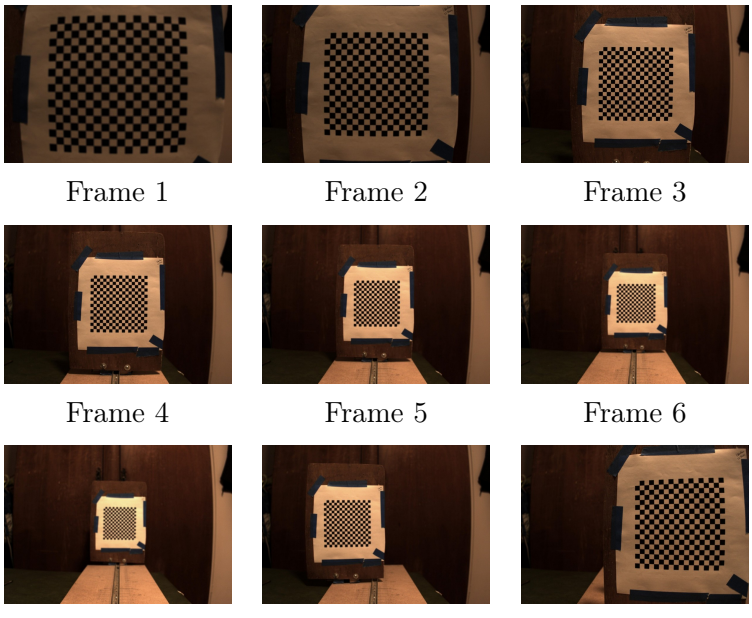
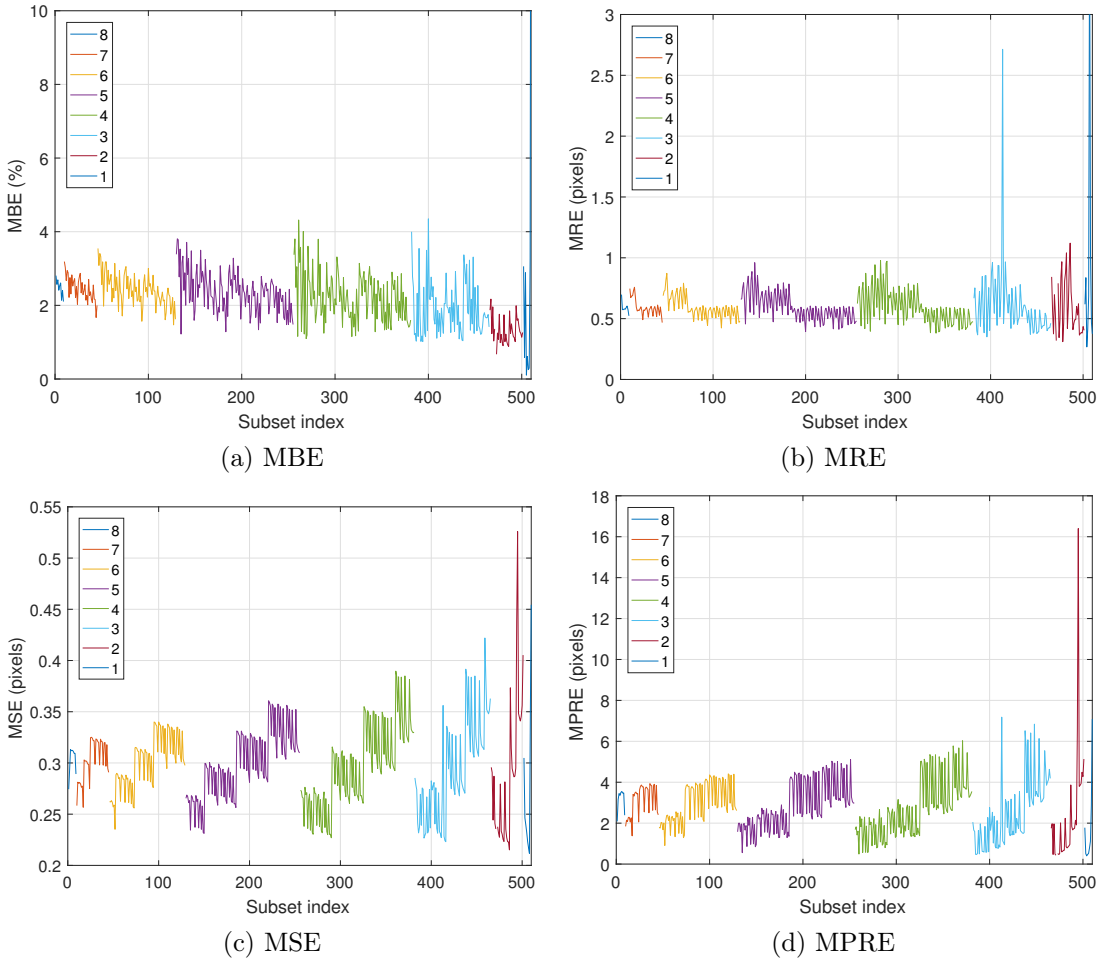


Figure 4.1: All images for dataset collected

4.1.3 Results and Discussion

4.1.3.1 Performance Metrics

Fig 4.2 shows all errors for all subsets. Subsets are ordered largest to smallest left to right, and the legend indicates which colour corresponds to which subset size. The first observation is that subsets with single images (large subset index) produce highly erratic errors. For MRE, MSE, and MPRE, there is a clear increase in variation as the subset size decreases. We also observe a strong correlation between MSE and MPRE in this experiment. The correlation coefficient is 0.9368 (calculated using Matlab's **corrcoef** function). This indicates that MPRE, which is the error the optimisation step uses, does well controlling for MSE but does not necessarily have similar influence over the other errors.



(a) MBE (b) MRE

(c) MSE (d) MPRE

Figure 4.2: All errors for the experiment conducted

The patterns visible in Fig 4.2 are due to the specific ordering of the combinations per subset size. To avoid misinterpreting effects due to this ordering, the sensitivity of each performance metric to the inclusion or exclusion of each frame was calculated separately. Table 4.2 shows the sensitivity of each performance metric to the inclusion of each frame. The sensitivity was measured by calculating the difference between the mean error for all datasets containing frame F and the mean error for all datasets not containing F. Subsets of size 1 were not included in this analysis, because it does not make sense to examine inclusion versus exclusion of a frame for cases where there is only one frame in the dataset.

A positive value indicates that the error is larger if the frame is included, and a negative value indicates that the error is smaller if the frame is included. The frame for which the largest absolute difference is observed is highlighted. This method of examining the sensitivity was chosen because it eliminate possible effects between frames which is useful for a preliminary stage of analysis. Similar analysis was performed on the data per subset size, showing the same behaviour as the aggregate results. The data per subset is presented in Fig 4.3.

Frame	MBE (%)	MRE (pix)	MSE (pix)	MPRE (pix)
1	0.1016	**-0.1582**	**0.0567**	**2.0497**
2	-0.0657	-0.0315	0.0212	0.3023
3	-0.2309	0.0518	-0.0153	-0.2789
4	0.0849	0.0365	-0.0158	-0.2015
5	-0.1158	-0.0052	-0.0169	-0.4779
6	-0.0578	-0.0520	-0.0152	-0.5065
7	0.4964	-0.0992	-0.0118	-0.3836
8	0.0486	0.0489	-0.0054	-0.2557
9	**0.8211**	0.1176	0.0262	1.1625

Table 4.2: Differences between mean error for all subsets that include a given frame versus mean error for all subsets that exclude the same given frame.

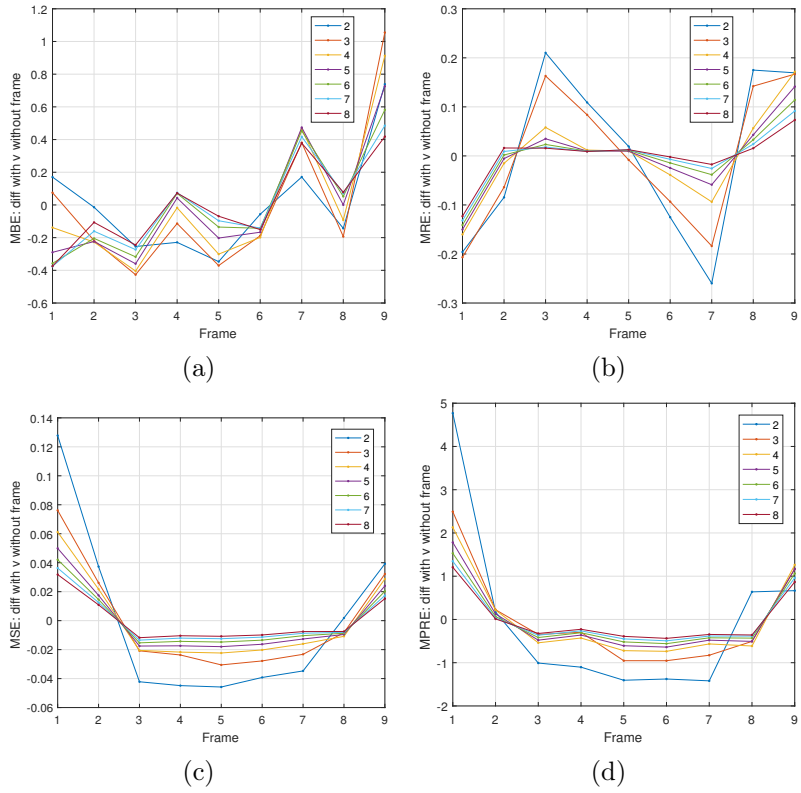


Figure 4.3: Sensitivities of a) MBE, b) MRE, c) MSE and d) MPRE to the inclusion versus exclusion of each frame for different subset sizes (best viewed in colour).

The values for MBE in Table 4.2 indicate that including frame 9 has the biggest effect on MBE, causing it it increase (compared to the individual effect of any other frame). There is no obvious interpretation of the sensitivity of MBE to other frames. This may

suggests that MBE is more affected by the combinations of frames present rather than the inclusion or exclusion of a single frame, which is not captured at this level of analysis. This hypothesis may be supported by observing that in Fig 4.2, MBE does not follow the trend of the other errors and become increasingly variable as the subset size decreases. This is demonstrated by examining the standard deviation of MBE for subset size 2, which is 0.3631, a value that is closest to the standard deviation for subset size 7 of 0.3406.

MRE is most affected by the inclusion of frame 1 which brings the error down. Frame 9 also has a large but opposite effect on MRE, with its inclusion increasing the error. This suggests that frames with large apparent size are good for this error, but also suggests that these frames should be on the optical axis rather than offset.

The results for MSE and MPRE show similar trends. They both indicate that frames with larger apparent size increase the error, as is shown by the large positive sensitivity to frames 1, 2, and 9. The effect may be due to lens distortion, since frames with larger apparent size are more affected by lens distortion by virtue of having feature points closer to the edge of the image, where lens distortion is more severe. The calibration method does use a radial distortion model: however, the checkerboard is initially detected on distorted sub-aperture images, and these detected corners are used to determine the 'ground-truth' disc data. The checkerboard detector used also appears to have its own distortion model, which cannot easily be turned off. Therefore, larger MSE and MPRE may be caused by some interaction between the lens distortion model used in calibration and the distortion model used by the checkerboard detector.

4.1.3.2 Parameter Estimates

Next, the parameter estimates were examined. Table 4.3 shows the results of sensitivity analysis of the intrinsic parameters, which was performed in the same manner as in the previous section. In this case, the sign of the entries in the table do not mean the estimate is better or worse - they indicate whether the effect of the frame on the estimate causes the estimate to be high (positive) or lower (negative). Fig 4.4 shows the plots of all intrinsic parameters for all subsets.

Frame	K_1	K_2 ($\times 10^4$)	f^x ($\times 10^4$)	f^y ($\times 10^4$)	c^x	c^y	k_1 ($\times 10^{-9}$)	k_2 ($\times 10^{-16}$)
1	-2.1670	0.6963	0.0847	0.0820	22.1689	-17.2223	0.1894	-0.6099
2	-0.8169	0.4241	0.0336	0.0329	0.2858	-17.7534	-0.0038	-0.3029
3	-1.6429	0.9080	0.1323	0.1437	6.9072	**27.7499**	0.0494	-0.1340
4	-1.6660	0.8036	0.0349	0.0397	8.7997	11.2549	-0.0248	0.2392
5	-1.0124	0.5409	0.0291	0.0311	-12.7053	-8.9772	-0.0819	0.2767
6	0.3458	-0.5237	-0.0444	-0.0414	-10.4444	-7.2789	-0.0673	0.2161
7	**-2.3639**	0.5076	-0.0479	-0.0452	-10.9188	-5.8011	-0.0090	-0.2790
8	-1.5198	-4.3760	-2.3528	-2.3422	**46.6692**	-21.2811	0.1570	-0.5626
9	-0.5499	**-4.5520**	**-2.3893**	**-2.3739**	-37.9024	1.9096	**0.4588**	**-0.8784**

Table 4.3: Differences between mean parameter estimate for all subsets that include a given frame versus mean parameter estimate for all subsets that exclude the same given frame.

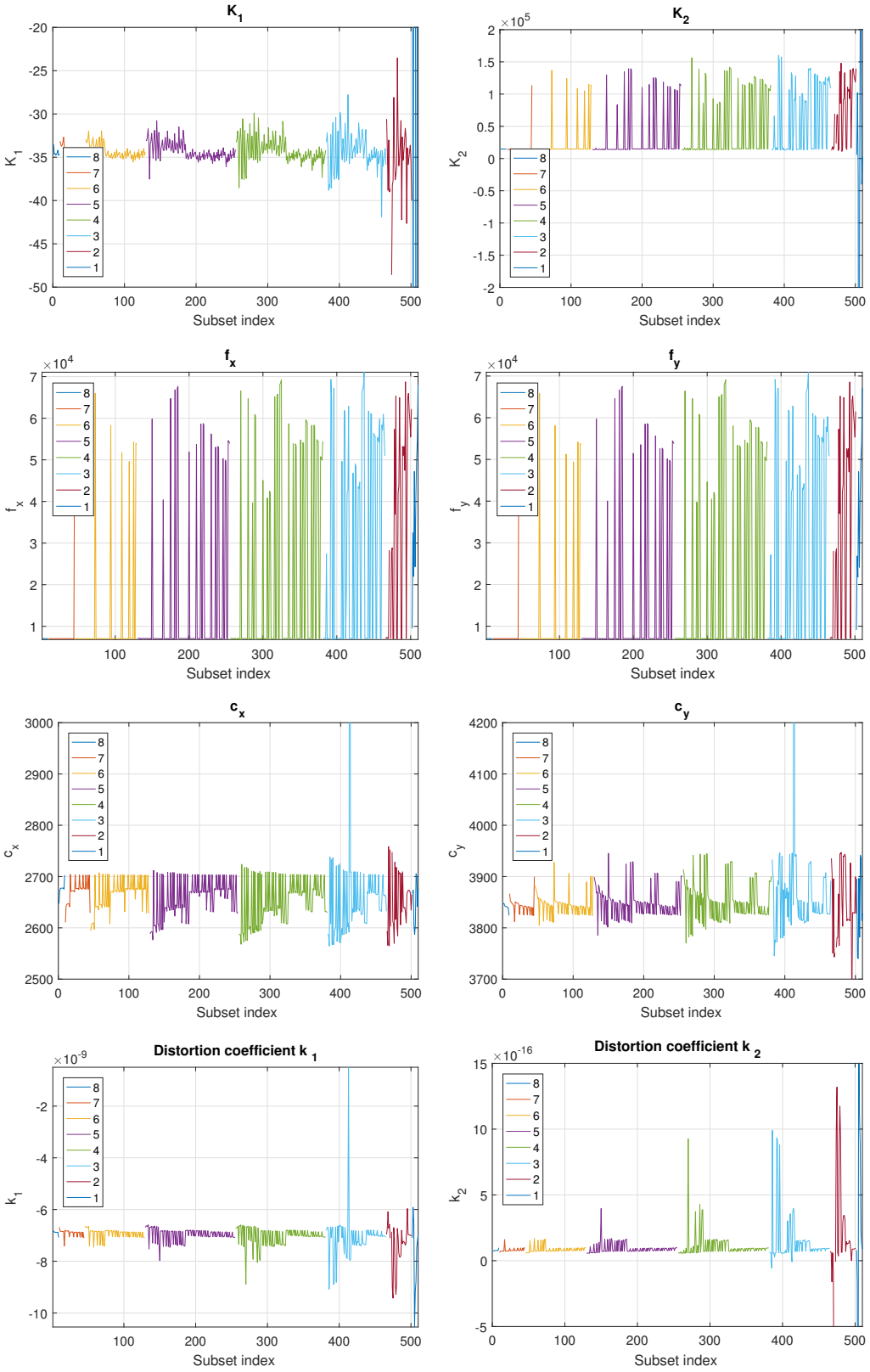


Figure 4.4: Intrinsic parameters for all subsets.

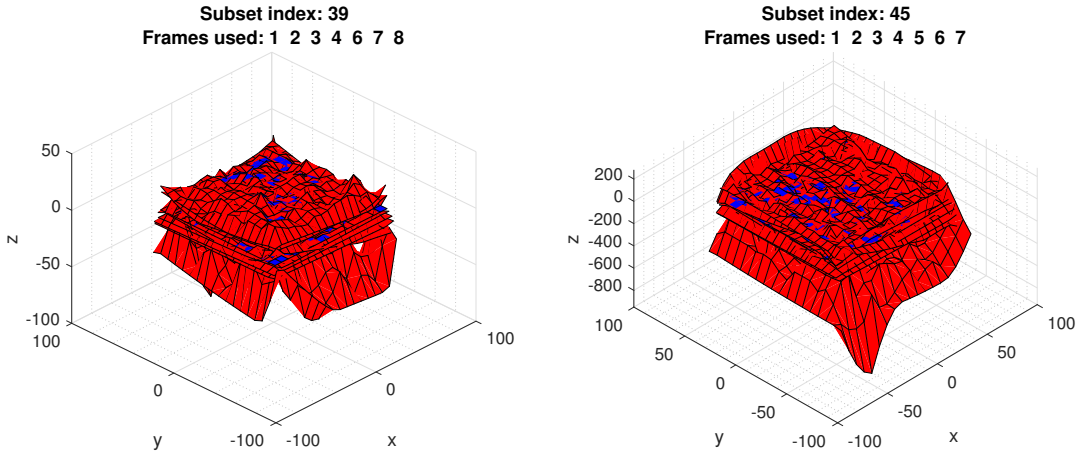
Table 4.3 indicates that K_1 is most sensitive to the inclusion of frame 7 and frame 1, which both decrease the value of the estimate. This may be a consequence of using a small

Frame	K_1	
1	-3.6232	-
2	-1.8877	-1.0293
3	-3.4053	-2.8008
4	-3.5290	-3.3660
5	-2.0798	-1.8512
6	0.9004	1.1847
7	-	-3.8189
8	-3.2053	-3.0501
9	-1.8046	-0.2025

Table 4.4: Sensitivity analysis of K_1 on datasets (a) without frame 7, (b) without frame 1.

dataset, since frame 1 and 7 are the closest and furthest frames respectively. To determine whether this is the case, we can remove datasets containing frame 7 from the sensitivity analysis and observe whether frame 6 shows the same behaviour as frame 7 in frame 7's absence. The same can be done for frame 1. Table 4.4 shows the results of this analysis, which indicate that frame 6 does not have a similar effect to frame 7 in frame 7's absence, because the relative effect of frame 6 does not change. Similarly, in the absence of frame 1, the relative effect of frame 2 remains unchanged. This suggests that K_1 is actually affected by a more complex factor than the inclusion or exclusion of a single frame, such as certain combinations of frames.

Table 4.3 indicates that this factor is the inclusion of frames 8 and 9 cause the K_2 estimates to be low, and in their absence, the K_2 estimate is high. This behaviour is visible in Fig 4.4. This would suggest that offset frames have a large effect on K_2. One way of determining which K_2 value is 'better', and therefore whether including offset frames is good, is to reconstruct the checkerboards. Fig 4.5 compares the 3-D reconstructions of the checkerboard for two subsets, one containing a single offset image, frame 8 (a) and one containing no offset images (b). Observing the z-axis scale, it is clear that the lower value of K_2 produces better checkerboard reconstructions, meaning including offset images produces better 3-D reconstructions of the checkerboard.



(a) Low K_2 estimate

(b) High K_2 estimate

Figure 4.5: Comparison of checkerboard reconstructions for low and high K_2 estimates

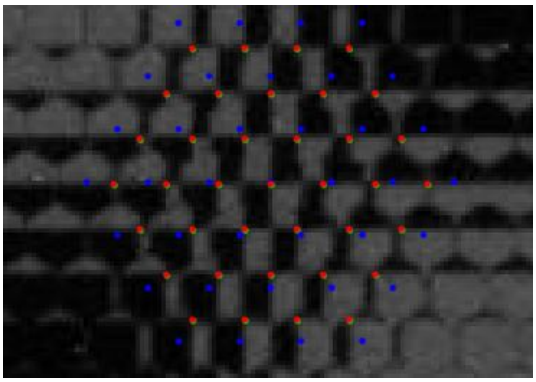


Figure 4.6: Corner pixel reprojections on the raw image using different parameters: green points are reprojections using a low K_2 estimate (subset 39), red points are reprojections using a high K_2 estimate (subset 45), and blue points are the lenslet centres.

We can also compare the reprojections of 3-D points onto the corresponding pixels on the raw image for these two subsets. Fig 4.6 compares the pixel reprojections for a corner point in frame 1 (common to both subset 39 and 45). The reprojections using parameter estimates from subset 39 (low K_2 estimate) are shown in green, while those using parameters estimated from subset 45 (high K_2 estimate) are shown in red. The lenslet centres are shown in blue. There is a small but constant offset between the pixel locations. However, the reprojections are reasonably close together. Table 4.5 shows why this may be the case. The parameter estimates f^x and f^y appear to be compensating for the larger K_2 estimate. If we look back at Table 4.3 we see that f^x and f^y have similar sensitivities, which supports the hypothesis that these parameters are simply compensating for the value of K_2. Overall, this small comparison highlights that the performance metrics do not tell the whole story, and that the parameter estimates need to be considered too.

	Subset 39	Subset 45
K_1	-34.7565	-34.8193
K_2	1.4839×10^4	1.1372×10^5
f^x	7.0118×10^3	5.3751×10^4
f^y	6.9896×10^3	5.3584×10^4
MBE (%)	2.1635	1.9983
MRE (pix)	0.5164	0.4675
MSE (pix)	0.2979	0.2908
MPRE (pix)	2.6297	2.4049

Table 4.5: Comparison of K_1, K_2, f^x, f^y, and performance metrics for subset 39 and 45. A full comparison showing all parameters and performance metrics is given in Appendix B Table B.1.

Examining the centre pixel parameters c^x and c^y shows that c^x is most sensitive to the inclusion of frames 8 and 9, while c^y is most sensitive to the inclusion of frames 3 and 8. It is possible that these effects are due to inaccuracies in the setup, where the centre of the checkerboard was not always located precisely in the centre of the images for frames 1 - 7. Lastly, the distortion coefficients k_1 and k_2 are unsurprisingly most sensitive to frames with feature points close to the edge of the image, with frame 9 having the strongest influence on both parameters, followed by frame 1.

4.1.3.3 Precision of Parameter Estimates

Finally, to address the hypothesis regarding number of frames to use, the standard deviation of the intrinsic parameters was calculated for the different subset sizes used (1 - 8). Fig 4.7 shows that single frame calibration is not adequate for producing precise parameter estimates, and that using more frames is always better if precise calibration parameter estimates are required. The exact number of frames required will depend on the desired precision of estimates, which will be determined by the application. The data used to produce these graphs is tabulated in Appendix B Table B.2. It should be noted that the confidence intervals for each datapoint are different, with the highest confidence values at frames 4 and 5 (the number of samples is included in Appendix B Table B.2).

There is interesting behaviour around dataset size 3 frames for f^x, f^y and c^x, c^y, parameters that are associated with the plenoptic disc centre part of the projection model. This could possibly be a statistical effect. The number of samples starts off low for 1 frame datasets, reaches a maximum at 4 and 5 frame datasets, and then decreases again. While the number of samples is changing, the sample size (number of frames) is also changing. Changing the number of samples has a different effect to changing the sample size, so it is possible that at datasets with 3 frames, there is some crossover between the two effects. A thorough statistical analysis would be required to determine the cause of this behaviour.

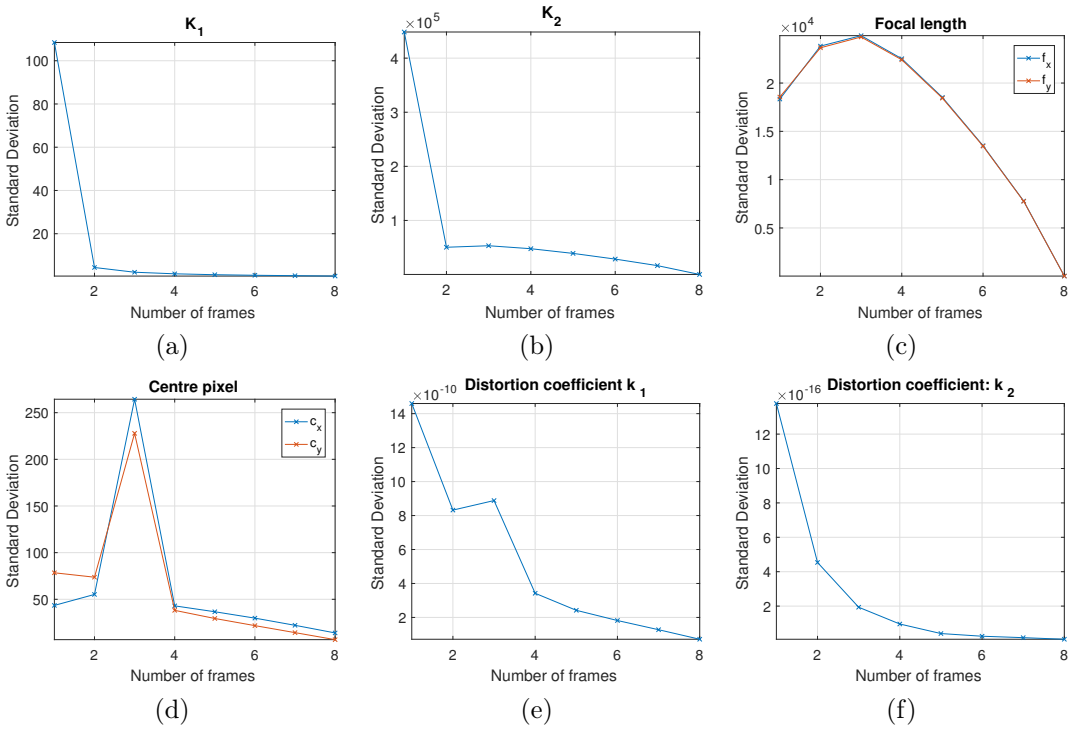


Figure 4.7: Standard deviation for intrinsic parameters (a) K_1, (b) K_2, (c) f^x and f^y, (d) c^x and c_y, (e) distortion coefficient k_1 and (f) distortion coefficient k_2.

In summary, the key results from this experimental series into the behaviour of errors and the consistency of parameter estimates were:

- Larger datasets increase the precision of parameter estimates.

- K_2 estimates are sensitive to whether the dataset contains images with an offset from the optical axis, while K_1 estimates appear to be most sensitive to frames at the ends of the range of depths in the dataset
- Smaller apparent size is better for MSE, MPRE
- The errors and the parameter estimates are not always correlated

4.2 Varying Apparent Size Using Different Grid Sizes

4.2.1 Motivation and Hypotheses

One of first choices made when collecting a calibration dataset is what calibration target to use, and what properties it has. In the case of checkerboards, this corresponds to the physical size of the checkers, how many checkers there are, and whether the checkerboard is symmetric or asymmetric. This experimental series focuses on the first property, grid size (in mm), in order to verify the findings relating to apparent size from the first experimental series.

As discussed in Section 2.1.3, the distance at which the main lens of the camera is focused determines the distance at which an object will appear 'in-focus'. This affects the characteristics of the plenoptic disc - out of focus images will have larger (absolute) plenoptic discs, and in focus images will have smaller plenoptic discs. In addition to this camera geometry, the distance an object is placed from the camera affects its apparent size in pixels in the raw image. This apparent size will affect the position of the plenoptic disc centres in the raw image small apparent grid sizes will have disc centres that are closer together than large apparent grid sizes. The first experimental series showed that smaller apparent grid size is better for MSE, MPRE, and worse for MRE (although it looked like offset interacted with an apparent size effect for MRE). In this experimental series, we want to verify these results by varying grid size, and then observing whether the relationship holds when focal distance is varied. If the observed effect really is caused by apparent size, we would expect larger grid sizes to produce worse MSE, MPRE, but possibly better MRE. Furthermore, we would expect this relationship to hold for any focal distance, since varying focal distance does not change the apparent size.

4.2.2 Experimental Design

The Lytro Illum was used for this experiment. The experiment was conducted by selecting six grid sizes and three focal distances and taking one dataset of 16 images per focal distance per grid size, giving a total of 18 datasets. This experimental design generates 'nested' variation in the two factors of interest: focal distance and grid size. The set of distances were held constant for all 18 datasets. Table 4.6 shows the complete dataset properties. In summary, the only (intended) difference between the 6 datasets per focal distance was the apparent size of the checkerboard in the raw image. The only additional difference between the focal distance sets was the value of the disc radii per image.

The focal distances were intended to be located at the front, middle, and back of the range of distances at which the target is placed. The focus settings on the Lytro indicated that the focal distances were correctly positioned when capturing the dataset, but after decoding, it turned out that the focal distances were actually located as shown in Fig 4.8.

The decision to have the checkerboard orthogonal to the optical axis in every frame was made to eliminate the potential effects of other factors. However, as discussed in Section 3.4, this may create an ill-conditioned calibration problem where there are multiple solutions. This means that the calibration outcome may be highly sensitive to very small variations in input data. However, as a preliminary investigation, it was preferred to hold

Property	Value
Dataset size	16 frames
Number of features	361 internal corners
Feature symmetry	20 x 20 checkers
Pattern size	3.5, 4.5, 5.5, 6.5, 7.5, 8.5 mm grid sizes
Pose set	No rotation around any axis. x and y offset always approximately 0 mm. z positions in 20 mm increments over 300 mm range starting at approximately 185 mm from the main lens aperture
Target type	Checkerboard
Focal distance	Approximately 265, 465, and 625 mm from the main lens aperture

Table 4.6: Dataset properties for the proposed experiment

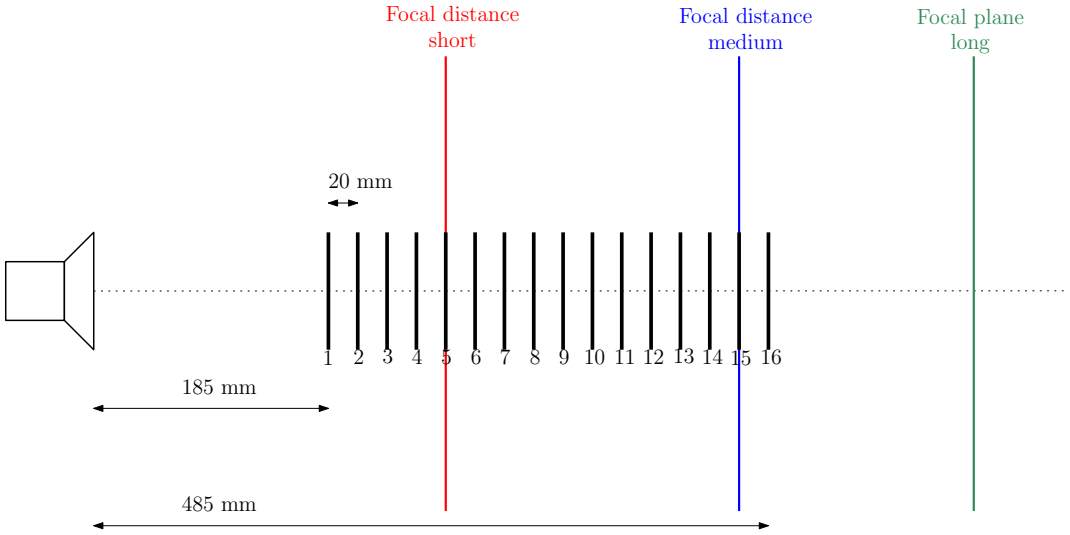


Figure 4.8: Visual illustration of the setup for this experiment

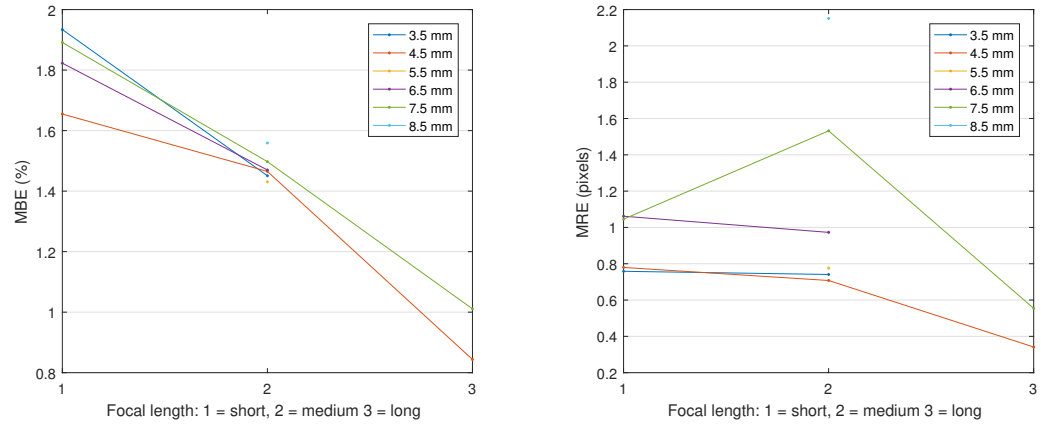
all other factors constant.

4.2.3 Results and Discussion

The performance metrics for each dataset are shown in Fig 4.9. A number of datapoints were removed because they produced unacceptably high MBE of over 100%. Closer inspection revealed that this was caused by parameter estimates for K_2, f^x, and f^y that were several orders of magnitude lower than they should have been, which also affected the extrinsic parameter estimates. This was most likely a result of the ill-conditioning discussed previously. The graphs for all datapoints including those with large MBE can be found in Appendix A Fig A.2. Grid sizes 5.5 mm and 8.5 mm were particularly susceptible, but the reason for this could not easily be determined without detailed analysis of the optimisation algorithm behaviour.

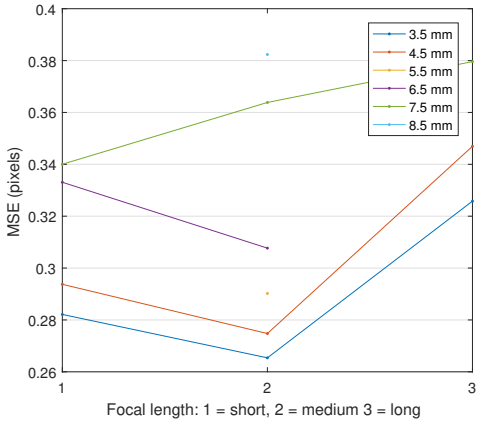
Fig 4.9 shows that for MSE and MPRE, smaller grid sizes are always better, independently of the focal distance. Since smaller grid sizes result in smaller apparent size, this supports the hypothesis that apparent size has a real effect on MSE and MPRE. The results for

MRE could arguably support the same conclusion, although the distinction between large and small grid sizes is less clear. inconclusive with respect to apparent size.

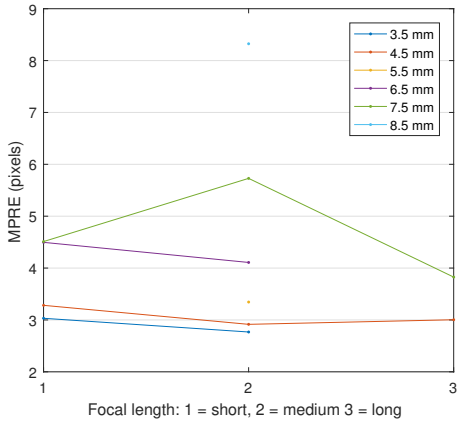


(a) MBE

(b) MRE



(c) MSE



(d) MPRE

Figure 4.9: Performance metrics for the experiment conducted. The different series represent different grid sizes (best viewed in colour)

There are some other interesting trends in the data presented relating to the focal distance. The MBE results indicate that the set of poses (16 orthogonal images at equally spaced distances) is best suited to cases where the focal plane is far behind the furthest frame. It is not clear why this would be the case. The MRE results do not clearly support this, although one could argue that the 4.5 mm series demonstrates similar behaviour to MBE. The MSE results indicate that dataset is best suited to the medium focal distance. There is also some sensitivity at the medium focal distance for MSE, where large grid sizes have higher error compared to the short focal distance, whereas smaller grid sizes have lower error compared to the short focal distance. This sensitivity around the medium focal distance is observed in MRE and MPRE too, but manifests differently, most likely due to the different ways the errors are measured.

The medium focal distance datasets were examined in more detail, since all datasets in this group produced acceptable MBE results. Fig 4.10 shows the errors and parameter estimates plotted against grid size for the medium focal distance dataset. The remaining

intrinsic parameters are shown in Appendix A, Table A.3. These results show the effect of grid size in the absence of variation in any other factor besides grid size. Apart from the small dip in MBE, all errors show similar behaviour, with lower errors for smaller grid sizes. Furthermore, we see that K_1 also has similar behaviour, with lower estimates for smaller grid sizes.

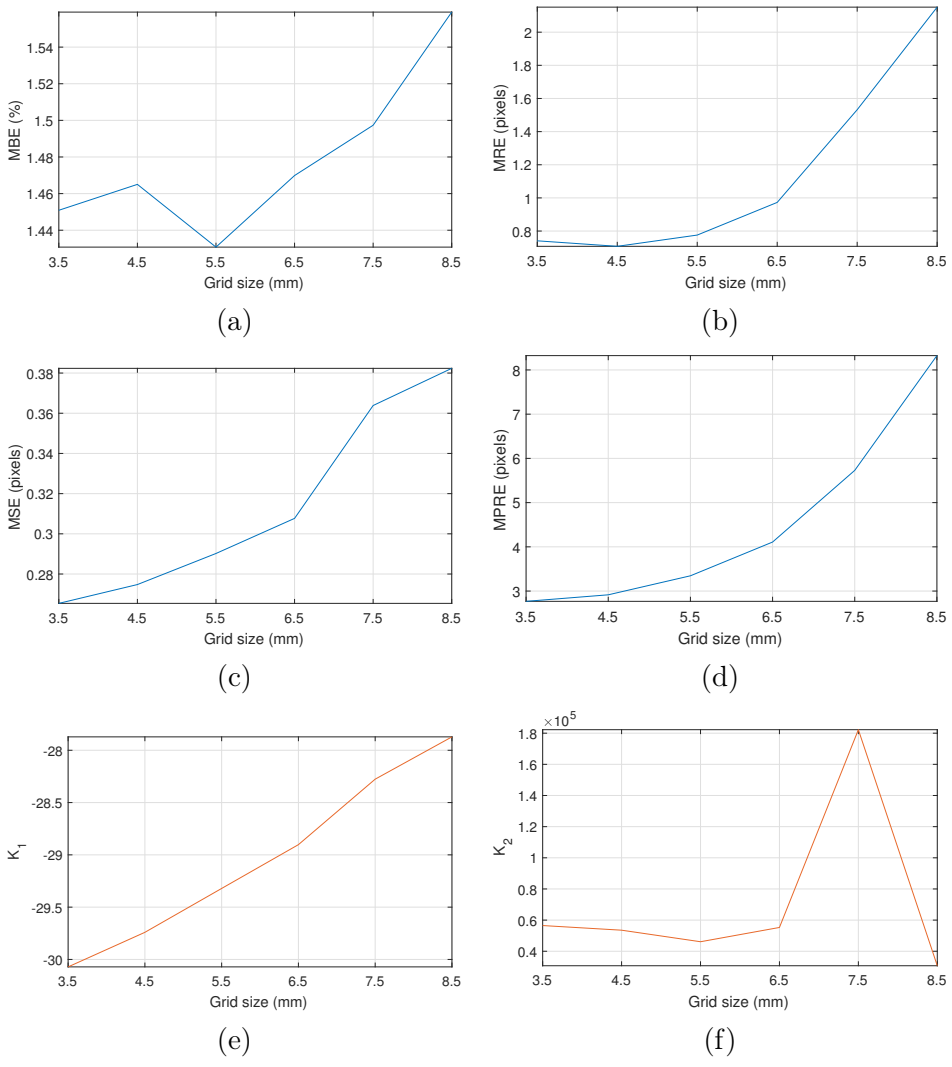


Figure 4.10: Medium focal distance performance metrics and parameter estimates: (a) MBE, (b) MRE, (c) MSE, (d) MPRE, (e) K_1, (f) K_2, (g) focal length of lenslets, and (h) centre pixel.

The behaviour of K_1 may be a product of holding constant too many factors in this experiment, but it does show that larger apparent size increases the parameter estimate. This may simply be a direct response to MPRE, since this is the error optimised for in the calibration routine. To investigate this effect further, a similar dataset would need to be examined where the focal distance is in the middle of the range of distances at which the target is placed. The estimate for K_2 appears to be sensitive to larger grid sizes but in a non-linear matter, which indicates that the results for 7.5 and 8.5 mm may be effects of ill-conditioning where the estimate is responding to very subtle differences in the larger grid size datasets.

Ideally, and experiment like this would be repeated using a camera that suffers negligible lens distortion, such as the Raytrix R42. This would allow us to pinpoint the cause of the apparent size effect. However, there are still key findings that can be taken from this experiment:

- When all frames are orthogonal to the camera's optical axis, the calibration problem becomes ill-conditioned which can manifest in MBE of over 100% and unusable calibration parameters
- With the exception of a few data points, smaller apparent size is associated with better performance metrics (independently of focal distance)

4.3 Angled Images

4.3.1 Motivation and Hypotheses

The introduction of angle into a dataset has potential effects that are of interest. By rotating the target round any axis (except the z-axis), depth variation within a single image can be achieved. It is not clear how this variation within an image contributes differently to depth variation between images. However, there is a clear conceptual difference between these two types of variation. Practically, it will nearly always be true that a more depths can be sampled by introducing angle into a single image than by using many parallel images taken at small depth increments. There is also potential for interaction between the distance at which the angled target is placed and how the angle 'translates' into disc radii variation, as seen in the first experimental series.

Several observations have informed the hypotheses for this experimental series. It has been observed that the calibration algorithm converges more often and with fewer iterations when the dataset contains angled images. It has also been observed that the frame MRE and MSE for individual angled images within datasets are higher than for orthogonal images in the same dataset. The aim of experimenting with angle is to investigate the impact of quantity and 'severity' of angle on the calibration performance measures. The broad hypothesis is that moderately angled images are sufficient for making the calibration problem 'well-conditioned' and reducing MBE to acceptable values, and that severely angled images in a dataset will increase MRE and MSE. We also want to know whether parameter estimates such as K_2 are also sensitive to angle in the same way they are sensitive to other factors (e.g. offset from the optical axis).

4.3.2 Experimental Design

The Lytro Illum was used for this experiment. The optical axis of the camera passed through the centre of the target in each image, and images were taken at 7 distances that crossed the focal plane. At each depth, three images were taken: one orthogonal frame (0° rotation), one frame rotated around the y-axis by 30°, and one rotated by 60°. Rotation around the y-axis was chosen for convenience, but it is important to note that it is not necessarily representative of equivalent rotation around the x-axis. A separate experiment would need to be conducted to determine whether y-axis and x-axis rotation have equivalent affects on calibration. In theory they should, but due to a number of real factors such as lens distortion, this may not be the case. Rotation around the z-axis was not examined at all in this experiment due to issues correctly detecting the orientation of the checkerboard within the calibration routine.

The intention with this experimental design was to calibrate subsets of the whole set. The set was divided into the three frame types: 0°, 30°, and 60°. In the first part of this experiment, 20 random pairs for each combination of frame type were taken. Table 4.8 shows the combinations of frame types in the order they appear in later figures. While calibrating pairs of images has been shown to produce more variable errors/less precise parameter estimates, it is a good place to start because the effects will be more noticeable if there is less data. Once we know what effects are the right ones to look at, a more targeted experiment can be designed that uses larger, more realistic datasets.

Property	Value
Dataset size	21 frames
Number of features	361 internal corners
Feature symmetry	20 x 20 checkers
Pattern size	3.5 mm grid size
Pose set	See Fig 4.11. No x-axis or y-axis offset. Depths of 20 mm increments over 120 mm range. Three angles at each depth, with 0°, 30°, and 60° rotation around y-axis.
Target type	Checkerboard
Focal distance	Approximately 265 mm from the main lens aperture

Table 4.7: Dataset properties for the proposed experiment

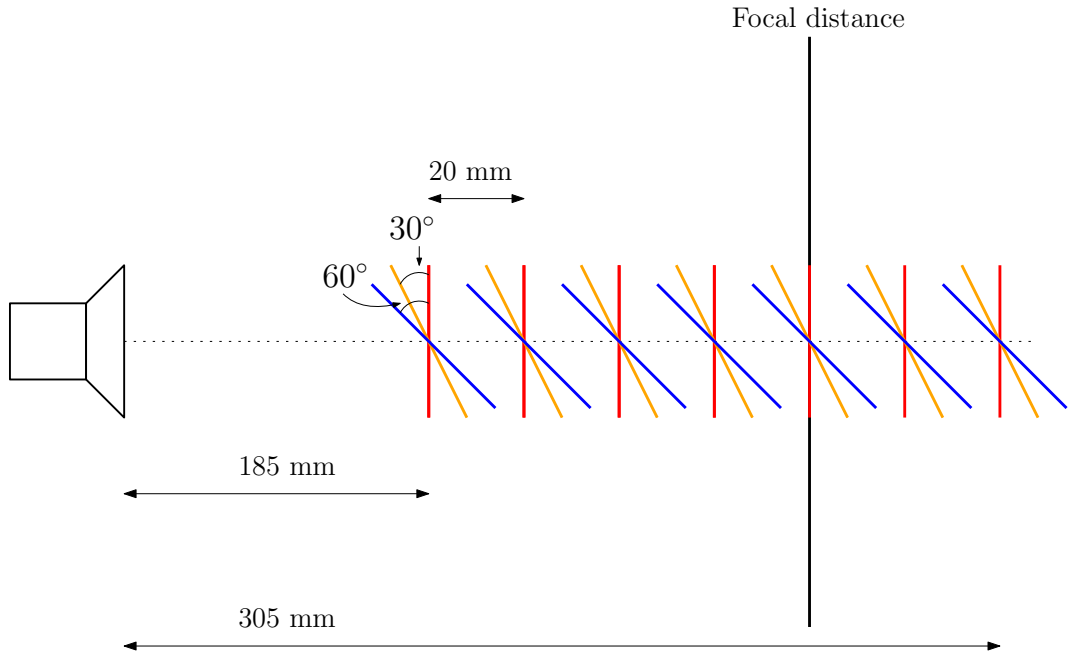


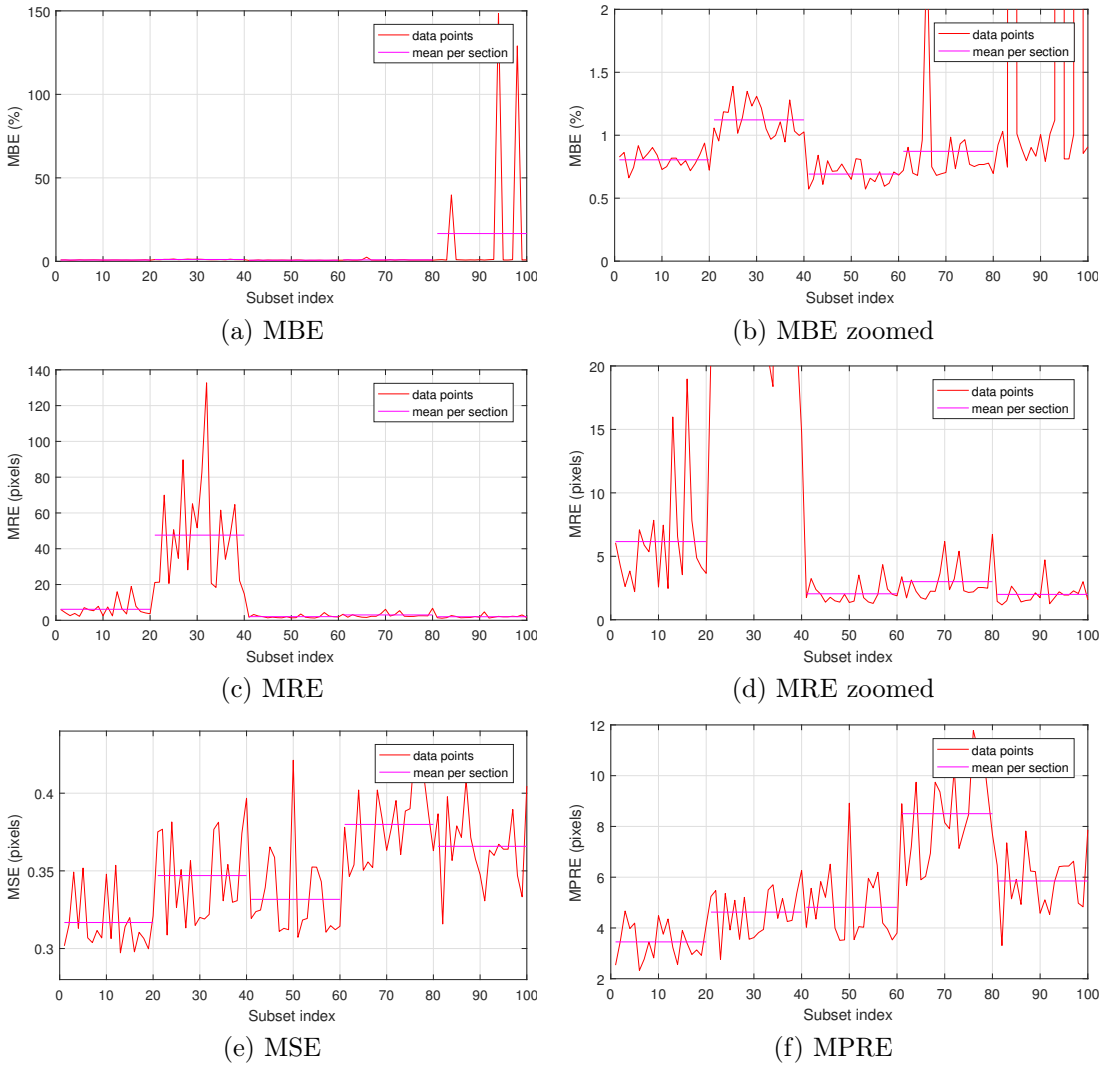
Figure 4.11: Visual illustration of the setup for this experiment

Subset index range	Frame 1 Type	Frame 2 Type
1 - 20	30°	30°
21 - 40	60°	60°
41 - 60	0°	30°
61 - 80	0°	60°
81 - 100	30°	60°

Table 4.8: Structure of frame type combination experiment

4.3.3 Results and Discussion

Fig 4.12 shows the results from the frame type combination experiment. The MBE results indicate that combinations of orthogonal and 30° frames are best, while combinations of 30° and 60° are worst. Further discussion of this behaviour is given later in this section. The best performing group for MRE is (marginally) the 30° and 60° group, which is surprising given how poorly only 60° frames do against MRE. MSE seems to preference the 30° frame only group, as does MPRE. MSE and MPRE are also highly correlated in this experimental series, with a correlation coefficient of 0.9115.



(a) MBE (b) MBE zoomed

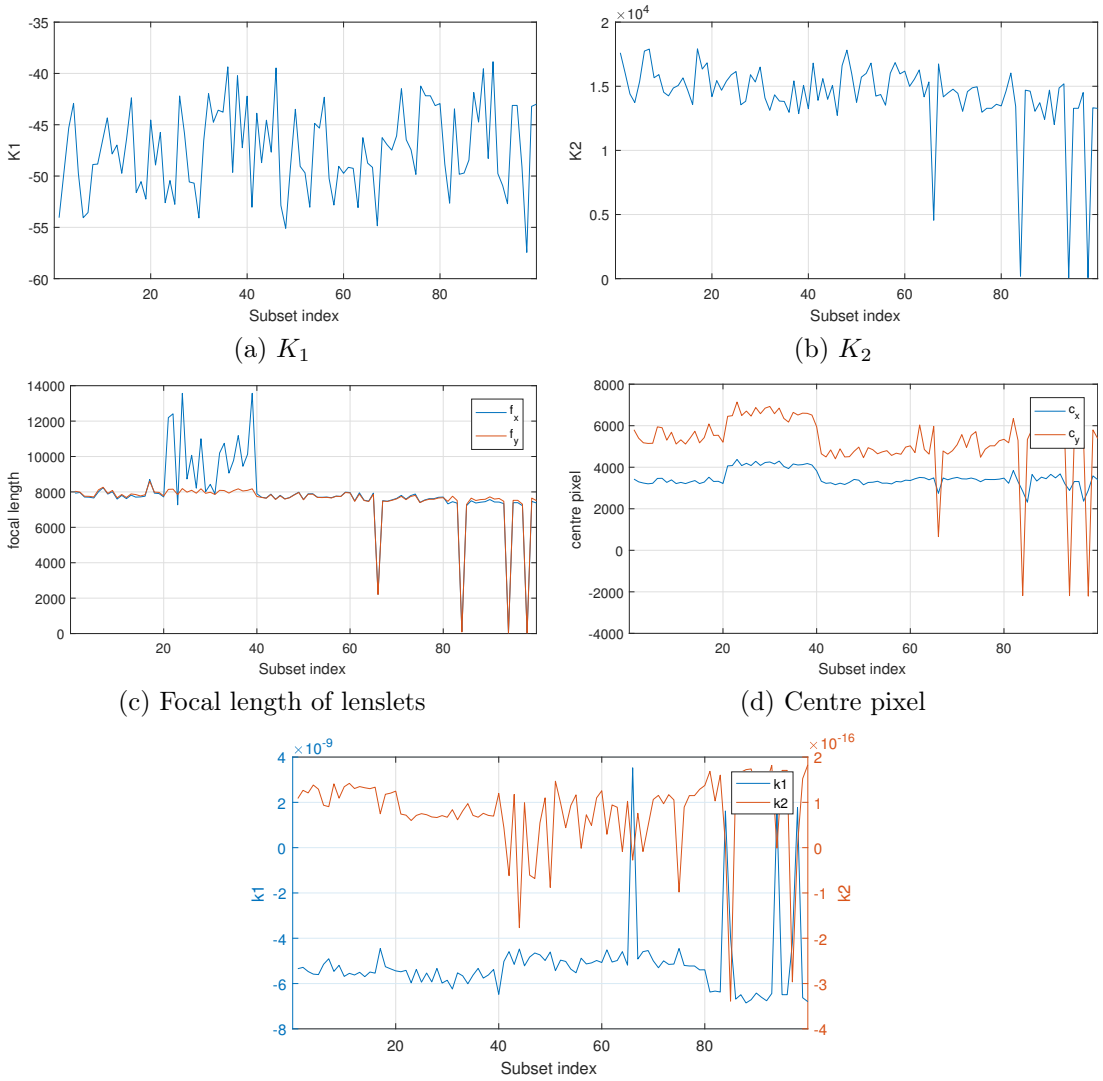
(c) MRE (d) MRE zoomed

(e) MSE (f) MPRE

Figure 4.12: All errors for the frame type combination experiment

We can also examine the parameter estimates, which are shown in Fig 4.13. The variation in K_1 is reasonably consistent for each section, indicating that K_1 is not sensitive to angled versus non-angled frames (at least for pairs of images). K_2 is stable for the first two combination types, which are the sets where samples are drawn from images with the same angle but different depths (e.g. all 30°, or all 60°). This may indicate that having only frames at similar depths is not good for K_2, since for the first two frame type combinations it is not possible to generate a pair where both frames are at the same depth. The last set of frame type combinations with only angle result in very low K_2 estimates, orders of magnitude below what they should be (the same phenomenon seen in Section 4.2). If we compare the locations of the low K_2 estimates we see that they align with large MBE of over 100%. This is a useful result because it shows that MBE will pick up certain inaccuracies in the parameters.

Looking at the other intrinsic parameters shows that K_2, f^x, f^y, c^x, c^y are compensating for one another, since they all spike in the same places (except for f^x, whose behaviour for



(a) K_1

(b) K_2

(c) Focal length of lenslets

(d) Centre pixel

(e) Distortion coefficients

Figure 4.13: All intrinsic parameters for the frame type combination experiment

the second subset group has no plausible explanation). The extrinsic parameters for these spiking datasets show that the extrinsics are also compensating, with estimated depths of less than 1 mm. The distortion coefficients spike at the same subsets as K_2, but the magnitudes are not proportional.

While some of the results were difficult to interpret, there are still several important conclusions that can be drawn:

- Combinations of moderately angled (30°) and orthogonal frames achieve a good trade-off across all errors.
- K_2, f^x, f^y, c_x, c^y estimates are sensitive to combinations of moderate and severely angled frames.
- MBE will pick up low K_2, f^x, f^y parameter estimates, provided they are orders of magnitude below the typical values.

Chapter 5

Primary Investigation

The purpose of the primary investigation is to verify the results obtained from the preliminary experiments in a real world application, depth estimation. Depth estimation extends upon the verification techniques already used such as examining the checkerboard reconstructions and reprojections. The checkerboard reconstructions are not sufficient to fully test the geometric accuracy of the calibration parameters, because the reconstructions rely on extrinsic parameter estimates, and as seen, inaccuracies in parameters can be compensated for by other parameters. Therefore, depth estimation has been selected as an independent task for verification of the results. Additionally, it is still unclear what method of calibration evaluation should be used for a given task. This section will describe the highly targeted experimental series and subsequent validation that was undertaken in order to determine what measure or parameter matters most for depth estimation.

5.1 Hypotheses

In Section 4.1, it was shown that K_2 is highly sensitive to the inclusion of images offset from the optical axis. Through the simple reconstruction of the checkerboard using the estimated intrinsic and extrinsic parameters, we observed that datasets containing offset images produced better reconstructions. The hypothesis is that K_2 estimates produced by datasets containing offset images also give better depth estimates when compared to K_2 estimates produced by datasets without any offset images.

The sensitivities of K_1 were harder to determine in the preliminary investigations, possibly due to its nonlinear relationship in the function relating 3-D points to disc data, and possibly due to the small dataset used. However, the current hypothesis is that K_1 is sensitive to factors relating to depth and apparent size. This is worth investigating in more detail, however, as will be discussed, effects due to K_2 will be more visible in a depth estimation application due to K_1 once again having a non-linear relationship in the depth equation. Therefore, the focus in this experimental series was on K_2.

A final goal of this experimental series was to determine whether calibration estimates that performed better in the depth estimation task could be correlated in some way with any of the performance measures. If the error measure can be correlated with MPRE in particular, this would add support for this error measure as a good choice for the error to minimise in the optimisation step.

5.2 Experimental Design

The Lytro Illum was used for this experiment. Table 5.1 gives the details of the dataset that was collected. The dataset was composed of four distinct groups of images, which are visually represented in Fig 5.1 (the relative distances reflect the actual setup). The different colours in this figure represent 'subsets' with a specific pattern. For example, the red frames form a set of 11 images, all orthogonal to the optical axis, at increasing distance from the camera.

Property	Value
Number of frames	42 frames
Number of features	361 internal corners
Symmetry	20 x 20 checkers
Pattern size	3.5 mm
Pose set	See Fig 5.1. No y-axis offset. Offset from x-axis of approximately 70 mm for all offset frames. Depths of 110, 140, 170, 200, 230, 260, 290, 330, 380, 470, 590 mm from main lens aperture
Target type	Checkerboard
Focal distance	Approximately 200 mm from main lens aperture

Table 5.1: Dataset details for the proposed experiment

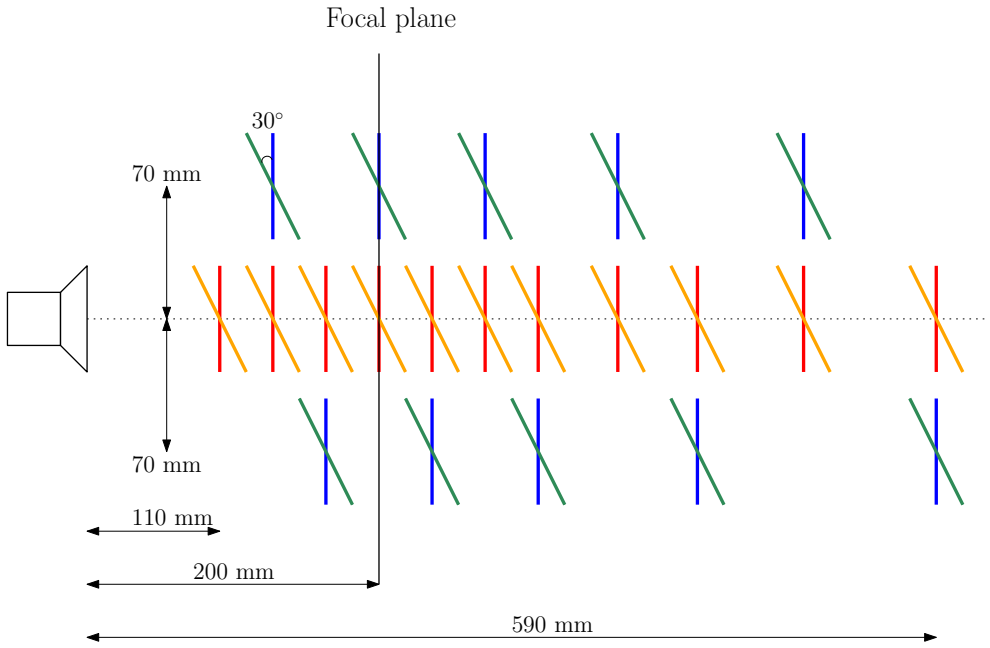


Figure 5.1: Visual representation of the data collected for this experiment. The different colours identify different groups (best viewed in colour)

The intention was to run many randomised subsets drawn from the four groups of images. The groupings of interest were (a) only images with no offset (i.e. all frames lying on the optical axis), (b) only images with offset (i.e. all frames lying off the optical axis), (c) all angled images (which would include both on and off axis images), and (d) and all orthogonal images (which would also include both on and off axis images). Due to the arrangement

of the depths and poses, if a statistically significant number of subsets can be calibrated, this would hopefully distinguish effects such as angle with respect to position (i.e. whether angled frames should be placed on the optical axis or offset, and angled towards or away from the optical axis). Analysis and results from the first two groupings were completed, and the full analysis of the remaining groupings is left as future work.

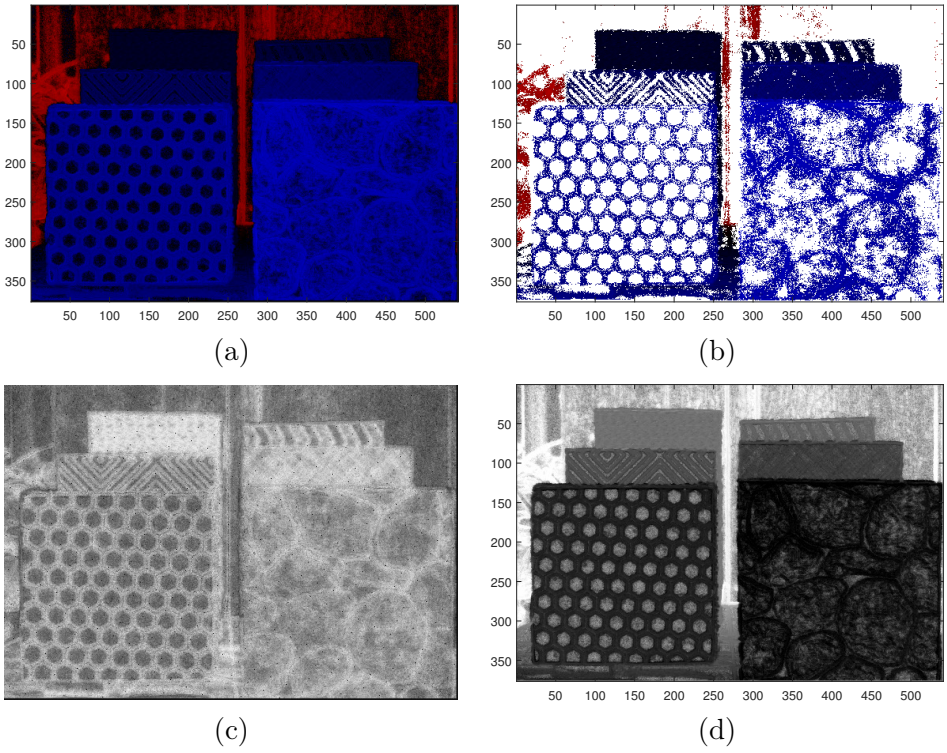


Figure 5.2: Examples of a) disparity map, b) disparity map with confidence < 0.6 filtered out, c) confidence map used to filter (dark regions indicate low confidence), and d) depth map. The different colours in a) indicate regions in front (blue) and behind (red) the focal plane (black).

As well as calibration datasets, images of the target with known depths (see Section 3.2) were also taken at different distances throughout the same range as the calibration data. Disparity maps were calculated for each of these images using the method described in [14]. The disparity map used for analysis is shown in Fig 5.2a. This particular image was chosen because it showed large differences in disparity between the cards. The regions containing these cards of known depth were manually segmented. To get real depth from the disparity map, the following equation was applied:

$$D = -\frac{K_2}{(K_1 + (2 \cdot r^2 \cdot \delta))} \tag{5.1}$$

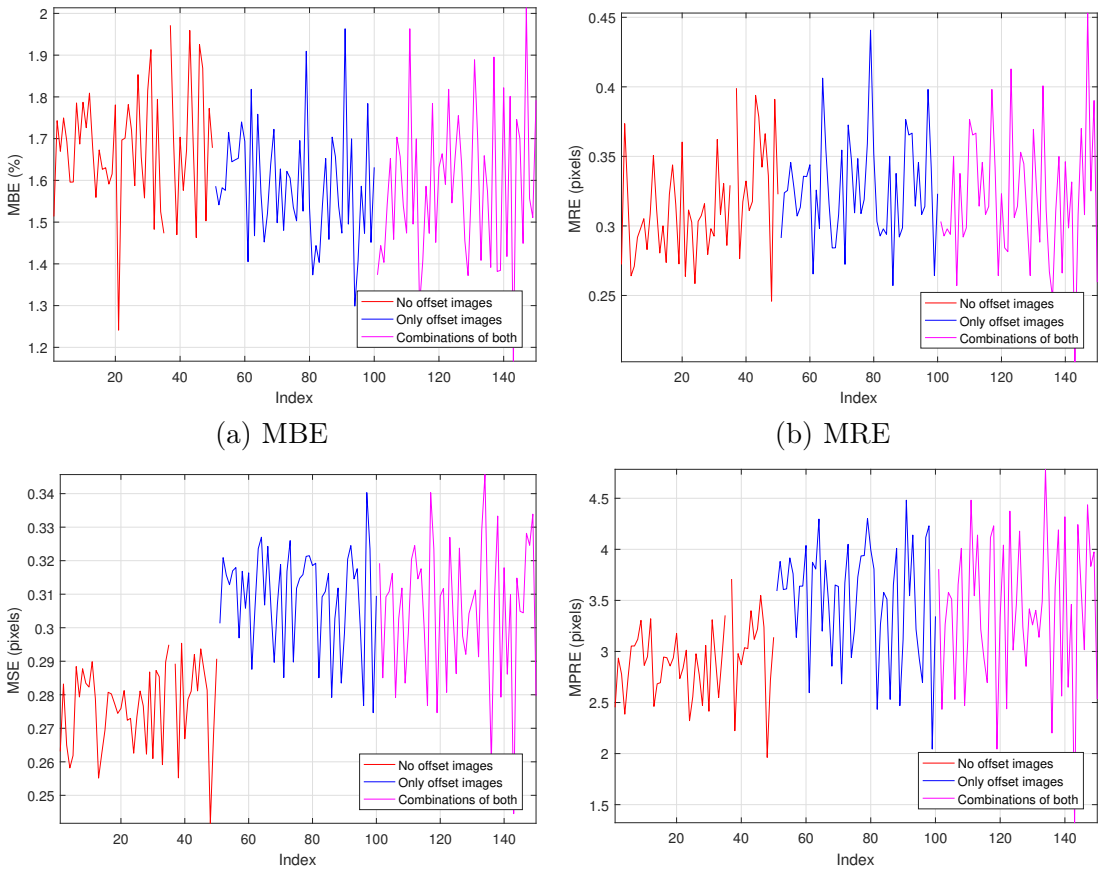
where D is depth, K_2 and K_1 are intrinsics estimated in the calibration algorithm, r is the subimage radius for the camera, and δ is disparity (which was determined using the mean disparity of each segmented region within each card). Disparity estimates with low confidence values were also filtered out. Fig 5.2c shows the confidence values, and Fig 5.2b shows the resulting filtered disparity map. When calculating depth using the

mean disparity, only disparities with confidences of 0.8 and higher were used, which filtered out the problematic disparity values. Once the depths had been calculated using Eqn 5.2, the best parameters were defined as those with the lowest sum of squared difference (SSD) error, where the difference was calculated between the truth depths between adjacent cards and the estimated depths between adjacent cards. This error is measured in mm^2.

5.3 Results and Discussion

5.3.1 Performance Metrics

While the focus of this experiment was not on examining the performance metrics in great detail, they are still worth displaying. Fig 5.3 shows all performance metrics. The behaviour of MSE and MPRE supports the hypothesis that the apparent size effect is related to distortion, since the dataset with no offset images would have the least distortion, which corresponds to systematically lower error measures according to MSE and MPRE. MBE arguably shows the opposite behaviour to the other errors, with the datasets containing no offset images producing systematically higher error according to MBE. The intrinsic parameter estimates are provided in Appendix A Fig A.4.



(a) MBE (b) MRE

(c) MSE (d) MPRE

Figure 5.3: Performance metrics for the data collected in Section 5 (Primary Investigations).

5.3.2 Depth Estimation

Parameters were generated by calibrating random combinations of frames from three dataset 'groups' or 'types'. The first group was all images with no offset from the optical axis (red and orange frames in Fig 5.1). The second group was only images with offset from the optical axis (blue and green frames in Fig 5.1). The last group included the entire dataset, and therefore included both offset and non-offset images.

Fig 5.4 shows the SSD error for the different parameter estimates produced in the manner stated above, with 50 datasets randomly selected. Subsets ranged in size from 10 to 16 images. The index on the x-axis has no meaning, the datasets are simply presented in the order they were generated. It is clear from Fig 5.4b that there is a systematic difference in the error produced by the different dataset types. On average, the parameters produced by datasets containing no offset images are associated with a higher SSD error than the other two dataset types, which both contain offset images. The mean SSD is tabulated in Table 5.2, as well as the minimum SSD error and the parameters that produced the lowest errors. This confirms the hypothesis that datasets containing offset data produce more geometrically accurate calibration parameter estimates.

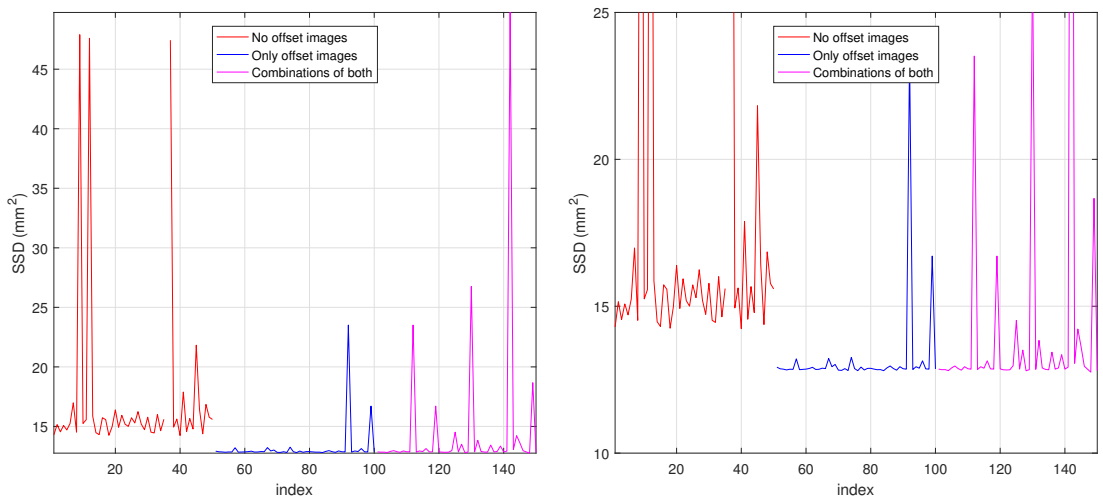
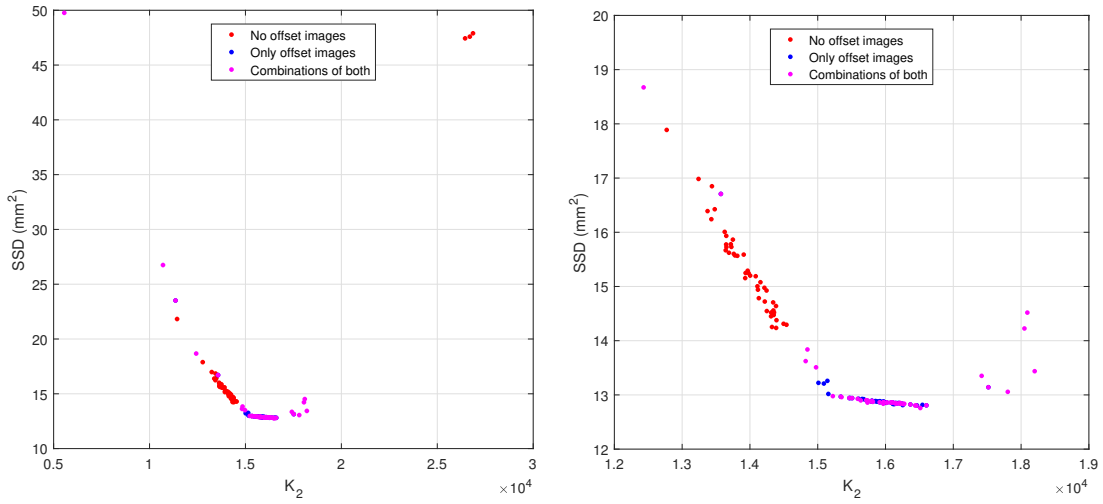


Figure 5.4: Comparison of SSD error for the three datasets types examined. The datasets used to generate these calibration parameter estimates ranged from 10 - 16 images in size.

We also explored the relationship between the estimate for K_2 and the SSD error. This relationship is plotted in Fig 5.5. This figure illustrates again that different dataset types produce systematically different K_2 values. It also shows that these K_2 values have a strong effect on the SSD error. There are a number of interesting observations that can be made about this graph. The first is that while it looks like there should be some symmetry in the shape of the curve, most points are clustered to one side (corresponding to lower K_2 estimates). It is unclear why this is the case, many more samples would need to be computed in order to understand this behaviour more fully.

Table 5.3 compares the estimated depths between adjacent cards with a similar experiment conducted by Bok *et al.* in [1]. It should be noted that no details were given in [1]



(a) (b)

Figure 5.5: K_2 estimates plotted against the corresponding SSD for the three dataset types shown in Fig 5.4.

	Mean SSD	Min. SSD	K_1	K_2
No offset images	17.4207	14.2358	-69.3467	1.4386×10^4
Only offset images	**13.1889**	12.8064	-69.6468	1.6464×10^4
Combinations of both	14.4265	**12.7581**	-69.0229	1.6513×10^4

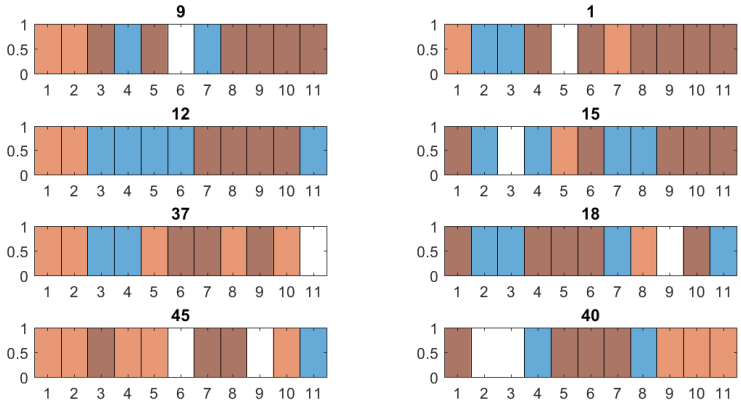
Table 5.2: Mean and minimum SSD for the different types of datasets. The parameters that produced the minimum SSD are also shown.

of the setup other than the relative depths being 15 mm. It is not known what the focal settings of the camera were, or where the target was positioned with respect to the camera. The actual size of the cards in the scene is also unknown. However, this shows that the parameters perform quite well compared with the ground truth values, with a maximum deviation of 4 mm, compared to 13 mm for [1]. Furthermore, it is likely given the pattern in the depth estimates that the target was placed on a slight angle with respect to the camera, which is why the depths between every second card are systematically lower. This may mean that the parameter estimates are more precise than the SSD error indicates.

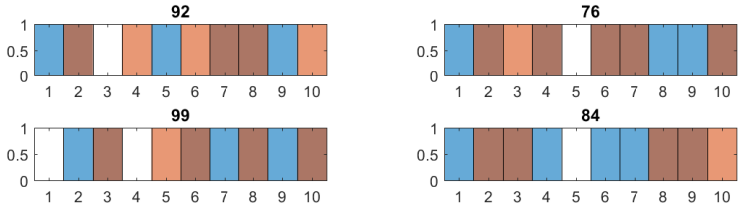
Adjacent Cards	Ground Truth (mm)	Bok *et al.* [1]	Best estimate (mm)
(1) - (2)	15	17	13.3271
(2) - (3)	15	14	16.6362
(3) - (4)	15	8	11.2232
(4) - (5)	15	4	20.6317
(5) - (6)	15	2	14.9595

Table 5.3: Comparison of estimated distances between adjacent cards (as seen in Fig 5.2) to a similar experiment in [1]. The best estimate uses the parameters from Table 5.2 in the 'Combinations of both' row.

So far we have seen clear evidence in support of including offset images in calibration datasets to achieve good depth estimation results. However there are still peaks evident



(a) High SSD (b) Low SSD



(c) High SSD (d) Low SSD

Figure 5.6: Histograms comparing the frames used for (a) high SSD errors for no offset dataset types, (b) low SSD errors for no offset dataset types, (c) high SSD errors for all offset dataset types, and (d) low SSD errors for all offset dataset types. Angled frames are orange, orthogonal frames are blue, and in instances where both the angled and orthogonal frame at the same distance were used, that region appears dark brown. Frame 1 is closest while frames 11 and 10 are furthest from the camera for the respective groups.

in the SSD error for all dataset types. We explored the peaks that occur in the first two dataset types by closely examining the frames used for the datasets that produced parameters that produced large SSD error. These frames were compared to those in datasets that produced parameters that produced very low SSD error. This comparison may not be representative of the overall behaviour, but it is a reasonable place to start. The results are presented in the form of several histograms in Fig 5.6, where the frame number is a proxy for the depth of the frame, and the colour indicates whether the frame is orthogonal or angled. This representation gives a sense of the distribution of frames both throughout the depth range, and whether there were multiple frames at any depths (these appear brown in the figure). The title number is the index number that can be related back to Fig 5.4.

Fig 5.6a and b compare the frames used for local maxima and minima in SSD for datasets without offset images. The high SSD datasets seem to have (a) only angled frames at the front of the depth range, and (b) angled frames at depths with no accompanying orthogonal frame at the same depth. If we look at Fig 5.6c and d, which make the same comparison but for datasets with only offset images, that trend is less clear. Fig 5.6c and d is harder to interpret, but may indicate that frames are more clustered (e.g. index 84) for the low SSD datasets and more 'sparse' for the high SSD datasets. It is not possible to draw concrete conclusions from this small, surface-level investigation, but these results do indicate that

there is some effect relating to the distribution of images that has a large effect on SSD, which itself is highly correlated with K_2.

Finally, we tried to relate the performance as measured by SSD in the depth estimation task to the original performance metrics. Table 5.4 shows the correlation coefficients per dataset type. The strongest correlations appear for the no offset datasets, with MPRE most strongly correlated with SSD. However, the correlations for the other two dataset types are low, and the rankings are not consistent. This tells us that either the errors do relate to SSD but in a non-linear way, or that the errors do not correlate with SSD very well. Ideally, these coefficients would be calculated on a much larger sample size, while also systematically varying factors such as dataset size over a wider range, but the current conclusion that can be drawn is that the errors do not correlate very well with SSD, and therefore are not a good way to evaluate the calibration outcome for a depth estimation task like the one presented.

	SSD	MBE	MRE	MSE	MPRE
No offset	1.0000	0.2368	0.1394	0.1728	**0.4162**
Only offset	1.0000	-0.1397	0.0530	0.0011	**-0.1422**
Combinations of both	1.0000	0.1182	**0.1252**	0.0631	-0.0040

Table 5.4: Correlation coefficients between SSD and calibration performance metrics for each dataset type

5.4 Summary

The experiments conducted in this section have shown that accuracy in a depth estimation task depends heavily on the estimates for K_2. It was previously concluded that parameters produced on datasets including offset data improved 3-D reconstructions when compared to the 3-D reconstructions produced by parameters obtained from datasets with no offset images. This verification in a depth estimation has shown that datasets containing offset images are almost always sufficient for producing accurate K_2 values. However, we did see that there are still other factors in play that can cause the K_2 values to be inaccurate. A small investigation points towards the distribution of 3-D points sampled by each frame in the dataset as the cause, but much more detailed analysis is required to determine the precise relationship.

Angle as a factor should have been investigated in tandem with offset in this experimental series, because the conclusions of this experiment rely on the assumption that the effect of angle is unchanged when an angled frame is moved off the optical axis (i.e. any change that is observed in the calibration is due to the frame being offset rather than being angled and offset). This assumption should have been tested. The first place to start with this testing would be to extract datasets where the only difference is angle (i.e. one dataset has an orthogonal frame at position X, and the other has an angled frame at position X, and all other frames are the same), and examine the difference in both calibration performance metrics and parameter estimates.

There is still a huge amount to discover with the dataset collected for this primary investigation. As discussed, two groups were not analysed at all (all orthogonal frames and all

angled frames) and could be examined in a similar way. The recommended approach is to run a statistically significant sample of random subsets of different sizes from these two groupings. After this is complete, the depth estimation task can be performed and the errors examined according to the same method used above. If a statistically significant number of datapoints can be generated, this will allow for more complex statistical analysis that can examine interactions between factors, which will provide very valuable insight.

Chapter 6

Conclusions and Future Work

The effects of certain calibration dataset properties on calibration outcomes have been investigated. One of the first findings was that different methods for evaluating calibration will tell you very different stories about the calibration quality. In particular, the performance metrics MBE, MRE, and MSE, all of which are commonly used, do not correlate well with the values of the calibration parameter estimates.

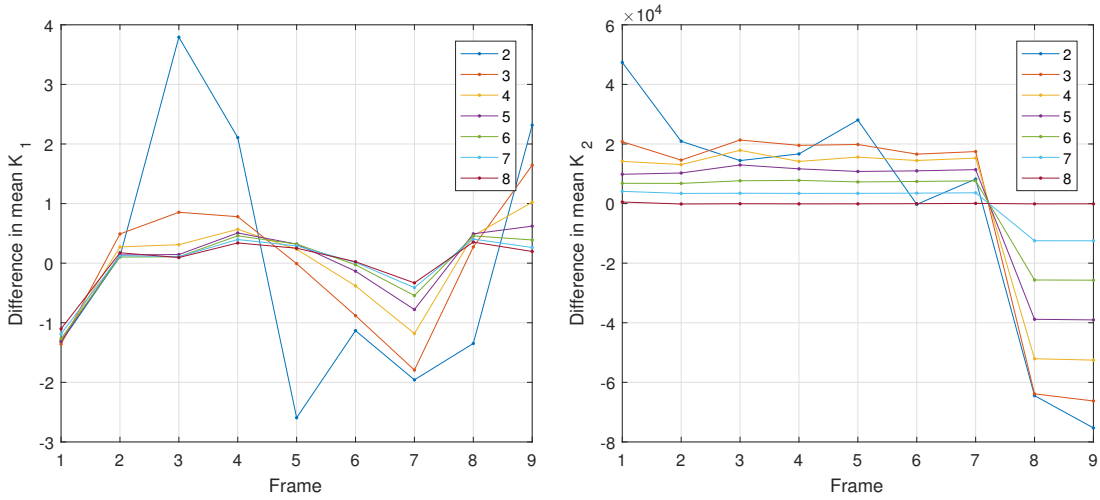
The recommendations that can be provided based on the research conducted so far are to use large datasets for calibration (provided that there is variety of pose in these datasets), and include many offset images at different distances. This should ensure that the parameter estimates are precise and accurate, although it may not produce particularly low error measures. Severe angles should be avoided, as they can cause large variation in the parameter estimates. While a concrete conclusion regarding the distribution of frames within the space in front of the camera cannot be drawn without more data and more detailed statistical analysis, there is clearly an effect most likely related to the distribution of 3-D points that the frames 'collectively' sample. As an preliminary recommendation, we advise that the dataset should contain frames that are near other frames, particular if one frame has large variation in one or more aspects of pose (by near, we refer to both depth and offset).

There are many areas of future work in this space. One of the first tasks would be to complete the analysis for the entire dataset collected for the primary investigation. From the primary investigation it was clear that the density and uniformity of the 3-D points sampled affects the K_2 estimates, which affect the accuracy in depth estimation tasks. This is a clear avenue for future work. It would involve a more detailed statistical approach, where we control for the density and uniformity of distribution of either points in 3-D space or some parametrization of these points that relates more specifically to checkerboard pose (such as angle around its central axis, position of its centre, and depth of its centre).

As noted, not all factors that relate to calibration datasets were investigated in this research. The target symmetry and target type would be useful to investigate to a similar level of detail. This would add value because practically speaking, these factors are easy for anyone performing calibrating to control. Another useful task would be replacing the current error minimised in the optimisation step with other error measures (specifically MBE, MSE, and MRE) and observing the effect on the calibration parameter estimates. This would be useful in determining whether a different error measure should be optimised to produce results that can be better correlated with the errors.

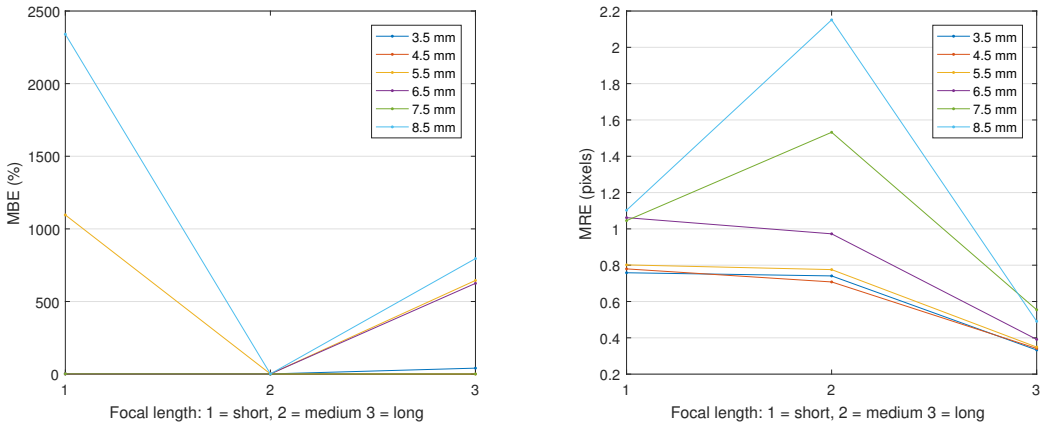
Lastly, different calibration methods could be used on the data collected to verify the findings of this research. This would add value by demonstrating whether the observed effects generalise, or are simply artefacts of the particular calibration method used. This would also assist in standardising the comparison of calibration methods, which is typically done on whatever datasets authors make available when publishing new calibration methods. As discussed at the very beginning of this report, it is conceivable that different methods will favour different datasets. Therefore, performing similar analysis on other calibration methods becomes very valuable future work because it could bring a greater level of standardisation to the comparisons of calibration methods.

Appendix A: Additional Figures



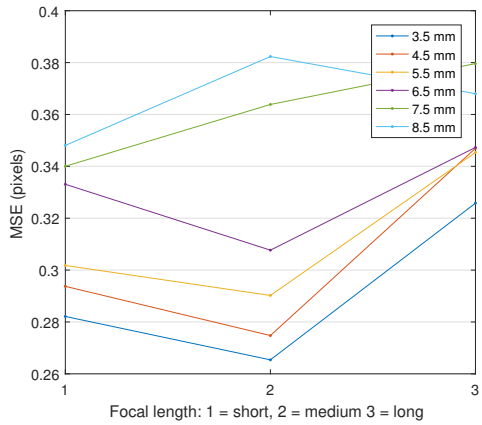
(a) K_1 (b)K_2

Figure A.1: Sensitivity graphs for (a) K_1 and (b) K_2 from Section 4.1.3.2. These graphs show the differences between mean parameter estimate for subsets that include a given frame versus mean parameter estimate for subsets that exclude the same given frame. The results are separated by subset size.

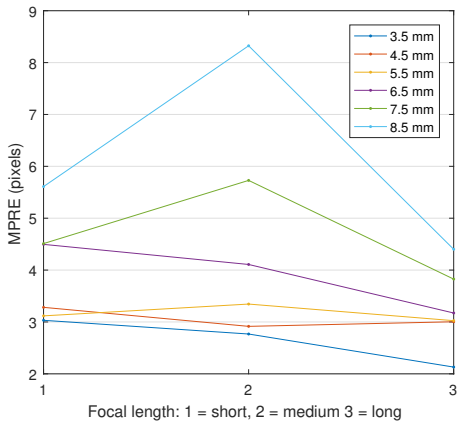


(a) MBE

(b) MRE



(c) MSE



(d) MPRE

Figure A.2: All performance metrics for all datapoints in the grid size experiment from Section 4.2. (best viewed in colour).

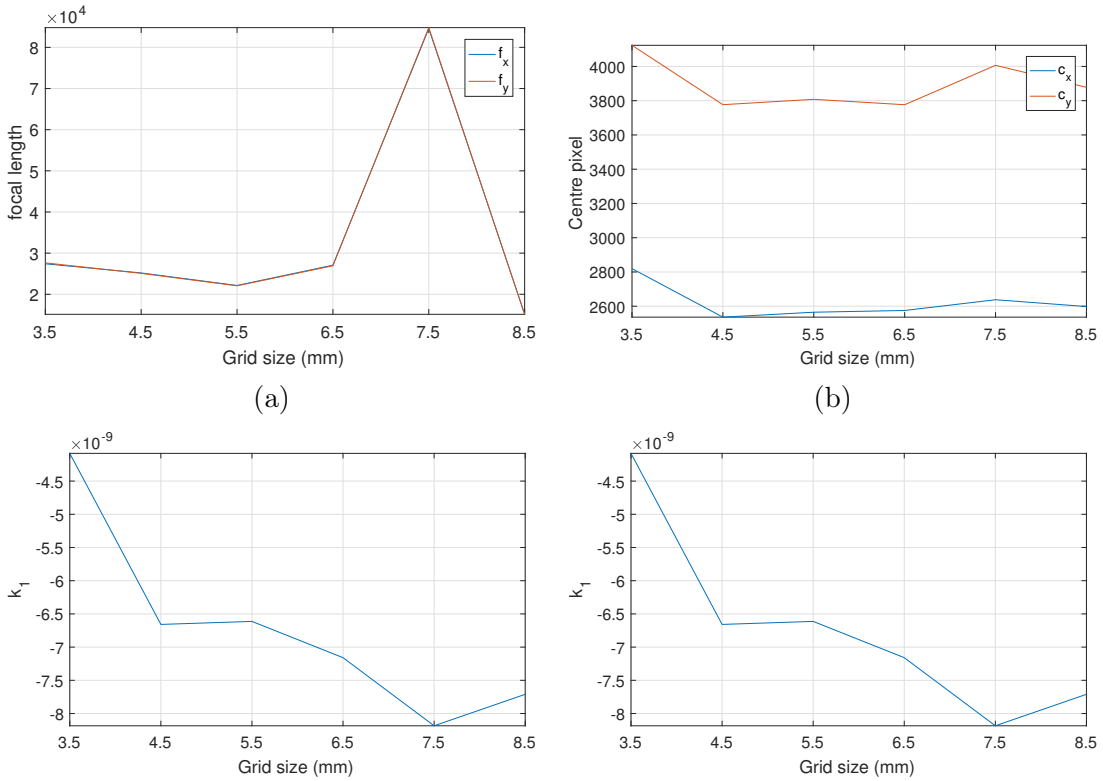
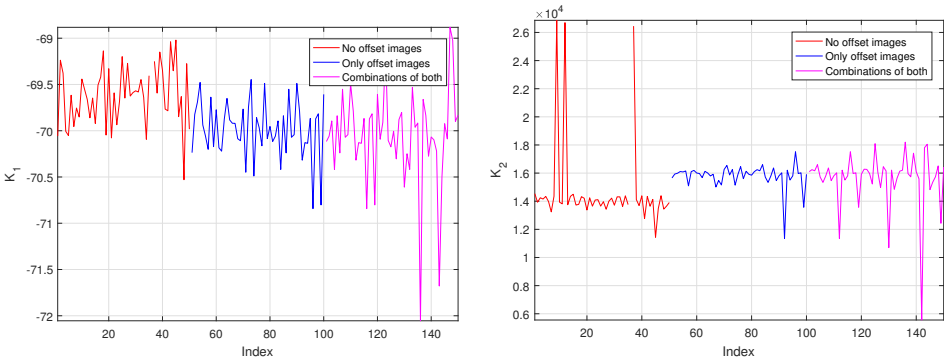
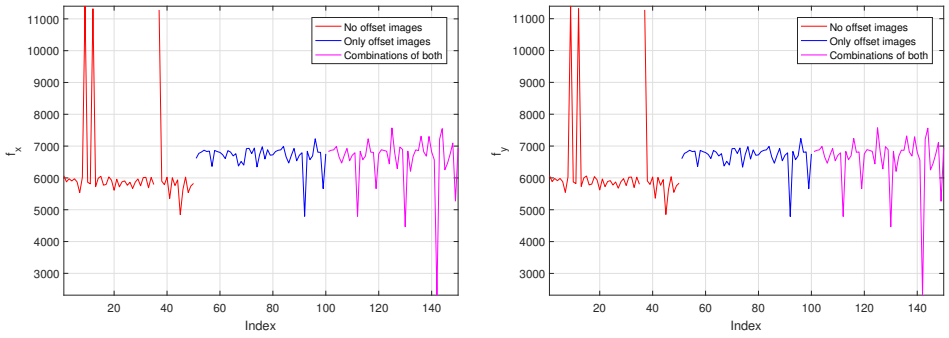


Figure A.3: Intrinsic parameters f^x, f^y, c^x, c^y, k_1, and k_2 from Fig 4.10 in Section 4.2



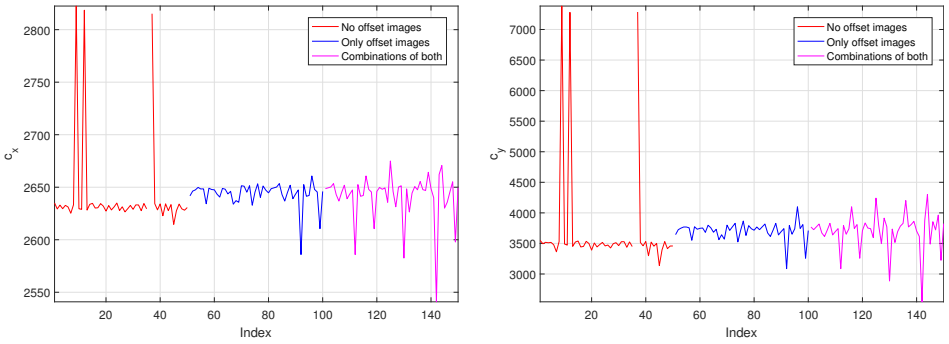
(a) K_1

(b) K_2



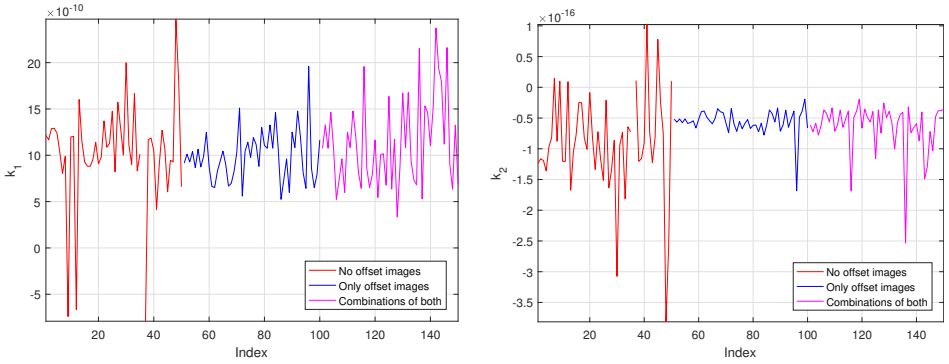
(c) f^x

(d) f^y



(c) c^x

(d) c^y



(c) k_1

(d) k_2

Figure A.4: All intrinsic parameters for the data collected in Section 5 (Primary Investigations).

Appendix B: Additional Tables

	Subset 39	Subset 45
Intrinsic Parameters		
K_1	-34.7565	-34.8193
K_2	1.4839×10^4	1.1372×10^5
f_x	7.0118×10^3	5.3751×10^4
f_y	6.9896×10^3	5.3584×10^4
c_x	2.7027×10^3	2.6291×10^3
c_y	3.8246×10^3	3.8998×10^3
k_1	-7.0711×10^{-9}	-7.0818×10^{-9}
k_2	9.6984×10^{-17}	9.8742×10^{-17}
Performance Metrics		
MBE (%)	2.1635	1.9983
MRE (pix)	0.5164	0.4675
MSE (pix)	0.2979	0.2908
MPRE (pix)	2.6297	2.4049

Table B.1: Comparison of all intrinsic parameters and performance metrics for subset 39 and 45 (from Section 4.1.3.2, extension of Table 4.5).

Subset size	1	2	3	4	5	6	7	8
Sample size	9	36	84	126	126	84	36	9
K_1	108.4332	4.3750	2.1889	1.4574	1.0434	0.7813	0.5891	0.4104
K_2 $(\times 10^5)$	4.4831	0.5045	0.5326	0.4777	0.3911	0.2858	0.1651	0.0018
f^x $(\times 10^4)$	1.8343	2.3826	2.4907	2.2522	1.8517	1.3518	0.7784	0.0019
f^y $(\times 10^4)$	1.8551	2.3653	2.4769	2.2410	1.8436	1.3468	0.7760	0.0019
c^x	43.4605	55.2068	264.4112	42.9745	36.6613	29.8212	22.1247	13.9639
c^y	78.4191	73.6810	227.8615	38.1634	29.4453	21.7885	14.2874	6.7613
k_1 $(\times 10^{-8})$	0.1460	0.0832	0.0888	0.0342	0.0242	0.0182	0.0128	0.0072
k_2 $(\times 10^{-14})$	0.1378	0.0454	0.0193	0.0096	0.0040	0.0025	0.0017	0.0007

Table B.2: Standard deviation of intrinsic parameters per subset size (from Fig 4.7 in Section 4.1.3.3).

Bibliography

1. Y. Bok, H.-G. Jeon, and I. S. Kweon. Geometric calibration of micro-lens-based light-field cameras using line features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(2):287–300, Feb. 2017. (cited on pages 4, 8, 10, 13, 17, 42, and 43)
2. D. G. Dansereau, O. Pizarro, and S. B. Williams. Decoding, calibration and rectification for lenselet-based plenoptic cameras. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1027–1034, June 2013. (cited on pages 4, 6, 8, 12, and 17)
3. A. Geiger, F. Moosmann, O. Car, and B. Schuster. Automatic camera and range sensor calibration using a single shot. *2012 IEEE International Conference on Robotics and Automation*, pages 3936–3943, 2012. (cited on page 12)
4. H. Ha, M. Perdoch, H. Alismail, I. S. Kweon, and Y. Sheikh. Deltille grids for geometric camera calibration. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5354–5362, Oct 2017. (cited on page 12)
5. C. Hahne. The standard plenoptic camera. http://www.plenoptic.info/, 2018. (cited on pages 5, 6, and 7)
6. C. Hahne, A. Aggoun, V. Velisavljevic, S. Fiebig, and M. Pesch. Baseline and triangulation geometry in a standard plenoptic camera. *International Journal of Computer Vision*, 126(1):21–35, Jan 2018. (cited on page 4)
7. C. Heinze, S. Spyropoulos, S. Hussmann, and C. Perwass. Automated robust metric calibration algorithm for multifocus plenoptic cameras. *IEEE Transactions on Instrumentation and Measurement*, 65(5):1197–1205, May 2016. (cited on page 17)
8. M. Levoy and P. Hanrahan. Light field rendering. In *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '96, pages 31–42, New York, NY, USA, 1996. ACM. (cited on pages 4 and 5)
9. LightField Forum. Lytro archive. http://lightfield-forum.com/lytro/lytro-archive/, 2019. (cited on pages 13 and 18)
10. MathWorks. Evaluating the accuracy of single camera calibration. https://au.mathworks.com/help/vision/examples/evaluating-the-accuracy-of-single-camera-calibration, 2019. (cited on page 13)
11. R. Ng, M. Levoy, M. BrÃľdif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report*, 2(11):144âĂŞ162, 2005. (cited on pages 5 and 6)
12. S. Nousias, F. Chadebecq, J. Pichat, P. Keane, S. Ourselin, and C. Bergeles. Corner-based geometric calibration of multi-focus plenoptic cameras. In *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017. (cited on pages 4, 5, 8, and 12)
13. S. O'Brien, J. Trumpf, V. Ila, and R. Mahony. Calibrating light-field cameras using plenoptic disc features. In *2018 International Conference on 3D Vision (3DV)*, pages 286–294, Sep. 2018. (cited on pages 4, 5, 7, 8, 11, and 14)
14. S. O'Brien, J. Trumpf, V. Ila, and R. Mahony. Disparity fields: A new method of depth estimation with light field cameras. In *2019 International Conference on Computer Vision (ICCV)*, 2019. **[Preprint]**. (cited on page 40)

15. A. L. Todor Georgiev. The multifocus plenoptic camera. volume 8299, 2012. (cited on page 5)
16. G. Wu, B. Masia, A. Jarabo, Y. Zhang, L. Wang, Q. Dai, T. Chai, and Y. Liu. Light field image processing: An overview. *IEEE Journal of Selected Topics in Signal Processing*, 11(7):926–954, Oct 2017. (cited on page 5)
17. S. Zhu, A. Lai, K. Eaton, P. Jin, and L. Gao. On the fundamental comparison between unfocused and focused light field cameras. *Appl. Opt.*, 57(1):A1–A11, Jan 2018. (cited on page 5)