Thesis Submitted for the Degree of
Doctor of Philosophy

Calibration of Non-Conventional Imaging Systems

Aubrey Keith Dunne, B.Eng., M.Eng.

Supervisor: Professor Paul F. Whelan

Dublin City University
School of Electronic Engineering
June 2009

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

Signed: _____ ID No.: 99279568

Candidate

Date: 19^{th} June 2009

Acknowledgement

Firstly, I wish to thank Prof. Paul Whelan and Dr. John Mallon for the continuous advice, guidance and help that they provided to me over the course of my PhD. Their technical and organisational input was invaluable to me during this time. Thanks also to all the members of the Vision Systems Group, who are always willing to offer help and impart their knowledge whenever asked.

I have thoroughly enjoyed doing my PhD over the past three and a half years, and for this I must thank my fellow research students and friends, Seán, Trish and Brendan. Also, my labmates Julia and Tony, who help to make the lab an enjoyable place to work. Lunch and coffee breaks are always something to look forward to and often provide some unexpected inspiration. For this I thank my friends in DCU - Paddy, Marie, John, Diane, Barry and Andrew. Many other people in the School and in DCU - lecturers, technicians, students and Friday footballers - helped me in a variety of ways and I wish to collectively thank all these people.

My family are always there to support and encourage me during my studies and for this I am very grateful. They helped me through the long days and always celebrated my successes. Thanks Dad, Mum, Cameron and Melanie. Thanks also to Faraday, whose joy at seeing me in the evenings always brightened my day.

Finally, I would like to thank IRCSET, the Irish Research Council for Science, Engineering and Technology, and the Irish State, under the National Development Plan, for generously funding my PhD studies.

Contents

Calibration of Non-Conventional Imaging Systems

Aubrey Keith Dunne

Abstract

This thesis investigates the calibration of cameras that do not conform to the standard pinhole-plus-distortion camera model. The ability to calibrate such camera systems is essential for the development of new, non-conventional, camera types. The major contributions of the thesis are the development of calibration algorithms for both central catadioptric cameras and non-central flexible mirror imaging systems. A key link between the calibrations of pinhole cameras and of general central cameras is expounded, upon which the proposed central calibration scheme is built. The proposed scheme incorporates constraints inherent to the geometry of central cameras, resulting in a method that is shown through rigorous experimentation to have superior performance to the standard generic method. An enabler for this calibration scheme is the use of active calibration grids, using structured light techniques, in place of standard calibration targets. An implementation and thorough characterisation of these active grids is presented. Nevertheless, the extraction of static calibration grid feature points from images with large distortions is still fundamentally important, in particular for experimentation with real images. To this end, a novel algorithm for the detection and consistent ordering of corners in such images of chessboard calibration grids is presented, and its capabilities are demonstrated. The flexible mirror imaging system investigated in this thesis consists of a fixed perspective camera viewing a deformable mirrored surface, which in general results in a non-central imager. A novel calibration technique is proposed that can model the flexibility of these systems, resulting in a calibration that is dynamically updated as the reflective surface flexes. Comprehensive testing for both constrained and unconstrained mirror deflections demonstrates that good calibration results can be obtained using this technique. Results for structure-from-motion and object reconstruction from a sequence of images from a calibrated flexible mirror camera are presented, establishing the capabilities of such calibrated non-conventional imaging systems.

Glossary of Acronyms

Acronym	–	**Explanation**
2D	–	Two Dimensional
3D	–	Three Dimensional
BA	–	Bundle Adjustment
CELECT	–	Corner Extraction via Local Edge Contour Tracing
CGSP	–	Central Generic Synthetic Pinhole calibration
DLT	–	Direct Linear Transform
DMD	–	Digital Micro-mirror Device
FOV	–	Field of View
IAC	–	Image of the Absolute Conic
RANSAC	–	RANdom SAmple Consensus
RMS	–	Root Mean Square
SD	–	Standard Deviation
SLR	–	Single Lens Reflex
$SPFC^2$	–	Scene Point based Flexible mirror Camera Calibration
SVD	–	Singular Value Decomposition
SVP	–	Single View Point
TFT	–	Thin Film Transistor

Chapter 1

Introduction

The human vision system enables humans to instantly perceive depth, colour, texture, and motion at high resolution, and to perform tracking and rapid focal length adjustments to account for relative scene motion. This exceptional versatility has led to the design of cameras and computer vision systems that replicate human vision. In particular, the perspective camera, which is the staple of computer and machine vision, is essentially a simple opto-electronic model of the human eye. However, with the continual increase in processing power and with the ever growing demand for application specific vision solutions, researchers have realised that it is advantageous to develop camera systems that stand apart from the human vision model of perspective cameras. New classes of cameras are emerging, with characteristics beyond those of the human vision system, which can achieve such properties as larger or reconfigurable fields-of-view (FOVs), high dynamic range imaging, and which in general vary the optical domain properties of cameras. This thesis is concerned with some of these emerging camera types, and with methods to apply to them to achieve one of the most fundamental tasks in computer vision — calibration.

In contrast to the many different camera models that have been proposed, the general model for capturing camera projection is applicable to any camera type without restriction. It can cope equally well with perspective cameras that have minimal distortion, and with omnidirectional cameras that have severe image distortion due to FOVs up to 360^o. Two basic methods have recently been proposed for general model calibration, and of these generic calibration is

the more flexible and more easily performed. The benefits of generic calibration for non-perspective central cameras are initially demonstrated in this thesis in order to provide a basis for the further development of the generic calibration method. Like many other camera calibration methods, corners extracted from images of planar chessboard grids are used to generate input for the generic calibration algorithm. This extraction can be problematic for high distortion cameras, since imaged grid corners are severely displaced from their canonical positions in a non-linear manner. A semi-automatic method for the extraction and ordering of grid corners in such images is investigated that circumvents the need for tedious manual corner selection. When performing high fidelity calibrations for the general camera model, the low density of chessboard grid corners can result in interpolation-induced bias. In order to acquire accurate and dense location data, the use of active grids – spatio-temporally varying patterns displayed on a monitor – for calibration is explored, and ultimately benchmarked against localisation for standard calibration grids. By using active grids, performance enhancements can be achieved for the standard generic calibration method by way of major modifications that are subsequently proposed. The central camera, in which all camera rays intersect at a single point, is by far the most common type of camera. However, the generic calibration of this camera type using the standard method is particularly complicated. Simplifications made possible by active grids and by the geometric constraints of central cameras are the inspiration for the improved generic calibration that is investigated in this thesis.

Reconfigurable cameras that view a scene via reflection in a flexible mirrored surface stand apart from most other camera types since they have few constraints and since the calibration changes after each mirror flexion. Thus it is difficult to efficiently apply existing calibration methods to them, and specific methods are required. The sole such existing method is accurate and applicable to video, but it demands an extensive offline calibration stage. Promoting flexible mirror imaging within the community as an advantageous imaging modality would be made easier if a more user-friendly calibration was demonstrated. This goal forms the basis for the final section of this thesis.

In summary, this thesis addresses the key issue of camera calibration within the specific domain of non-conventional cameras, with the key novelty being the development of practical methods and aids for calibration.

1.1 Background and Motivation

This section gives a brief overview of the principal concepts that are dealt with in this thesis, namely camera calibration, and non-conventional imaging. The aim is to highlight the importance of calibration for non-conventional cameras, and to show how such cameras can contribute to progress in computer vision.

1.1.1 Camera Calibration

The process of camera calibration is fundamental to all computer vision tasks involving metric image information. Calibration is necessary to allow faithful measurement and analysis of the abstracted world scene information that is available in a camera image. It describes the relationship between the image space and the world space, or at the most basic level it relates 3-space points to image space pixel locations. In general the closer the camera is to the theoretical pinhole model, the easier it is to perform calibration. As the camera model becomes more complicated, principally due to increased non-linear optical effects, the number of model parameters typically increases, requiring a more involved calibration. As with many engineering problems, linear solutions to calibration are desirable but often not sufficiently accurate. Non-linear refinement of calibration parameters, in the form of bundle adjustment, is a drain on computational and time resources, but is often necessary to improve the accuracy of a linear calibration solution. In summary, the task of calibration is to select an appropriate camera model, and to determine the parameters or values of this model to an accuracy acceptable to the current application with the minimum amount of computation, and in a manner that can be conveniently performed by the practitioner.

1.1.2 Non-Conventional Imaging Systems

Perspective cameras are the most established and most common camera type in use today. This is principally because they have many advantageous properties such as ease of manufacture, simple analytic projection scheme, low distortion and compact form. However, as researchers strive to develop and realise ever more increasing accuracy and efficiency in computer vision and machine vision

tasks, the shortcomings of the perspective camera become a problem for some applications. For example, perspective cameras have a FOV that is fixed, that is relatively narrow, and that is necessarily continuous. In addition, perspective cameras are central, meaning that there is a scale ambiguity when they are applied to certain structure and motion tasks.

In response to these shortcomings, recent years have seen a trend in computer vision towards non-perspective cameras that can overcome some of the perspective camera disadvantages, and that are designed for specific vision tasks. Conventionally the input to the computer vision pipeline has been the camera image, whereas increasingly the input to the process is the scene or object being viewed by the camera. Therefore problems that originally were being solved in the image domain using image processing are now being partly addressed in the optical domain before scene information is incident on the camera sensor, through the use of non-standard camera designs. The camera systems that result from this trend are collectively termed non-conventional cameras. A non-conventional camera can be considered to be a camera in which the optical path between scene and sensor is intentionally modified in a specific way in order to achieve a certain goal or property that cannot be achieved with a perspective camera.

Non-conventional cameras have been successfully implemented in various fields within computer vision, such as mobile robot navigation, security, telepresence, and virtual reality. Examples of non-conventional cameras in mobile robot navigation include their use for visual topological map building (Silpa-Anan and Hartley, 2005), for autonomous unmanned aerial vehicle navigation through urban canyons (Hrabar and Sukhatme, 2004), and for obstacle avoidance by several teams that partook in the DARPA Urban Challenge 2007 for autonomous vehicle navigation[1]. Other interesting recent applications of non-conventional cameras include a visual maneuvering aid for large vehicles (Ehlgen and Pajdla, 2007), omnidirectional endoscopic attachments for medical applications (Sagawa et al., 2008), and low-cost immersive displays for virtual reality (Johnson et al., 2007), where the projector can be seen as a non-conventional camera in reverse. In Kuthirummal and Nayar (2006) a single image multiview method for object frontal 3D structure recovery is made possible by the use of a truncated conical mirror and perspective camera, whilst

[1] http://www.darpa.mil/GRANDCHALLENGE/ (Accessed September 2008)

Nomura et al. (2007) use a flexible camera array to generate dynamic scene collages that enable a new and interesting way to experience large dynamic scenes.

An example of an image from a non-conventional camera is shown in Fig. 1.1(a). This image, which is from a camera with a 360^o horizontal FOV, demonstrates by example that the FOV limitation of the perspective camera can be overcome by applying a new camera design. The richness of scene information captured in this image could never be replicated in a single image by a perspective camera alone. On the other hand, the resolution of the image shown in Fig. 1.1(b), which is a version of the omnidirectional image remapped to improve its perceptibility to the human vision system, varies from the top to the bottom of the image. Clearly, despite their many benefits, non-conventional cameras also have inherent drawbacks when compared to perspective cameras, undesired variable resolution being only one example. One of the most significant drawbacks is the difficulty in calibrating them, since perspective calibration techniques are not normally applicable. This difficulty in calibration is the core problem that is addressed in the thesis. By providing methods for the calibration of non-conventional cameras to enable their wider use in computer vision, the practitioner is afforded the opportunity to select from a broader range of camera types when determining the most appropriate camera setup for any particular application.

1.1.3 Thesis Goals

The work presented in this thesis is motivated by the desire to develop efficient and practical methods to calibrate non-conventional imaging systems so that metric information may be derived from images produced by such systems. Two types of non-conventional camera are considered, firstly wide-angle and omnidirectional cameras, and secondly flexible mirror cameras. Complete practical calibration methods enable non-conventional cameras to be used in a wider range of applications and facilitate the development of novel imaging systems using such modalities.

Accordingly, the goals of the presented research are

- to enable the easier application of chessboard grids to the calibration of

(a)

(b)

Fig. 1.1: Omnidirectional image (a) before, and (b) after cylindrical unwarping.

cameras with severe distortion;

- to demonstrate the suitability of active grids for use in non-conventional camera calibration;
- to improve calibration accuracy and efficiency for central generic calibration;
- to develop a practical calibration method for flexible mirror cameras
- to demonstrate the suitability and applicability of flexible mirror cameras for practical computer vision tasks

1.2 Literature Review

Possibly camera calibration has received more attention in the literature than any other aspect of computer vision. In particular since the advent of the digital camera, researchers have endeavoured to develop accurate calibration methods that are convenient to implement on-site, away from the laboratory. With such a large body of work, the literature survey presented here concentrates on the significant publications immediately related to the stated goals of the thesis. A more detailed examination of camera calibration, and of the progression of calibration methods to keep pace with the emergence of non-conventional cameras, is provided in Chapter 2.

The following review is conducted under headings that correspond to each of the main thesis chapters. In each case, it provides a context for the work in the related chapter, and covers the current solutions to the problem under examination. Where necessary, further references and more details of related work are cited in the relevant chapter.

1.2.1 Corner Extraction and Ordering

Camera calibration methods generally fall into one of two categories: those that require a-priori knowledge of the scene, and those that do not require a-priori scene knowledge. The first category primarily uses geometric primitives such as lines, corners and conics in the image to help with the calibration process, whereas the second category consists of self-calibration methods that make use of the epipolar constraint. This thesis deals with the first category, which encompasses methods for calibrating perspective, fisheye, omnidirectional and non-central cameras (Zhang, 2000, Kannala and Brandt, 2006, Scaramuzza et al., 2006, Sturm and Ramalingam, 2003). Images of planar calibration targets in different orientations are used for calibrating in these methods, enabling calibration to be easily performed by the practitioner. The most popular target pattern for planar calibration is the chessboard pattern, in which the dominant features are corners, while circular patterns, in which the features to be extracted are the circle centroids, are less common. A recent study by Mallon and Whelan (2007*b*) compares chessboard and circular grid patterns in order to determine the accuracy with which feature points in each can be recovered

in images containing distortion. They show that chessboard calibration grid features are invariant to both perspective bias and distortion bias, and so are superior to circular patterns for the generation of bias-free control points.

The practical problem of extracting corners from images of chessboard calibration grids is extremely important in the context of accuracy, as significant noise can be introduced into the calibration by poor feature extraction. As a result, several automatic or semi-automatic methods of corner detection and ordering from images of chessboard grids have been proposed. Extraction of corners in distortion free chessboard images is relatively straightforward, with the main steps necessary being corner detection, false corner removal, and corner ordering. Corner detection can be achieved using a robust corner detector, such as Canny, SUSAN, or Harris, or by performing template matching. False positive corners can be removed by a symmetry test on each candidate corner, and by enforcing a minimum distance between corners. Ordering of corners is based on their geometric proximity. However, the large non-linear distortions typically present in images from non-conventional cameras makes corner extraction and ordering much more difficult.

Recently Wang et al. (2007) presented a detection and ordering scheme that makes use of vanishing points and vanishing lines to detect and order chessboard corners in perspective camera images. Corners are detected with a Harris corner detector and tested for symmetry. The vanishing points for the grid rows and columns are then determined, and corners that do not lie at vanishing line intersections are removed. However, the use of vanishing lines assumes linearity of the grid edges, and thus the method is only suitable for images containing very low distortion.

The method of Mallon and Whelan (2007*a*) is to extract all corners in the image automatically using any standard corner detector, and then to remove outliers by applying a symmetry measure to the candidate corners. Ordering of the points is performed by determining the top left and top right corner points, and then by extracting the remaining corners in the row as the closest N points to the line between these two points, where N is the number of grid columns. Column ordering is based on point distance from the origin corner. Extracted corners are then removed from the unordered set and the process is repeated until all corners are ordered. The method of Lucchese (2005) is very similar in principle, although it requires the user to manually select the

four outer corners of the chessboard grid in the image at the beginning of the process. The ordering of the internal grid corners proceeds iteratively by determining the corner pairs whose join is close to parallel to the join of corner pairs in the same columns but in the previous row, and then selecting the closest resulting corner.

Shu et al. (2003) exploit the topological structure of chessboard grids using Delaunay triangulation to group the grid corner points, detected with a Harris corner detector, into a triangular mesh. Neighbouring triangles with similar colour are merged into quadrilaterals, the corners of which are the grid point corners. This method has the advantage that partial grids can be detected in images. However, it can fail when applied to images with severe distortion, since the compression of the grid squares in these images can result in an incorrect triangulation. Test results for images with large distortion are not presented.

The OpenCV computer vision library[2] includes a *cvFindChessBoardCorner-Guesses()* function, developed by Vladimir Vezhnevets[3], that can be used to automatically detect and order chessboard corner points. The function adaptively thresholds the input image and then successively erodes the image and searches for image contours. All contours with exactly four sides are classified as grid square quadrangles, and their corners are the grid corners. Ordering is performed by pairing each corner based on its proximity to all other detected corners. However, because the function will only operate correctly if all of the grid squares are visible in the image, only complete, and not partial, grids can be extracted. The function was evaluated by Fiala and Shu (2008) with real chessboard images from cameras with varying parameters and it was found to perform poorly. Poor performance of the function for images with high distortion has also been reported by Fiala and Shu (2008) and by Rufli et al. (2008).

Mühlich and Aach (2007) present a technique to detect chessboard grid corners using multi-steerable filters. Candidate corner regions are determined using a Harris corner detector, and then matched filtering is applied to these regions to detect corners with pixel accuracy. Corners are subsequently localised by

[2]http://sourceforge.net/projects/opencvlibrary/ (Accessed December 2006)

[3]http://graphics.cs.msu.ru/en/research/calibration/opencv.html (Accessed November 2008)

parabola fitting. However, the method focuses on detecting corners in perspectively warped images, and results for non-linearly distorted images are limited to a single example. The ordering of the corners is not discussed.

A common camera calibration approach is to use the Camera Calibration Toolbox for MATLAB[4]. The corner extraction stage of this toolbox requires the user to select the location of each of the four outer corners of the grid. An initial estimate of corner locations is made by mapping the metric grid to the four extreme grid corners under a planar homography, which simultaneously fixes the ordering of the corners. When the result of this mapping is not sufficiently accurate, the user must iteratively guess the value of a single radial distortion parameter until a suitable result is obtained. The low-order radial distortion model cannot model complex distortions accurately, and cannot model distortion that is not radial.

Rather than extracting localised feature points from chessboard grids, Fiala and Shu (2008) propose using self-identifying planar patterns for plane-based calibration. The ARTag fiducial marker system is employed, which consists of a library of bi-tonal square markers that each encodes 2D location using a 6×6 binary array at the marker centre. By forming a planar array of the ARTag markers, the location of the markers can be independently determined using decoding software. This has the advantage of being able to extract partial grids for calibration, and since the marker locations are immediately available, there is no need for an ordering stage. Results presented for calibration with the marker system show improved accuracy relative to the OpenCV method, and additionally improved robustness to perspective and to non-linear distortion. However, results for omnidirectional or non-central cameras are not presented - calibration experiments were conducted for cameras with a distortion model containing only two radial and two tangential terms. The localisation of the marker centre, which is performed by intersecting marker quadrilateral diagonals, is also susceptible to bias induced by distortion. The ARTag marker libraries and decoding software are not currently freely available.

The ARTag fiducial marker system is an example of a spatially encoded grid. Various spatial encoding schemes have been presented in the literature, principally in the area of structured light. For this purpose the spatially encoded patterns are typically projected onto a scene in order to solve the correspon-

[4]http://131.215.134.19/bouguetj/calib_doc/index.html (Accessed September 2007)

dence problem in stereo vision and thus enable accurate point triangulation. However, they can be equally employed for the purpose of plane-based camera calibration. Vuylsteke and Oosterlinck (1990) presented a spatially redundent binary pattern for depth estimation based on an underlying chessboard grid structure, where the encoded points exist at the grid corners. The grid pattern is modulated by a dark or light spot at each grid corner, and the location on the grid of each encoded point is determined by the binary colours of these spots in a 2×3 window. Spatial redundancy is achieved by using overlapping patterns. Many other spatial encoding strategies have been presented in the literature, and summaries and comparisons of some of these methods can be found in Mouaddib et al. (1997) and Salvi et al. (2004). However, when used for calibrating non-conventional cameras that induce severe image distortion, spatially encoded grids have the disadvantage that they require image continuity and that the grid image must be locally regular. Regularity is necessary since the location of each grid feature is encoded in a local area around that feature. For large local distortions, it may not be possible to decode the feature locations.

In contrast to the methods described above, the OcamCalib Toolbox for MATLAB (Scaramuzza et al., 2006) is designed specifically for the calibration of omnidirectional cameras, which produce images with severe distortion, using planar chessboard grids. The calibration method does not assume prior knowledge of the mirror surface shape, therefore every corner point of the calibration grid must be selected manually and in a consistent fashion. This directly determines the corner ordering, but in practice it is slow and tedious.

In order to improve the efficiency of calibration with the above toolbox, Rufli et al. (2008) recently presented a fully automatic region based segmentation approach to detecting and ordering chessboard grid corners that is specifically designed for images from catadioptric cameras. This solution to the detection and ordering problem was developed simultaneously but independently of the method that is proposed in this thesis. Input parameters required are the number of corners in each grid row and in each grid column. The algorithm modifies the method implemented in the OpenCV *cvFindChessBoardCornerGuesses()* function by making five adaptations. The most important of these are a new kernel for the erosion of the quadrangle regions to prevent excessive corner rounding, and a new heuristic for determining the linking between quadrangle corners. Test results for the modified algorithm show that it outperforms

the OpenCV method for non-perspective images, and for low-resolution and blurred images. While automatic detection is an advantage for perspective cameras, in omnidirectional images with large FOVs, corners near the image edges are often too small, too distorted, or too blurred to be localised reliably, and typically they are dropped from further consideration. In these cases, some manual input is required to discard these corners.

Approaches to corner detection and ordering based on some prior knowledge of camera deformation have been proposed in the literature. Sturm and Barreto (2008) and Mei and Rives (2007) present methods that require the manual selection of 12 and 7 points, respectively, on the calibration grid in order to fully extract the grid corners. The selected points are used for partial camera calibration in order to obtain a model for the camera distortion. Both methods are parts of calibration frameworks, and they can only be applied to central catadioptric cameras. The disadvantage with using camera deformation related priors is that such an approach is unsuitable for chessboard corner detection and ordering in images from some non-central cameras, since for these cameras deformation types are so varied as to make a-priori modelling of them impractical.

With the exception of the independently developed method of Rufli et al. (2008), none of the above automatic or semi-automatic methods has been shown to correctly and consistently detect and order corners in images of chessboard grids from both central and non-central cameras that contain significant distortion. The level of distortion at which each method fails varies, although generally the distortion must be at least sufficient to cause corners in the second grid row/column to be closer to a line joining the outer corners than corners in the first grid row/column. Thus, at the time that the grid corner detection and ordering problem for distorted images was considered, manual corner selection was the only available option. This is redressed in Chapter 3, which presents a novel semi-automatic method for chessboard corner detection and ordering that is applicable to images from a broad range of non-conventional camera types that contain significant non-linear distortion. The proposed method takes a different approach to solving the problem than that taken by Rufli et al. (2008), although the goals of both methods are similar.

1.2.2 Central Generic Calibration

The standard generic calibration method proposed by Sturm and Ramalingam (2003) and developed further in (Sturm and Ramalingam, 2004, Ramalingam et al., 2005*a*,*b*,*c*) is applicable to any camera geometry. It is used to calibrate the general camera model (Grossberg and Nayar, 2001) based on the constraint that all 3-space points seen by the same camera pixel for a fixed camera position are collinear. The general camera model is described in Chapter 2. The complete generic calibration method for central cameras is detailed in Section 2.4, along with the basis for its development and its benefits in relation to other existing calibration methods. In this section, advancements to the method, that have been proposed in the literature, are reviewed.

Generic calibration was developed such that both central and non-central cameras can be calibrated within the same framework. Consequently, existing parametric calibration techniques for central cameras are not used in the generic calibration framework. Calibration of central cameras is of significant practical use in vision applications, since the common classes of perspective, fisheye, and central catadioptric cameras are all either central or approximately central, and thus central generic calibration warrants further attention. Nevertheless, a survey of existing literature has highlighted a lack of research into central generic calibration. Use of generic calibration for vision tasks such as structure from motion has been proposed (Ramalingam et al., 2006), but only very limited work has been carried out on furthering generic calibration itself. This may be due partly to the short period of time that has elapsed since generic calibration was originally proposed by Sturm and Ramalingam (2003), and it may also be due partly to the fact that a specific algorithm for central generic calibration from planar grids was proposed and verified at that time. However, the method of Sturm and Ramalingam (2003) for central calibration differs principally from their method of non-central calibration in the analytical determination of the calibration values, which are the grid poses and the effective camera centre. Their central method is primarily presented because the non-central equations can not be applied to the central case due to indeterminacy related to ray coincidency at the camera centre. In their generic calibration framework, the geometry of the central problem is considered to be just a specific case of the general camera geometry.

Gonçalves and Araújo (2005) presented a modification to the non-central

generic calibration method of Sturm and Ramalingam (2003) that uses the known perspective camera geometry to reduce the minimum number of input images required for non-central catadioptric calibration from three to two. This was achieved by assuming a-priori knowledge of the intrinsic calibration parameters of the perspective camera viewing the mirror, so that the required third point seen by each pixel is parameterised as a point along a known perspective camera ray. The use of the three-point collinearity constraint still forms the basis of the calibration, as in the original method.

Self-calibration of general cameras under constrained motion has been proposed by Ramalingam et al. (2005*c*). They show that a central camera can be calibrated using images resulting from two pure rotations of the camera and one pure translation. Nistér et al. (2005) present a theoretical analysis of the limits of central general camera self-calibration based on the image flows induced by infinitesimal and finite rotations and translations. They concentrate on the case of three instantaneous flows resulting from infinitesimal rotations observed at a finite number of points in the distorted image, and show that in this case the camera can be calibrated up to a projective ambiguity. Grossmann et al. (2006) further show that Euclidean calibration can be obtained from two dense flows corresponding to pure rotations of the camera around non-collinear axes. A corresponding calibration approach is presented. Espuny and Gil (2008) also present a method of general camera self-calibration that requires optical flow data from only two pure rotations about non-collinear axes. However, these self-calibration methods that are based on camera rotations only are not yet mature, either because of ambiguity in their solutions or because of high sensitivity to noise and to camera type. The self-calibration methods for general cameras do not provide any route for the further development of the standard central generic calibration.

Recently, Ramalingam and Sturm (2008) presented a minimal solution for generic calibration of central cameras using planar calibration targets. By considering the theoretical minimum of four rays intersecting three calibration grids, they derive a calibration solution that is shown to be more robust to noise than the original method. The solution does not introduce any further constraints on the calibration due to centrality, but rather is based on the same collinearity constraint as the original with additional analytical reductions.

Active grids consist of a sequence of structured light patterns displayed on a flat

screen monitor. In Chapter 4 the application of active grids to central generic calibration in order to improve calibration accuracy is examined. Structured light has previously been used in calibration methods as a means of generating a more dense set of planar feature points than standard planar chessboard grids. However, its application as a means of directly achieving calibration has not received significant attention. Although not for complete camera calibration, Sagawa et al. (2005) propose a distortion calibration method for wide-angle lenses in which an active grid is imaged by the camera so that accurate correspondences between active grid locations and camera pixels are determined. A resulting lookup table describing the mapping between distorted and undistorted pixels is formed, which, unlike most distortion calibration methods, is non-parametric. In this way it is similar to the non-parametric generic calibration method. Undistortion is then easily applied by using the determined lookup table, after normalisation and after some linear interpolation to refine the accuracy. Presented results demonstrate the improved distortion correction performance of the method over standard parametric methods.

The above modifications that have been proposed to generic calibration focus either on the non-central case, or on further analysis of the equations of the original central generic method. Central cameras have a geometry that is uniquely defined by the existence of a single effective centre of projection, and as a result they have many properties in common with pinhole cameras. However, no attempt has been made to identify links between pinhole calibration and generic calibration, and so the potentially advantageous properties of the former have not been applied to the latter. Chapter 4 presents a new method for central generic calibration that achieves improved accuracy when compared to the existing central generic method by exploring commonality between pinhole and central generic calibration. Active grids are shown to enable an efficient calibration framework in which complete central generic calibration is achieved, and for which central camera geometry is fully exploited.

1.2.3 Flexible Mirror Imaging

Flexible mirror imaging systems, recently introduced by Kuthirummal and Nayar (2007), are non-conventional catadioptric cameras typically consisting of a perspective camera viewing a scene reflected in a flexible, or bendable, mirror. The characteristics of the camera image can be continuously and non-linearly

altered by directly flexing the mirror. This enables control over both the FOV of the image and the resolution of the imaged scene. By suitable flexing of the mirror, certain portions of the scene being imaged can be attributed more sensor resources than other portions, and the FOV can be altered easily as required. Thus flexible mirror imagers may be beneficial in active vision monitoring and security applications, where higher resolution could be obtained in image regions containing objects of interest without sacrificing FOV, as would normally be the case in pan-tilt-zoom camera systems.

The field of programmable imaging is related to that of flexible mirror imaging, although the methods whereby the images are captured differ significantly. Nayar et al. (2006) presented a programmable imager based on a perspective camera viewing a digital micro-mirror device (DMD) similar to that used in state-of-the-art projector systems. By individually varying the orientations of the micro-mirrors, variable FOVs and image resolutions can be obtained. However, current DMDs have only two possible active orientations, so that programmable imagers cannot produce continuously variable FOVs and thus are severely limited with respect to flexible mirror imagers. Nomura et al. (2007) presented a multi-perspective flexible camera array that they use for forming dynamic scene collages. The imager consists of a set of 20 perspective cameras that are attached to a bendable plastic frame. The authors focus on image matching and stitching for improved viewer interpretation rather than on actually calibrating the camera array, and so there is no crossover with flexible mirror imager calibration.

The introduction of flexible mirror imaging systems naturally leads to the requirement for methods for their calibration, so that tasks such as tracking in surveillance can be accomplished. Calibration of such systems is difficult firstly because the mirror deflection is generally unconstrained, and secondly because the camera configuration, and thus calibration, alters each time that the mirror is flexed. The calibration is intrinsically linked to the estimation of the flexible mirror surface itself, since knowledge of local mirror surface shape allows surface normals and thus reflected rays to be determined for given incident rays. However, since the mirror shape can change, catadioptric calibration methods that assume some prior knowledge of mirror shape (Tardif et al., 2006, Thirthala and Pollefeys, 2005*a*, Scaramuzza et al., 2006) are not applicable, and therefore methods that can estimate the mirror shape are required.

Much work has been presented in the literature on the recovery of surface shape from images of a diffuse surface, using either structured light or epipolar geometry. These methods are often grouped into the field of shape-from-X, which has received significant attention outside the field of calibration. The recovery of specularly reflective surface shape has received relatively less attention from the vision community. This situation differs from the diffuse case since 'features' seen in the image are virtual features, caused by reflection, that do not obey the epipolar constraint. Halstead et al. (1996) presented one of the first techniques for accurately determining specular surface shape from surface reflection in a single image. They used a bespoke conical calibration object and concentric camera to determine corneal surface estimates based on a set of estimated surface normals. Savarese et al. (2005) describe a method for specular surface recovery from a single image of a planar calibration target when at least two local orientations are available at each target point.

Several approaches to estimating mirror shape from motion have also been presented. Swaminathan et al. (2002) examine the dependence on surface geometry of specularities in static scenes with constant velocity camera motion, whilst Roth and Black (2006) estimate the specular surface geometry from specular flow and then use the result to improve surface estimation from diffuse flow. Oren and Nayar (1997) recover information on surface profile from specular reflections whose paths overlap under constrained camera motions. However, all of these methods assume a fixed mirror shape with relative mirror scene motion, and so are not suitable for flexible mirror imaging systems calibration. Where relative camera scene motion is used, the motion is assumed rigid, whereas for flexible mirror systems the mirror motion due to flexing is non-rigid.

Calibration methods for the general camera model (Grossberg and Nayar, 2001, Sturm and Ramalingam, 2004) can be used to determine reflected ray directions and thus calibrate any catadioptric camera, but these methods are time consuming, and when applied to a flexible mirror imager, they would require the entire calibration process to be repeated each time the mirror is flexed. As mentioned in the previous section, Gonçalves and Araújo (2005) simplify the method of Sturm and Ramalingam (2004) in the case of known perspective camera calibration. Points on the estimated mirror surface are available from their solution, but two images for each mirror position are required, and information on the previous positions of the mirror is not used to inform the

calibration. Gonçalves and Araújo (2007) also present a method for calibrating a catadioptric camera from a single image of a calibration target, but it is only applicable when the mirror surface can be described by a quadric. A completely non-parametric method for specular surface estimation based on voxel carving has been presented by Bonfire and Sturm (2003), in which normal vectors are accumulated for each scene voxel, and the voxels with the normals in best agreement are considered to be on the specular surface. The method requires images of the reflective surface from many different viewpoints to achieve good results. Bonfire et al. (2006) present a method for specular surface estimation that is an extension of the generic calibration method. Using images of a planar grid, located in at least two different positions, reflected in the specular surface, the surface is recovered by triangulation. All the above methods assume a fixed specular surface geometry. If they were to be applied to calibrate a flexible imaging system, the system would have to be recalibrated completely after each mirror flexion, which is not practical in the envisaged system.

The field of adaptive optics is concerned with the development of flexible mirror arrays primarily for use in astronomical observation applications. By applying small deformations in a controlled way across the surface of the mirror in an astronomical telescope, wavefront aberrations in the incident light can be corrected for optically. The aberrations are the result of atmospheric disturbances that distort the wavefront as it passes through the earth's atmosphere – by removing this distortion, more faithful astronomical observations are achieved. Recently, Papavasiliou and Olivier (2006) have used nanolaminate reflective materials in conjunction with micro-electro-mechanical electrostatic actuators in order to create deformable mirrors of various sizes for use in astronomical imaging. However, the magnitudes of the deformations applied in adoptive optics systems are of the order of $10\mu m$, and so these systems do not provide the scope to allow variable camera FOV, and are thus not suitable in their current form for use in flexible mirror imagers in computer vision. Due to the small deformations and to the direct control of the deformations using the actuators, the calibration methods for adoptive optics systems do not have significant crossover with flexible mirror imager calibration.

Kuthirummal and Nayar (2007), who introduced the concept of flexible mirror imaging with a nominally planar flexible mirror surface, presented the only existing calibration method specifically for such systems. Their technique is

dependent on a unique mirror boundary to mirror surface mapping. Calibration is performed offline by acquiring an image of the mirror boundary, and an accompanying mirror surface shape measurement, for each mirror deformation that is likely to be imposed. The one-to-one mapping between the mirror surface shape and a descriptor of its boundary is then stored in a look-up table, so that the calibration is immediately available when the mirror outline is visible in the image. They present results for a calibration with greater than $30,000$ descriptor-surface pairs. Although this method has demonstrated good accuracy, it requires a significant offline calibration stage that necessitates equipment for directly determining accurate surface shape information.

The calibration methods reviewed above, excepting that of Kuthirummal and Nayar (2007), are generally either unsatisfactory or inapplicable for direct application to flexible mirror camera calibration. They assume a fixed mirror surface and thus necessitate recalibration after each mirror flexion, which becomes impractical for multiple successive mirror shape changes. On the other hand, the single existing method of flexible mirror camera calibration is designed to calibrate a continuously altering mirror surface, but requires an extensive pre-calibration stage. If flexible mirror cameras are to grow as an imaging modality for computer vision, then their calibration should be achieved easily by using standard calibration tools such as those typically used for perspective camera calibration. Chapter 5 presents a calibration approach designed specifically for flexible mirror cameras that is motivated by this desire for accessible and practical calibration.

1.3 Mathematical Notation

The real and projective spaces are represented by $\mathbb{R}^n$ and $\mathbb{P}^n$, respectively, where n is the dimension of the space. General sets are denoted by $\mathbb{A} = \{a_1, a_2, a_3, \ldots a_n\}$ where $a_1, a_2, a_3, \ldots a_n$ are the set elements.

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

transformation. Rotations, R, and homographies, H, are represented similarly.

Vectors and points in $\mathbb{P}^2$ are denoted in bold lower case as $\mathbf{a} = [a_1\ a_2\ a_3 \ldots a_n]^T$ with elements $a_1, a_2, a_3, \ldots a_n$. Points in $\mathbb{P}^3$ are represented in bold upper case as $\mathbf{A} = [A_1\ A_2\ A_3 \ldots A_n]^T$ with elements $A_1, A_2, A_3, \ldots A_n$. Point $\mathbf{A}_{[i\ j\ \ldots\ k]}$ is a new point formed by selecting elements $i, j, \ldots, k$ from point $\mathbf{A}$. The inhomogeneous equivalent of point $\mathbf{A}$ is denoted as $\tilde{\mathbf{A}}$. Line A in $\mathbb{P}^2$ has orientation $\angle A$ with respect to the positive x-axis.

Estimates are denoted with a caret as $\hat{A}$, and updates are denoted with a prime as A'. Equality up to a non-zero multiple is denoted with the $\simeq$ symbol. Finally, an image point $\mathbf{p}$ after distortion is represented by $\breve{\mathbf{p}}$.

1.4 Contributions

The major contributions of this thesis correspond to the main thesis chapters. Lesser contributions, which enable the achievement of the major contributions but have a lower significance, are described in the minor contributions section. Together, these contributions encapsulate the most important work of the thesis.

Major Contributions

The major contributions are reflected in the publications resulting from the work in this thesis, which are listed in full in Section 6.2. Each of the three major contributions is summarised below.

- An edge-tracing based semi-automatic method for detecting and ordering chessboard grid corners that is applicable to images containing significant non-linear distortion is presented. The method is termed CELECT - Corner Extraction via Local Edge Contour Tracing. An a-priori model is not used, so the CELECT method has general applicability to any image containing continuous distortion. It is shown to extract corners successfully from distorted images for which, until recently, only manual methods operated successfully. It is a practical aide to planar chessboard camera calibration techniques for non-conventional cameras.

- A significantly improved complete method for the generic calibration of central cameras is proposed. Benefits are achieved through the use of a synthetic pinhole image plane in the linear estimation stage that provides a key link between pinhole calibration and non-pinhole calibration of central cameras, which had not previously been identified. The method is termed CGSP - Central Generic Synthetic Pinhole calibration. This method is less complicated and more accurate than the existing central generic calibration method.

- A novel calibration framework for the calibration of flexible mirror imaging systems is presented. Based on this, a calibration method that operates on scene points is proposed, which is termed SPFC^2 - Scene Point based Flexible mirror Camera Calibration. The SPFC^2 method provides several benefits over the only existing method specifically for calibrating flexible mirror imagers. Changes in the mirror shape are estimated through a dynamic calibration stage, so that the calibration is updated after mirror deflection rather than being completely recalibrated.

Minor Contributions

Minor contributions are subsidiary considerations that were key to the achievement of the main thesis objectives, but that do not stand alone as significant contributions.

- A study and comparison of the relative performance and applicability of standard central generic calibration versus standard perspective camera calibration is conducted. The results demonstrate the invariance of generic calibration of the general camera model to increasing non-linear camera distortion. For mid to high distortion the standard generic method is clearly shown to be more effective than the pinhole-plus-distortion calibration.

- A performance characterisation of active grids with respect to static grid corner detection methods for generating accurate calibration input data is conducted. Despite the frequent use of active grids in calibration, the relative performance had not previously been examined in the literature.

- The application of a linear pose estimation technique to central general

cameras for improved generic calibration efficiency is expounded. The linear technique provides speed and accuracy improvements over the pose estimation method of standard generic calibration, and is more amenable to inclusion in a RANSAC framework.

- Structure and motion estimation and object reconstruction for a flexible mirror camera are implemented and evaluated through experimentation, demonstrating how existing multiview techniques can be applied to flexible mirror imaging systems.
- Modifications to the SPFC^2 calibration method are made to investigate the possibility of flexible mirror camera self-calibration. Experimental results towards achieving this are presented and discussed.

Key elements of code relating to the most important contributions will be made freely available on the Vision Systems Group code repository[5].

1.5 Thesis Organisation

Chapter 2 examines existing camera models and their treatments of camera distortion for pinhole, perspective, wide-angle, omnidirectional, and non-central cameras. The general camera model is described in detail and is contrasted with the parametric camera distortion models. The chapter concludes with a comparative study of central generic calibration versus standard perspective camera calibration for the calibration of simulated and real wide-angle cameras. The accuracy and robustness to increasing non-linear distortion of the calibration methods are examined.

Chapter 3 firstly addresses the extraction and ordering of chessboard grid corners from images for subsequent camera calibration. CELECT, a novel semi-automatic method of corner detection and ordering that can extract chessboard grid corners from images with severe non-linear distortion, is presented and evaluated. This chapter secondly presents a performance characterisation of active grids for use in camera calibration. The accuracy and robustness of active grids is benchmarked against two existing methods of corner localisation typically applied to chessboard grid images during calibration.

[5]http://www.vsg.dcu.ie/code.html

Chapter 4 deals with the application of generic calibration to central cameras. Major improvements to the existing method are proposed, resulting in the CGSP method, that are shown through both simulated and real experiments to enhance significantly the accuracy and robustness of central generic calibration. A new linear estimation stage utilising active grids and an alternative pose estimation method enable the improvements. A side by side comparison with the standard generic method is conducted with respect to a ray-point error metric, to calibration parameters, to camera centrality and to distortion correction for both a real hyperboloidal catadioptric camera and a real fisheye camera.

Chapter 5 investigates the calibration of flexible mirror imaging systems. The SPFC^2 method, a novel two-stage calibration method that acts to update the current calibration, is presented. The method requires only two images of a chessboard calibration grid for calibration after each deflection. Extensive evaluation is performed on the SPFC^2 method for distortion correction, structure-from-motion, and object reconstruction tasks. The chapter concludes with an examination of possible extensions of the method towards self-calibration.

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

Chapter 2

General Camera Model and Generic Calibration

Camera calibration is possibly the most fundamental task in computer vision. It is the camera specific process of determining how the 3-space world relates to the 2D image space of the camera. The requirement of deriving any type of metric information from an image necessitates prior camera calibration. Although the phrase camera calibration is universally used, the more complete phrase is camera model calibration. This reveals the fact that it is a model of the camera that is actually being calibrated, and highlights the important role that models play in the calibration process. By camera model is meant the mathematical structure that describes the projection operation of a camera. Calibration can be considered to be the determination of the numerical quantities that complete the camera model.

This chapter begins by introducing some of the varied camera models that have been developed by the computer vision community, beginning with the pinhole model in Section 2.1. The pinhole camera model is fundamental to many camera calibration schemes, as it provides a convenient projection scheme that is linear (up to scale) in first order homogeneous coordinates. However, pinhole cameras are poor standalone models for cameras that contain distortion. The non-conventional cameras that are the focus of this thesis all contain significant non-linear distortion, and so camera models that move beyond the pinhole model are considered. Some such models advance the pinhole model by augmenting it with distortion terms to capture the non-linear camera distortion,

whilst for more severe distortion entirely new models are necessary. Section 2.2 examines the camera models for practical perspective and non-conventional cameras by way of their associated distortions.

It is seen that many of the camera models are tied to specific camera types, and that the number of parameters varies between different models and different cameras. A unifying camera model for all camera types would make camera specific model selection redundant and allow a consistent approach to the calibration of any camera. Such a model is the general camera model that was proposed by Grossberg and Nayar (2001). This non-parametric model is described in detail in Section 2.3. General camera modelling is eminently suitable for application to the non-conventional imaging systems in this thesis, and so the task of calibrating the general camera model must be investigated.

Section 2.3 describes the most basic general model calibration method, which is conceptually simple but requires a very precise experimental setup in practice. An alternative and less constrained calibration method, termed generic camera calibration, has more recently been proposed for the calibration of general cameras. The method, its component steps, and the implementation of the key steps for the case of central cameras are described in detail in Section 2.4. Having detailed the generic calibration method, a question arises as to the accuracy and performance benefits of the generic calibration method and general model over more established pinhole-plus-distortion techniques. Section 2.5 attempts to answer this question by providing a comprehensive evaluation of central generic calibration with respect to a standard planar calibration method. Experiments are conducted on simulated and real data, with the firm conclusion that the generic calibration method has the capability to outperform the standard parametric approach for imaging systems with significant distortion. The results demonstrate the benefits of generic calibration for non-conventional central cameras.

2.1 Pinhole Camera Model

The most basic camera model is the pinhole model. It assumes that all light rays captured by the camera are concurrent at a single point, called the camera centre, or equivalently that the camera aperture is infinitely small. The cam-

era obscura is the first recorded example of a pinhole camera. The concept of the camera obscura was explored as early as 500BC, and it gained popularity with European artists during the Renaissance in the 16th century. It was constructed by placing a small hole in a sheet of opaque material, allowing light rays reflected from the scene to pass through the pinhole and form an inverted image on a plane on the opposite side of the sheet. Artists then used the image to produce accurate perspective drawings of the scene. A true pinhole camera has no distortion but is not possible to realise in practice due to the infinitely small aperture requirement. However, since many perspective cameras with low distortion can be approximately modelled as pinhole cameras, it is beneficial to examine the pinhole model in more detail.

Pinhole camera properties that are independent of the camera position and orientation are termed intrinsic parameters, and are captured by the camera calibration matrix, K

$$K = \begin{bmatrix} \alpha f & s & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2.1}$$

where f is the focal length of the camera, p_x and p_y are the principal point offsets that describe the location of the intersection of the principal axis with the image plane in image coordinates, s is the camera skew, which accounts for skewness in the camera sensor elements, and α is the aspect ratio of the pixel height to the pixel width of the camera pixels. If knowledge of the pinhole camera pose is available then 3-space scene points, $\mathbf{X}_i$, can be mapped to image coordinates, $\mathbf{x}_i$, through the camera projection matrix, P, as

$$\mathbf{x}_i \simeq P\mathbf{X}_i \tag{2.2}$$

where

$$P = KR\left[I_{3\times 3}| - \tilde{\mathbf{C}}\right], \tag{2.3}$$

R is the rotation of the camera in the world coordinate frame, and $\tilde{\mathbf{C}}$ is the camera centre location in world coordinates. R and $\tilde{\mathbf{C}}$ are the extrinsic camera parameters. A pinhole camera is completely calibrated when all the intrinsic and extrinsic camera parameters are known.

2.2 Camera Distortions

Images from all practical cameras contain some image distortion due to the camera's physical construction. Image distortion is geometric distortion that displaces pixels from their correct image position but that does not alter the pixels in any other way. It is solely due to the properties of the camera used to capture the image. Random aberrations in the camera image sensor, such as those caused by thermal noise and quantisation, are considered to be image noise and not image distortions. Furthermore, in the examination of image distortion in this thesis, distortion due to chromatic aberration, and defocus blur due to mirror curvature in catadioptric systems, are not considered.

The distortion introduced into images from dioptric cameras, which are cameras that incorporate refractive-only optical elements, will be referred to as lens distortion. Due to this distortion such cameras do not exactly fit the pinhole camera model. Catadioptric camera systems, which incorporate both reflective and refractive optical elements, can introduce both lens and mirror distortion into the image. Camera calibration must capture these distortions if it is to faithfully describe the mapping between 3-space points and image coordinates. It is therefore beneficial to briefly examine the image distortions resulting from practical cameras, and in particular non-conventional cameras. In the area of photogrammetry, where the focus is on high precision calibration, high order models are employed in order to describe accurately the distortions of practical cameras. However, the level of accuracy required in most computer vision tasks is generally not as demanding, leading to the use of approximate models that are easier and faster to incorporate into calibration and multi-view frameworks. The wide variety of camera types finding practical application has led to a diverse range of camera distortion models.

2.2.1 Perspective Camera Distortion

Many perspective camera calibration methods model the camera with a pinhole model augmented by distortion terms that describe the non-linear lens distortion. The two principal lens distortions present in most perspective cameras are radial lens distortion, caused by camera lens optical non-linearities, and decentering lens distortion caused by misalignments between the camera lenses.

However, in most cameras the radial distortion is of significantly larger magnitude than the tangential distortion (Wei and Ma, 1994, Tsai, 1987), leading many perspective camera models to drop the tangential distortion terms.

Radial distortion is typically modelled using an image displacement approach, whereby the displacement of an image pixel from its distortion-free location is a non-linear function of the pixel's distance from the distortion centre of the image. Possibly the simplest such model is the one-parameter divisional model of Fitzgibbon (2001). As it is designed for the simultaneous estimation of radial distortion and multi-view geometry, and in particular for the task of matching across image frames in a sequence, it does not achieve high accuracy. The dominant forward model for radial distortion, derived from the wave aberration equation (Mallon, 2005, Mallon and Whelan, 2007*a*), is

$$\breve{\mathbf{p}} = \mathbf{p} + \mathbf{p}(k_1 r^2 + k_2 r^4 + \ldots) \tag{2.4}$$

where $\mathbf{p}$ and $\breve{\mathbf{p}}$ are the distortion-free and distorted lens-centric image coordinates, respectively, r is the Euclidean distance of $\mathbf{p}$ from the distortion centre, and k_i are the radial distortion coefficients. The polynomial order of the distortion model is chosen in accordance with the desired accuracy. However, for perspective cameras, the accuracy of the model is dominated by the first coefficient. The popular method of Zhang (1998) employs a two parameter radial distortion model of the form of Eqn. 2.4. Heikkila and Silven (1997) present a camera model that also utilises this distortion model, but it is augmented with a two parameter tangential distortion model. Hartley and Saxena (1997) capture radial distortion by a cubic rational polynomial model with 80 parameters.

2.2.2 Wide-Angle and Fisheye Distortion

Wide-angle cameras are dioptric cameras that are designed to have increased FOV over perspective cameras. Wide-angle cameras that have very large FOVs, possibly exceeding 180^o, are generally termed fisheye cameras. Whereas for perspective cameras the lens distortion present is not designed in and is undesirable, for wide-angle and fisheye cameras the lens distortion is carefully designed into the camera as a means to increase the image FOV. The distortion is once again predominantly radial, is non-linear, and is typically too severe to

allow the application of the low order distortion models discussed in Section 2.2.1.

Two approaches to dealing with lens distortion in wide-angle and fisheye camera models appear in the literature. The first approach treats the distortion as non-linear deviation from the pinhole model, and thus is an image displacement approach. In the second approach, the pinhole model is discarded in favour of a specific fisheye projection model, so that there is a transformation directly from incident rays to fisheye image points. The more severe the distortion, the more accurate is the latter approach. The image displacement approach will fail for fisheye cameras with FOVs greater than or equal to 180^o, since in these cases there is no single plane with which all the camera rays can intersect.

The image displacement approaches to distortion simply extend the models of Section 2.2.1 by increasing the number of parameters used. The divisional model (Fitzgibbon, 2001), with between 3 and 8 parameters, is used by Thirthala and Pollefeys (2005*b*) for modelling radially symmetric wide-angle cameras. Shah and Aggarwal (1994) separately model radial and tangential distortion for fisheye cameras using $5th$ order polynomials, resulting in a 10 parameter model. Claus and Fitzgibbon (2005) presented a rational function distortion model for fisheye cameras, requiring 18 parameters, that modifies the pinhole projection model of Eqn. 2.2 to operate on the lifted 6-dimensional space inhabited by the zeroth, first, and second order image points.

Specific fisheye camera models are required for the approaches to modelling lens distortion that are not based on the pinhole model. Ideal fisheye cameras are radially symmetric, and so they have a camera model of the form

$$\mathbf{p} = r(\theta) \begin{bmatrix} \cos\varphi \\ \sin\varphi \end{bmatrix} \tag{2.5}$$

where θ is the angle between the incoming ray and the camera principal axis, $r(\theta)$ is the projection function, and φ is the angle of the incoming ray about the principal axis. The projection function effectively models the radial distortion. Three common fisheye projection functions (Kannala and Brandt, 2006) are

$$Equidistance: \quad r(\theta) = f\theta \tag{2.6}$$

$$Equisolid\ angle: \quad r(\theta) = 2f \sin\frac{\theta}{2} \tag{2.7}$$

$$Stereographic: \quad r(\theta) = 2f \tan\frac{\theta}{2} \tag{2.8}$$

These functions are convenient for modelling fisheye distortion in experiments with synthetic data as they enable the extent of fisheye distortion to be easily related to the FOV. Precise radial undistortion with these models is also easily accomplished. However, in practice the distortion that exists in fisheye images requires a more complicated model to capture the non-linearities in camera and lens assemblies. The single parameter FOV model of Devernay and Faugeras (2001), which is based on the equidistance model of Eqn. 2.6, is an exception, although it still requires application of the radial distortion model of Eqn. 2.4 if significant distortion is present. In order to achieve improved accuracy while still using a low order model, Micusik and Pajdla (2006) present two fisheye camera models that augment Eqn. 2.6 with an additional parameter. Their models are determined for specific fisheye lenses as a compromise between high accuracy and a small number of parameters. Kannala and Brandt (2006) presented the following fisheye distortion model

$$\breve{\mathbf{p}} = \mathbf{p} + \Delta_r(\theta, \varphi) + \Delta_t(\theta, \varphi) \tag{2.9}$$

It extends the equidistance projection function as an odd-order polynomial function of θ, and includes an explicit radial distortion term, Δ_r, and a tangential distortion term, Δ_t. The complete model contains between 6 and 23 parameters, depending on the accuracy required and the amount of distortion present.

2.2.3 Catadioptric Camera Distortion

The dominant type of catadioptric camera is that consisting of a single perspective camera imaging a scene reflected in one of several specific continuous quadric mirror surfaces. Such cameras have FOVs that are greater than a hemisphere and so are often called omnidirectional cameras. Depending on the location of the camera relative to the mirror, catadioptric cameras are either central or non-central (Baker and Navar, 1998) (see Section 2.3.1 for a discussion on camera centrality).

The extent of the non-linear distortion in omnidirectional cameras is greater than that in fisheye cameras in order to achieve the 360^o FOV. As a consequence, many models for specific catadioptric configurations of omnidirectional cameras have been proposed, with the general trend in the literature being to model the camera directly using the mirror surface geometry. Svoboda and

Pajdla (2002) present such a modelling approach, in which a separate quadric representation is used to model each of the three possible non-trivial central catadioptric camera types (elliptical, paraboloidal, hyperboloidal). Geyer and Daniilidis (1999) propose a similar set of camera models for catadioptric cameras that are based on the mirror shape employed. In the case of a paracatadioptric camera (Geyer and Daniilidis, 2002), the mirror focal length and the mirror centre are the shape specific model parameters. Associating catadioptric cameras with mirror geometry implies knowledge of the geometry, and assumes that the mirror surface does not deviate or aberrate from this geometry.

Other researchers have taken a more general approach to omnidirectional distortion, by using a high order model to approximate any of the omnidirectional cameras. Tardif et al. (2006) propose a camera model based on varying focal length that is applicable to any camera whose camera rays are radially symmetric. Distortion is captured by a distortion function that directly maps image coordinates to camera rays. They use a full-order polynomial divisional model for the distortion function, where a 6 parameter model was employed in the presented experiments. The omnidirectional camera model of Scaramuzza et al. (2006) is possibly the most commonly used omnidirectional model, principally due to its inclusion in a freely available omnidirectional camera calibration toolbox[1]. Due to the generality of the model, it can be applied to both fisheye and central omnidirectional cameras, with the number of model parameters incremented until the magnitude of the resulting reprojection error is acceptable. The distortion model is integrated into the projection model by using a general polynomial, so that a camera sensor point, $\breve{\mathbf{p}}'$, is mapped directly to its associated camera ray, $\mathbf{r}$, through

$$\mathbf{r} = g\left(\breve{\mathbf{p}}'\right) \simeq \begin{bmatrix} \breve{\mathbf{p}}' \\ a_0 + a_2\|\breve{\mathbf{p}}'\|^2 + a_3\|\breve{\mathbf{p}}'\|^3 + \ldots + a_n\|\breve{\mathbf{p}}'\|^n \end{bmatrix} \tag{2.10}$$

where $g(\breve{\mathbf{p}}')$ is the projection function and $\{a_0, a_2, a_3, \ldots, a_n\}$ are the model parameters. Sensor points $\breve{\mathbf{p}}'$ are related to image points $\breve{\mathbf{p}}$ through an affine transformation.

Geyer and Daniilidis (2000) develop a camera model based on the observation that any central catadioptric camera projection can be represented by a central

[1] http://asl.epfl.ch/∼scaramuz/OCamCalibration/Scaramuzza_OCamCalib.zip (Accessed September 2007)

projection onto a sphere, with the sphere centre as the centre of projection, followed by a perspective projection onto a plane. The model parameters depend on the mirror shape. Barreto (2006) extends this model to all central cameras by incorporating single parameter radial distortion via the divisional model into the central spherical projection stage. Ying and Hu (2004) also extend the model of (Geyer and Daniilidis, 2000) to include fisheye cameras. Mei and Rives (2007) augment the same model with radial and tangential terms, containing 3 and 2 parameters respectively, in order to capture the distortion inherent to the perspective projection in the model.

The catadioptric camera distortion models discussed up to this point are not directly applicable to non-central catadioptric cameras. Non-central catadioptric cameras can be discontinuous, such as a catadioptric stereo system consisting of a perspective camera viewing multiple planar mirrors with skewed normals, or they can be continuous such as a perspective camera viewing a smooth non-planar mirrored surface, where, if the surface is quadric, the camera is not placed at a focus of the quadric. A flexible mirror imager is an example of a continuous non-central catadioptric camera for which the types of distortion are essentially unconstrained. Non-central camera models for spherical, paraboloidal and hyperboloidal cameras have been presented by Micusik and Pajdla (2004) for the case of inexact alignment between the perspective camera and the mirror. Gonçalves and Araújo (2007) present a method for calibration of fully non-central camera configurations involving quadric mirrors that is based on a 10 parameter model, where the model captures the mirror shape in addition to the relative pose of the mirror and perspective camera viewing it. However, parametric models for non-central catadioptric cameras whose mirror geometries are non-quadric have not been presented in the literature.

2.2.4 Exotic Cameras and Distortions

Many camera types with exotic projection properties and that are difficult to collectively categorise have been proposed and analysed in recent years. Such camera types stand between central cameras with parametric models and completely general cameras that do not have parametric models, and they induce distortions that are typically very specific to the camera type.

Crossed-slits, or X-slits, camera models describe cameras that consist of two

non-concurrent planes, each containing a slit, placed arbitrarily in front of the camera image plane. Perspective projection and linear pushbroom projection are both special cases of the X-slits projection. The projection of every 3-space point for an X-slits camera is defined by the line that passes through the point and intersects both slits. The distortion in X-slits images is depth dependent, and due to occlusions there is no way to remove it even if the scene geometry is known. Zomet et al. (2003) studied the X-slits camera model and presented methods for synthesising X-slits image mosaics from a series of perspective camera images with ego-motion (image-based rendering). Feldman et al. (2003) have developed the epipolar geometry of the X-slits projection model, and present the analogue of the fundamental matrix for X-slits cameras. Pless (2003) examined the epipolar geometry of general cameras consisting of a network of arbitrarily located perspective cameras. Sturm (2005) analysed the multi-view geometry of axial cameras, which are cameras in which all rays touch a single line called the camera axis. Menem and Pajdla (2004) combine the pinhole camera model with a dedicated circular panorama model in order to develop the two-view relations between perspective and circular panoramic images. Each of these cameras have unique distortions that are camera specific, and so distortion removal must be considered through an examination of the associated camera projection scheme.

2.3 General Camera Model

The previous section outlines the type of distortion present in each category of camera. The variety of camera distortion types that must be considered for camera calibration is shown to have led to the development of a large selection of parametric camera models that are tailored to specific camera and projection types. As the camera distortion increases, the parametric models become more complex and the associated calibration methods are increasingly specific. Naturally it would be beneficial to have a camera model that is cross-camera compatible. The general camera model is such a model.

Grossberg and Nayar (2001) promoted the general camera model as a model that can be applied to any camera, regardless of its inherent distortion. The concept of the general model is extremely simple - each image pixel is mapped to its corresponding 3-space camera ray via a lookup table, as shown in Fig. 2.1.

In principle, the ray direction for each pixel is completely independent of the ray directions of the surrounding pixels. General model calibration is then the task of determining this mapping in order to generate the lookup table. Whilst the augmented pinhole models require increasing numbers of parameters for increasing accuracy, the general model is completely non-parametric since it consists only of the lookup table. Thus it is rendered generally applicable, and is even capable of catering for discontinuous cameras. Any camera, regardless of its dioptric or catadioptric configuration, and independently of its centrality, can be described completely by the exact same general camera model. Due to the non-parametric formulation of the general model, it can capture camera distortions caused by mirror surface imperfections or lens defects, a capability that is not available with other methods. Grossberg and Nayar (2001) proposed a raxel, or ray pixel, to describe the virtual photosensitive element that measures the light energy of a compact bundle of rays represented by the single incoming ray that is associated with a pixel in the general model. Each raxel has a set of geometric, radiometric, and optical properties. However, in this thesis only geometric properties are considered, and thus the term camera rays rather than raxels is utilised. When only considering the geometric camera properties, the general model is similar to the model used by Champertous et al. (1992).

2.3.1 Central and Non-Central Cameras

When moving away from the pinhole and pinhole-plus-distortion models to discuss the general camera model, the concept of camera centrality must be considered. A central camera has all camera rays concurrent at a single effective point (Baker and Navar, 1998) called the camera centre or the optical centre. Each camera ray of a central camera is therefore determined by its normalised direction in conjunction with the single camera centre. Such cameras are sometimes termed single viewpoint (SVP) cameras (Tardif and Sturm, 2005). For non-central cameras, there is no limitation on either the concurrency or the intersection of camera rays. Therefore to describe a camera ray from a non-central camera requires both the ray direction and a point on that ray, termed the anchor point, as shown in Fig. 2.1. Every anchor point is distinct in the most general non-central camera.

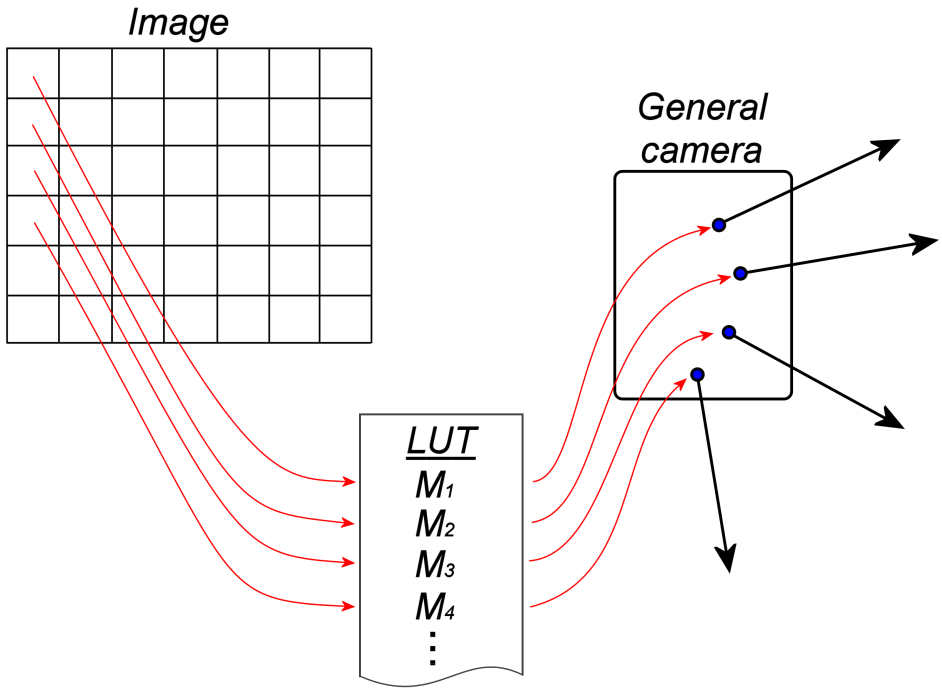


Fig. 2.1: The general camera model. Image pixels map to rays in 3-space via a lookup table of Plücker matrices, M_i.

2.3.2 Plücker Line Representations

Rays in 3-space are conveniently described homogeneously by Plücker representations. For any two non-concurrent homogeneous points, $\mathbf{A}$ and $\mathbf{B}$, the Plücker matrix representation, M, of the line passing through both points is a 4×4 matrix that is calculated as

$$M \simeq \mathbf{A}\mathbf{B}^T - \mathbf{B}\mathbf{A}^T \tag{2.11}$$

This representation has the properties of being invariant to the points chosen on the line, and to being skew-symmetric. Plücker matrices are used to describe viewing rays in the general camera model. It will be seen in Chapters 4 and 5 that an alternative Plücker representation of lines is more convenient for structure-from-motion tasks with general cameras. Plücker coordinate vectors, L, are 6×1 vectors that can be derived directly from their corresponding

Plücker matrices as

$$N = \begin{bmatrix} M_{41} - M_{14} \\ M_{42} - M_{24} \\ M_{43} - M_{34} \\ M_{32} - M_{23} \\ M_{13} - M_{31} \\ M_{21} - M_{12} \end{bmatrix} \tag{2.12}$$

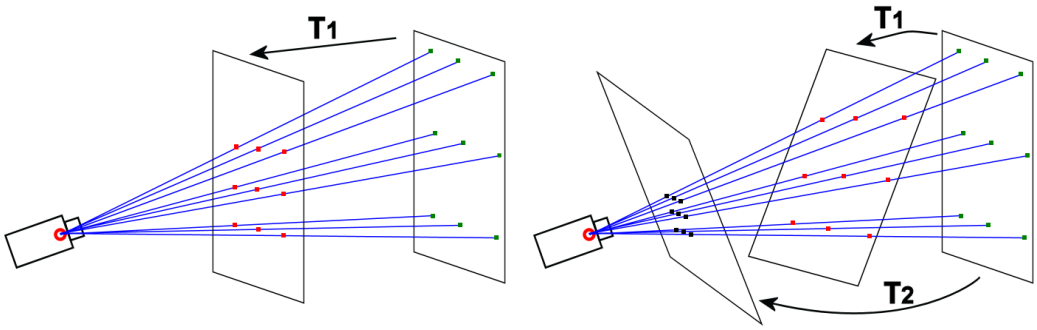
Note that if a ray passes through the origin of the coordinate system in which it is being represented, then the last three components of its Plücker coordinate vector are 0.

2.3.3 General Model Calibration

The general camera model, being non-parametric, requires a completely different method of calibration than methods proposed for the camera models described in Section 2.2. Along with their proposal of the general camera model, Grossberg and Nayar (2001) presented a method for calibrating the general model. The method is conceptually straightforward: for each image pixel, determine two points in 3-space that this pixel sees, then calculate the join of these two points to get the camera ray associated with that pixel, and finally determine the ray's Plücker matrix and store it in the lookup table. In practice, the 3-space points seen by each image pixel are determined by imaging a planar calibration grid in two known positions and locating the points on the grid in each position seen by each pixel. A schematic of the calibration process is shown in Fig. 2.2(a). The relative transformation (usually a translation) between the two grid positions, T_1, must be accurately known in order to achieve a good calibration. The convenience of the calibration method is thus compromised by the need for precise motion control. This contrasts with the perspective camera calibration methods for which the experimental setup consists only of a planar calibration grid in general positions.

2.4 Generic Calibration

Generic camera calibration was proposed by Sturm and Ramalingam (2003) to circumvent the need for precise motion control when calibrating general



(a) (b)

Fig. 2.2: Calibration configurations for (a) the general calibration method; (b) the generic calibration method. The case of central cameras is shown for simplicity.

cameras. As with the calibration method described in the previous section, the goal is to determine the ray direction corresponding to each image pixel. This is achieved in generic calibration by determining the points seen by a pixel on each of three differently orientated calibration grids[2]. Once each ray direction has been calculated, its Plücker matrix is computed and stored in a lookup table as before. Although generic calibration requires an extra grid, the grids can be located in general position without any a-priori knowledge of their poses in the world coordinate system. Thus there is no need for precise motion control in generic calibration, leading to an easier calibration configuration. The generic calibration setup is illustrated in Fig. 2.2(b), where T_1 and T_2 are initially unknown but are determined during the calibration process.

The generic calibration process proposed by Sturm and Ramalingam (2003, 2004) is summarised in Algorithm 1. This calibration method will hereafter be referred to as the standard generic method. The focus of Chapter 4 of this thesis is generic calibration for central cameras, and thus a more detailed discussion of stages 3 and 5 of the central version of the standard generic method is presented in the following sections.

[2]Other versions of generic calibration that operate on non-planar calibration targets are also presented by Sturm and Ramalingam (2003). However, while planar calibration grids are easily formed by attaching a grid printout from a desktop printer onto a planar substrate, accurate 3D calibration targets must be made with machining tools. In the interest of achieving efficient and practical calibration, this thesis focusses on generic calibration with planar grids.

Algorithm 1 Standard generic camera calibration using planar calibration targets

1. Image a calibration grid in a minimum of three different poses such that there is overlap between grid pairs in the image. Take additional overlapping images of the calibration grid in various poses so as to completely cover the image FOV
2. Determine the location seen by each pixel on each grid
3. **Linearly estimate** the poses of the calibration grids using this data and the known constraints. If the camera is central, linearly estimate the effective camera centre.
4. Bundle adjust in order to refine the poses of the initial grids and the ray directions
5. **Estimate the pose** of each of the additional grids, and subsequently the ray directions of additional rays, using geometric constraints followed by bundle adjustment
6. Store all the ray directions as Plücker matrices in a look-up table

2.4.1 Linear Estimation

The linear estimation stage of the standard generic method for central cameras is based on a collinearity constraint: for each camera ray, the camera centre and the world coordinates of the intersection point of that ray with each of the initial calibration grids are collinear. The world coordinate frame is chosen to be coincident with the local coordinate frame of the first grid, and that grid is termed the base grid. A mathematical expression for the constraint is derived by first stacking the world homogeneous coordinate for the centre and for the intersection points of a single ray with each calibration grid in a 4×4 matrix

$$\begin{bmatrix} C_x & Q_x^1 & R_{11}^2 Q_x^2 + R_{12}^2 Q_y^2 + t_1^2 Q_w^2 & R_{11}^3 Q_x^3 + R_{12}^3 Q_y^3 + t_1^3 Q_w^3 \\ C_y & Q_y^1 & R_{21}^2 Q_x^2 + R_{22}^2 Q_y^2 + t_2^2 Q_w^2 & R_{21}^3 Q_x^3 + R_{22}^3 Q_y^3 + t_2^3 Q_w^3 \\ C_z & 0 & R_{31}^2 Q_x^2 + R_{32}^2 Q_y^2 + t_3^2 Q_w^2 & R_{31}^3 Q_x^3 + R_{32}^3 Q_y^3 + t_3^3 Q_w^3 \\ C_w & Q_w^1 & Q_w^2 & Q_w^3 \end{bmatrix} \tag{2.13}$$

where $\mathbf{C} = [C_x \; C_y \; C_z \; C_w]^T$ is the unknown camera centre, R_{mn}^i is the element in row m and column n of the unknown rotation matrix that describes the pose of grid i, $\mathbf{t}^i = \begin{bmatrix} t_x^i & t_y^i & t_z^i \end{bmatrix}^T$ is the unknown translation of grid i, and $\mathbf{Q}^i = \begin{bmatrix} Q_x^i & Q_y^i & Q_z^i & Q_w^i \end{bmatrix}^T$ is the known intersection point of the ray with grid i in

Table 2.1: Coefficients used in Eqns. 2.14

i	j	k	V_i	W_i	M_i	N_i
1	x	x	0	R^2_{31}	0	R^3_{31}
2	x	y	0	R^2_{32}	0	R^3_{32}
3	x	w	0	$-C_z + t^1_z$	0	$-C_z + t^2_z$
4	y	x	R^2_{31}	0	R^3_{31}	0
5	y	y	R^2_{32}	0	R^3_{32}	0
6	y	w	$-C_z + \mathbf{t}^2_z$	0	$-C_z + \mathbf{t}^3_z$	0
7	w	x	$-C_y R^2_{31} + C_z R^2_{21}$	$-C_x R^2_{31} + C_z R^2_{11}$	$-C_y R^3_{31} + C_z R^3_{21}$	$-C_x R^3_{31} + C_z R^3_{11}$
8	w	y	$-C_y R^2_{32} + C_z R^2_{22}$	$-C_x R^2_{32} + C_z R^2_{12}$	$-C_y R^3_{32} + C_z R^3_{22}$	$-C_x R^3_{32} + C_z R^3_{12}$
9	w	w	$-C_y \mathbf{t}^2_z + C_z \mathbf{t}^2_y$	$-C_x \mathbf{t}^2_z + C_z \mathbf{t}^2_x$	$-C_y \mathbf{t}^3_z + C_z \mathbf{t}^3_y$	$-C_x \mathbf{t}^3_z + C_z \mathbf{t}^3_x$

grid is local coordinate frame. Q^i_z values do not appear in Exp. 2.13 as the intersection points by definition lie on the intersecting plane. Exp. 2.13 can be extended by including an additional column for each additional calibration grid with which the ray intersects, but only the minimal case of three grids is shown here for presentation clarity. The collinearity constraint is enforced by ensuring that all 3×3 subdeterminants of the matrix in Exp. 2.13 equal 0. The subdeterminants that include the first 2 columns and the last 2 rows are used to form the following homogeneous equations

$$\begin{aligned} \sum_{i=1}^{9} Q^1_j Q^2_k V_i &= 0 \\ \sum_{i=1}^{9} Q^1_j Q^2_k W_i &= 0 \\ \sum_{i=1}^{9} Q^1_j Q^3_k M_i &= 0 \\ \sum_{i=1}^{9} Q^1_j Q^3_k N_i &= 0 \end{aligned} \tag{2.14}$$

where the values of vectors $\mathbf{V}$, $\mathbf{W}$, $\mathbf{M}$ and $\mathbf{N}$, shown in Table 2.1, are computed up to scale using least squares.

The coefficient pairs are brought to common scale factors λ_1 and λ_2 as

$$\mathbf{V}' = \lambda_1 \mathbf{V} \tag{2.15}$$

$$\mathbf{W}' = \lambda_1 \left(\frac{V_4}{W_1} \right) \mathbf{W} \tag{2.16}$$

$$\mathbf{M}' = \lambda_2 \mathbf{M} \tag{2.17}$$

$$\mathbf{N}' = \lambda_2 \left(\frac{M_4}{N_1} \right) \mathbf{N} \tag{2.18}$$

and by multiplying variables from Table 2.1 the matrix equation of Eqn. 2.24 is formed, which is rewritten as

$$A\mathbf{u} = \mathbf{x} \tag{2.19}$$

A is rank 12, so $\mathbf{u}$ can be extracted in a subspace spanned by the particular solution and two basis vectors

$$\mathbf{u} = \mathbf{a} + l_1\mathbf{b} + l_2\mathbf{c} \tag{2.20}$$

Next constraints on the rotational elements, resulting from the orthonormality of rotation matrices, are applied to form a new matrix equation, given in Eqn. 2.25, which is rewritten as

$$D\mathbf{v} = \mathbf{y} \tag{2.21}$$

D is rank 5 and so $\mathbf{v}$ can be extracted in a subspace spanned by 4 vectors

$$\mathbf{v} = \mathbf{d} + g_1\mathbf{e} + g_2\mathbf{f} + g_3\mathbf{h} \tag{2.22}$$

By simulation it has been shown that certain elements of these vectors are always 0, allowing a solution to be obtained for $\{l_1, l_2\}$ and hence $\mathbf{u}$. Finally, the solutions for the camera centre and unknown poses are obtained as

$$\begin{array}{lcllcl}
C_x & = & u_1 & C_y & = & u_2 \\
C_z & = & \pm\sqrt{\frac{u_6u_6+u_7u_8}{-V_4'V_5'}} & & & \\
\lambda_1 & = & \pm\frac{\sqrt{u_5^2+u_7^2+\left(V_4'C_z\right)^2}}{C_z} & \lambda_2 & = & \pm\frac{\sqrt{u_{11}^2+u_{13}^2+\left(M_4'C_z\right)^2}}{C_z} \\
R_{11}^2 & = & \frac{u_5}{\lambda_1 C_z} & R_{12}^2 & = & \frac{u_6}{\lambda_1 C_z} \\
R_{21}^2 & = & \frac{u_7}{\lambda_1 C_z} & R_{22}^2 & = & \frac{u_8}{\lambda_1 C_z} \\
R_{31}^2 & = & \frac{V_4'}{\lambda_1} & R_{32}^2 & = & \frac{V_5'}{\lambda_1} \\
R_{11}^3 & = & \frac{u_{11}}{\lambda_2 C_z} & R_{12}^3 & = & \frac{u_{12}}{\lambda_2 C_z} \\
R_{21}^3 & = & \frac{u_{13}}{\lambda_2 C_z} & R_{22}^3 & = & \frac{u_{14}}{\lambda_2 C_z} \\
R_{31}^3 & = & \frac{M_4'}{\lambda_2} & R_{32}^3 & = & \frac{M_5'}{\lambda_2} \\
t_x^2 & = & \frac{u_3+\lambda_1 C_x C_z}{\lambda 1 C_z} & t_y^2 & = & \frac{u_4+\lambda_1 C_y C_z}{\lambda 1 C_z} \\
t_z^2 & = & \frac{\lambda_1 C_z t_y^2 - V_9}{\lambda_1 C_y} & t_x^3 & = & \frac{u_9+\lambda_2 C_x C_z}{\lambda_2 C_z} \\
t_y^3 & = & \frac{u_{10}+\lambda_2 C_y C_z}{\lambda_2 C_z} & t_z^3 & = & \frac{\lambda_2 C_z t_y^3 - M_9}{\lambda_2 C_y}
\end{array} \tag{2.23}$$

$$
\begin{bmatrix}
0 & -V'_4 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -V'_5 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -V'_6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-V'_4 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-V'_5 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-V'_6 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -M'_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & -M'_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & -M'_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
-M'_4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
-M'_5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
-M'_6 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
C_x \\
C_y \\
\lambda_1 C_z\left(t_x^2 - C_x\right) \\
\lambda_1 C_z\left(t_y^2 - C_y\right) \\
\lambda_1 C_z R_{11}^2 \\
\lambda_1 C_z R_{12}^2 \\
\lambda_1 C_z R_{21}^2 \\
\lambda_1 C_z R_{22}^2 \\
\lambda_2 C_z\left(t_x^3 - C_x\right) \\
\lambda_2 C_z\left(t_y^3 - C_y\right) \\
\lambda_2 C_z R_{11}^3 \\
\lambda_2 C_z R_{12}^3 \\
\lambda_2 C_z R_{21}^3 \\
\lambda_2 C_z R_{22}^3
\end{bmatrix}
=
\begin{bmatrix}
V'_7 \\ V'_8 \\ V'_9 \\ W'_7 \\ W'_8 \\ W'_9 \\ M'_7 \\ M'_8 \\ M'_9 \\ N'_7 \\ N'_8 \\ N'_9
\end{bmatrix}
\tag{2.24}
$$

$$
\begin{bmatrix}
a_5b_6 + a_6b_5 + a_7b_8 + a_8b_7 & a_5c_6 + a_6c_5 + a_7c_8 + a_8c_7 & b_5c_6 + b_6c_5 + b_7c_8 + b_8c_7 & b_5b_6 + b_7b_8 & c_5c_6 + c_7c_8 & V'_4V'_5 & 0 & 0 \\
a_{11}b_{12} + a_{12}b_{11} + a_{13}b_{14} + a_{14}b_{13} & a_{11}c_{12} + a_{12}c_{11} + a_{13}c_{14} + a_{14}c_{13} & b_{11}c_{12} + b_{12}c_{11} + b_{13}c_{14} + b_{14}c_{13} & b_{11}b_{12} + b_{13}b_{14} & c_{11}c_{12} + c_{13}c_{14} & M'_4M'_5 & 0 & 0 \\
2a_5b_5 + 2a_7b_7 & 2a_5c_5 + 2a_7c_7 & 2b_5c_5 + 2b_7c_7 & b_5^2 + b_7^2 & c_5^2 + c_7^2 & V'^2_4 & -1 & 0 \\
2a_6b_6 + 2a_8b_8 & 2a_6c_6 + 2a_8c_8 & 2b_6c_6 + 2b_8c_8 & b_6^2 + b_8^2 & c_6^2 + c_8^2 & V'^2_5 & -1 & 0 \\
2a_{11}b_{11} + 2a_{13}b_{13} & 2a_{11}c_{11} + 2a_{13}c_{13} & 2b_{11}c_{11} + 2b_{13}c_{13} & b_{11}^2 + b_{13}^2 & c_{11}^2 + c_{13}^2 & M'^2_4 & 0 & -1 \\
2a_{12}b_{12} + 2a_{14}b_{14} & 2a_{12}c_{12} + 2a_{14}c_{14} & 2b_{12}c_{12} + 2b_{14}c_{14} & b_{12}^2 + b_{14}^2 & c_{12}^2 + c_{14}^2 & M'^2_5 & 0 & -1
\end{bmatrix}
\begin{bmatrix}
l_1 \\ l_2 \\ l_1l_2 \\ l_1^2 \\ l_2^2 \\ C_z^2 \\ \lambda_1^2C_z^2 \\ \lambda_2^2C_z^2
\end{bmatrix}
=
\begin{bmatrix}
a_5a_6 + a_7a_8 \\
a_{11}a_{12} + a_{13}a_{14} \\
a_5^2 + a_7^2 \\
a_6^2 + a_8^2 \\
a_{11}^2 + a_{13}^2 \\
a_{12}^2 + a_{14}^2
\end{bmatrix}
\tag{2.25}
$$

The final columns of R^2 and R^3 are calculated as the cross products of their first two columns.

2.4.2 Pose Estimation

Pose estimation is necessary in generic calibration in order to allow the entire FOV of the camera to be calibrated. After completion of the linear stage, the Plücker matrices for the rays that intersect any of the three initial grids are known, but the remaining ray directions have still to be determined. By imaging an additional calibration grid in a position such that some of the previously calibrated rays intersect with it, the grid's pose can be estimated. Then the unknown ray directions can be determined directly from their intersections with the additional grid and from the pose of the additional grid.

The pose estimation scheme of the standard generic method, which can be applied to both central and non-central cameras, is based on solving a set of simultaneous second order equations. Given calibrated rays i and j, the distance d_{ij} between their intersections with the calibration grid of unknown pose can be determined directly. The depths λ_i and λ_j of the intersection points, measured along the ray directions $\{\mathbf{r}_i, \mathbf{r}_j\}$ from the ray anchor points $\{\mathbf{A}_i, \mathbf{A}_j\}$, can be computed by simultaneously solving

$$||\mathbf{A}_i + \lambda_i \mathbf{r}_i - \mathbf{A}_j - \lambda_j \mathbf{r}_j||^2 = d_{ij}^2 \tag{2.26}$$

for $i, j = \{l, m, n\}$, $i \neq j$. Additional rays are used to determine the correct pose from the eight possible solutions (four sign-reflected pairs).

2.5 Performance Comparison of Generic and Pinhole-plus-Distortion Camera Calibration

This section provides a benchmark of the performance of the central generic calibration method with respect to the well known and well understood standard perspective camera calibration method of Zhang (1998). The goal is to inform the practitioner of the level of precision to be expected with the generic

calibration strategy. As the comparison is conducted with respect to a radially distorted perspective model, this is the imaging modality that is utilised. Favorably, due to the properties of the general model, the same performance levels can be equally extended to other imaging modalities.

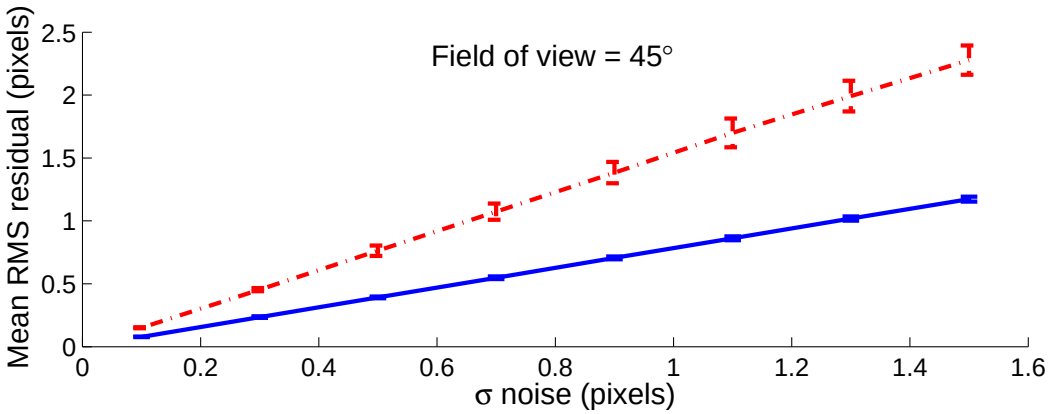
The calibration work of Sturm and Maybank (1999) and of Zhang (2000) were the first that fully exploited multiple views of planar grids taken from unknown viewpoints. These works describe how to obtain linear constraints on the intrinsic parameters of the camera from homographies between the scene and image planes. Mainly due to an executable distributed by Zhang[3] and an open source implementation within the OpenCV library, this plane-based approach has become a standard tool for calibrating perspective cameras, and will hereafter be referred to as the standard perspective method. Thus, as this method is widely used, and its performance has been well characterised, it is used to benchmark the performance of the central generic calibration method. Lens distortion is modelled in the standard perspective calibration method using Eqn. 2.4 in a final full nonlinear estimation process.

The results are primarily compiled over simulated data. These findings are subsequently validated on real image samples. For the simulated data, comparative experiments are designed to characterise the sensitivity to noisy input data, and to assess the nonlinearity removal with respect to increasing lens distortion. With real images, three samples with increasing FOV are used. These are similarly analysed for residual distortion levels after performing distortion correction, and for a motion estimation experiment.

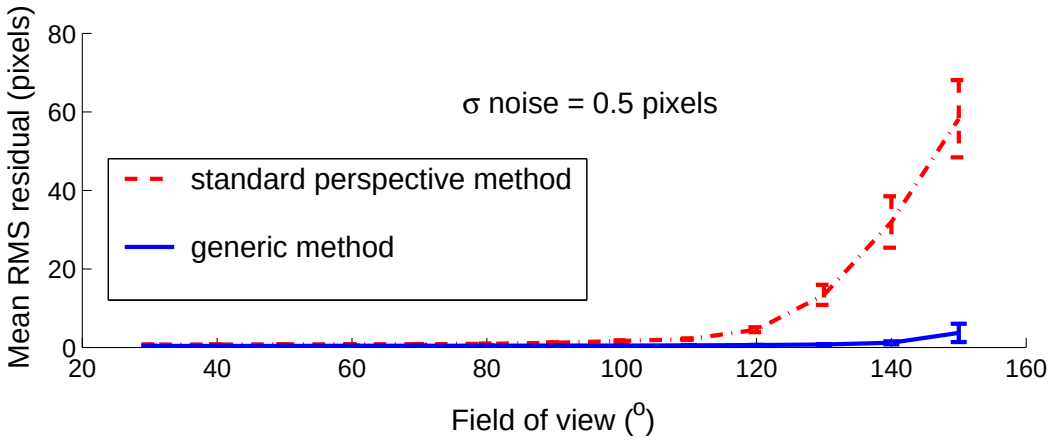
2.5.1 Experiments with Synthetic Data

Two experiments are conducted using data for a synthetic camera with a 640×640 resolution. Firstly, the sensitivity to noise of the central generic calibration method is compared with the sensitivity to noise of the standard perspective method. Increasing levels of Gaussian noise are induced in the locations of 999 random image points, and then each calibration technique is implemented using the resulting data. This process is conducted 50 times for each level of noise, whereupon the distortion residual statistics shown in Fig. 2.3(a) are computed. Note that for the central generic method the noise causes millimetre

[3] http://research.microsoft.com/∼zhang/calib (Accessed November 2006)



(a)



(b)

Fig. 2.3: Mean and standard deviation of distortion residuals for the standard perspective and central generic methods under (a) increasing noise; and (b) increasing distortion.

error in the ray-plane intersections, which is computed by back-projection of the noisy pixel data onto each plane. As with the standard perspective method, the effect of noise in the central generic method is shown to be linearly proportional.

The second experiment aims to investigate the precision of the calibration in terms of removing nonlinearities due to lens distortion. Increasing levels of lens distortion are simulated with the equidistance distortion function of Eqn. 2.6, which is chosen to be different from the distortion model used in the standard perspective calibration method. The model is manipulated to give an increasing field of view from 30^o to 150^o. Distortion residuals following calibration are compiled over 50 trials with 999 random points. Fig. 2.3(b)

shows the resulting statistics. As expected, for high distortion the residuals of the standard perspective method increase. However, the generic method residuals maintain a low mean and standard deviation throughout. This level of performance can be expected regardless of the imaging modality due to the non-parametric nature of the general model. The increase in the mean and the standard deviation of the residual at 150^o for the generic method can be attributed to the infinite planes used in the synthetic experiments. For certain camera rays and plane pose configurations the acute angles between the rays and the calibration planes with which they intersect can be very small. Consequently the local coordinates of these intersection points are more sensitive to noise than those intersection points for which the rays are less acute. When using infinite planes the noise is thus magnified at the periphery of the camera FOV, and it increases as the FOV increases. In calibrations with real data the calibration grids have finite extent, thus limiting the minimum acute angle between the rays and the grids.

Forming distortion corrected images is a 3-space operation within the generic calibration approach. As each pixel maps to a ray, distortion correction amounts to determining the intersections of these rays with a synthetic image plane. In the presented experiments, the synthetic plane is chosen as the plane that passes closest to the grids used in the calibration and that is perpendicular to the principal ray. As the location of this plane affects the scale and location of the corrected image points, a homography is applied to map these intersected points to their known metric positions. Since the homography is a projective transformation it preserves collinearity, and thus it does not affect the evaluation of the generic method for capturing non-linear distortions.

2.5.2 Experiments with Real Data

Three real images are analysed for each calibration method with respect to distortion residuals, and one image set is analysed with respect to a motion estimation task. Sample input images are shown in Fig. 2.4 that indicate the levels of distortion present. These were taken at three different zoom levels using a Nikon CoolPix 4500 camera fitted with a FC-E8 fisheye converter (183^o FOV).

Following calibration with both the standard perspective and central generic

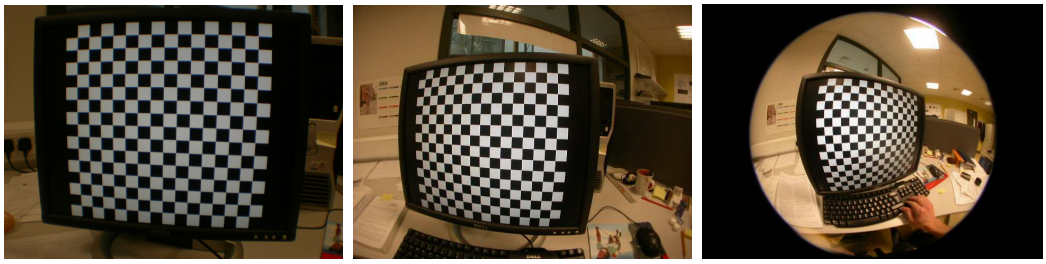


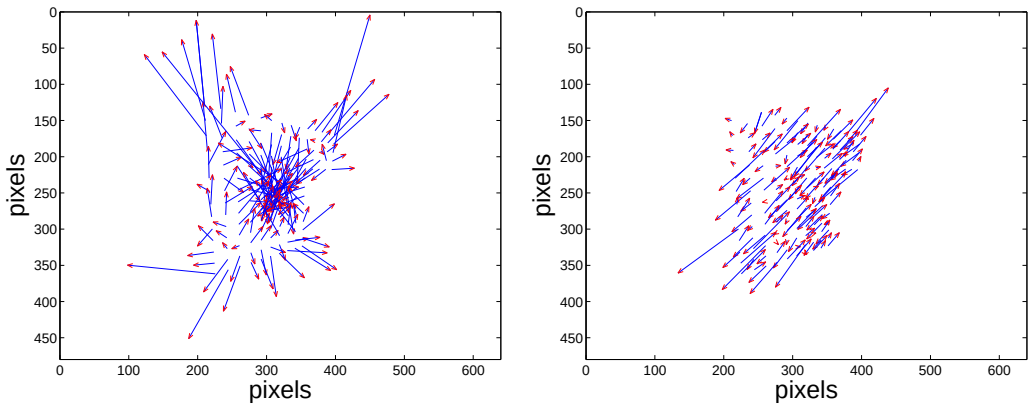
Fig. 2.4: Three levels of distorted images used in the experiments with real data.

methods, the distortion residuals are measured after applying homographies between the known metric grid structure and the perspectively corrected images determined from the calibration information. The statistics of the resulting residuals are presented in Table 2.2. The results show that the accuracy of the generic calibration is less than that of the standard perspective method for low distortion levels. As the level of distortion increases, the magnitude and standard deviation of the errors for the standard perspective method increase, while the generic method maintains its accuracy throughout. For image set 3, a vector plot of the residuals for each method is shown in Fig. 2.5. This shows a classic distortion bias pattern for the standard perspective method, caused by the least-squares determination of the fitting homography, whereby the parametric model overcompensates for distortion towards the image (distortion) centre, and undercompensates towards the image periphery. In contrast, although there is a systematic error in the generic method residuals, there is no distortion bias present. The systematic error is possibly due to error in the estimate of the camera centre. Chapter 4 proposes improvements to the method of generic calibration that aim to reduce this error by way of improved camera centre and grid pose estimation.

These results are in broad agreement with the simulated results - the error in the standard perspective method increases as the distortion increases, whereas the generic method is not sensitive to changes in distortion. The difference in error magnitudes between the simulated data and real data results can be attributed partly to the images used in the real experiments. Only the areas covered by the grids in the images are undistorted, and thus distortion residuals for the periphery of the images, where the distortion is greatest, are not calculated. Grids do not cover complete images due to the difficulty in accurately extracting corners of severely distorted grids. This issue is addressed

Table 2.2: Mean and standard deviation of the distortion residuals after perspective correction for three real images. Error is measured in pixels.

Method	Error type	Image set		
		1	2	3
Standard perspective	RMS	0.09	0.13	1.20
	SD	0.03	0.06	0.67
Central generic	RMS	0.65	0.72	0.68
	SD	0.35	0.39	0.38



(a) (b)

Fig. 2.5: Vector plots of residuals for (a) the standard perspective method; and (b) the central generic method. Vectors are scaled $\times 50$.

in detail in Chapter 3, where a corner extraction method designed for severely distorted grid images is presented.

A second experiment is also conducted to assess the calibration precision. This involves calibrating with each method, and then performing a motion estimation task. Images are captured of a planar grid attached to a linear motion controller. Five images are taken with translation increments of exactly $25mm$. Homographies are calculated between each step and subsequently decomposed to recover the motion. For calibrated images the homography H can be decomposed as $H = \left(R_{[1\ 2]}|R\mathbf{t}\right)$, where R is the relative rotation and $\mathbf{t}$ is the translation. These translation vectors are plotted end to end in Fig. 2.6. For visualisation purposes, the difference between each vector and the average vector is scaled by 20. As can be seen, the generic method outperforms the standard perspective method in the recovery of the translation component.

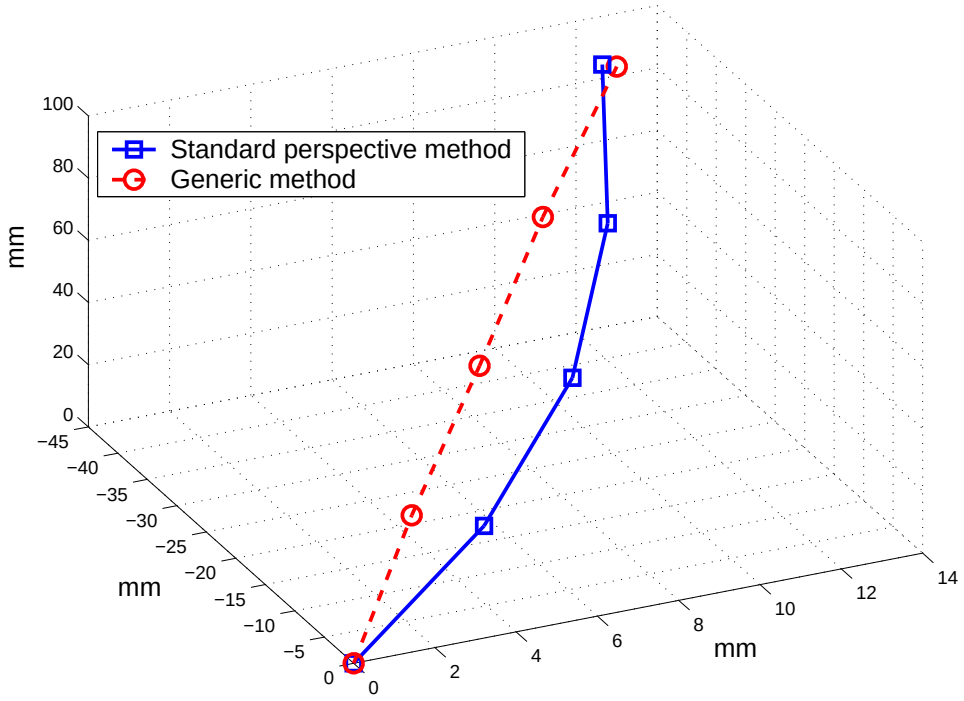


Fig. 2.6: Translation estimation using calibration data from the standard perspective method and from the generic method. Vector errors are scaled ×20.

2.5.3 Distortion Correction

Samples of perspectively corrected fisheye images formed using the results from each calibration method are illustrated in Fig. 2.7. The camera zoom setting corresponds to that of image set 2 in Table 2.2. Again the generic method is seen to outperform the standard perspective method around the periphery of the image. World vertical lines are off-vertical in Fig. 2.7(e) due to the synthetic plane used for the distortion correction, as discussed in Section 2.5.1. Nevertheless, straight world lines appear straight in the image, indicating precise correction of distortion.

2.6 Discussion

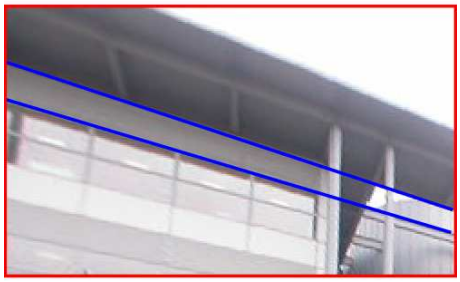
This chapter introduces the fundamentals of camera calibration for application in computer vision, and examines the current calibration approaches for a range of camera types by way of the camera and distortion models. It is shown that the majority of camera models in the literature are specific to certain cameras. In contrast, the general camera model is applicable to any camera

(a)

(b)

(c)

(d)



(e)

Fig. 2.7: (a) Original fisheye image; (b) perspectively corrected image using standard perspective method calibration data; (c) enlargement of region contained within red rectangle in (b), showing residual non-linear distortion; (d) perspectively corrected image using generic method calibration data; (e) enlargement of region contained within red rectangle in (d), showing removal of all visible non-linear distortion.

and so can overcome the difficulty of selecting the correct model and the correct parameter set to most accurately capture distortion. The general model and the generic method for its calibration are discussed, and further examined through a performance comparison of central generic calibration and standard perspective camera calibration.

Camera distortion inherent to perspective cameras is small, thus allowing the pinhole model to encapsulate its operation. The importance of the pinhole model, as discussed in Section 2.1, lies in its preservation of collinearity relationships and in the simplicity of its associated projection scheme. Deviation from the pinhole model is induced by lens non-linearity for dioptric cameras, and by lens and mirror non-linearity for catadioptric cameras. Section 2.2 discusses a hierarchy of camera distortions that are characterised by the camera type. For perspective cameras, a simple polynomial radial distortion model, using one or two parameters, is reported in the literature to remove the majority of lens distortion. Tangential distortion in perspective cameras contributes minimally to overall image distortion and is generally excluded from models for such cameras. Image distortion necessarily escalates as the FOV of a camera increases. For wide-angle and fisheye modalities the distortion is a desired property that is designed into the associated camera lenses. Although some proposed wide-angle camera models are simply extensions of the perspective model augmented by additional radial distortion terms, for fisheye cameras specific fisheye projection functions are generally employed that directly incorporate radial distortion. Catadioptric cameras are modelled using either a geometric description of the mirror shape, or using a more general polynomial approach. Relying on camera geometry implies that there is no aberration in the mirror surface. Parametric models for describing non-central catadioptric cameras have been presented for quadric mirror surfaces in which the camera and mirror are misaligned with each other, but for general non-central cameras no parametric models exist.

The brief review of camera models and distortion indicates that increasing numbers of parameters, and increasingly complex parametric camera models, are required as the camera type deviates from ideal pinhole. Section 2.3 discusses the general camera model proposed by Grossberg and Nayar (2001). It is a unifying model that can be equally applied to all types of cameras, independently of the distortion present in them, the mirror shape, or the radial symmetry of the lenses. A lookup table captures the mapping between cam-

era rays and image pixels, so the model is non-parametric. The versatility of the model clearly makes it attractive for the calibration of non-conventional cameras.

Two methods of calibration for the general camera model have previously been proposed. The general calibration method, outlined in Section 2.3.3, is conceptually simple but requires precise motion control during the calibration process. Generic calibration was proposed as a more flexible alternative in which the only requirement for calibration is a minimum of three images of a planar calibration grid in different but unknown poses, and where the second and third grids each have common overlap with the first grid. Generic calibration algorithms vary slightly depending on the camera centrality, and Section 2.4 details the algorithm for generically calibrating central cameras. In particular, the linear estimation and pose estimation stages of the algorithm are comprehensively presented, as these will be dealt with further in Chapter 4.

Section 2.5 provides a side by side comparison of central generic calibration with the well established standard perspective camera calibration method of Zhang (1998). Experiments are conducted with simulated data for sensitivity to noise and to increasing distortion. Experiments with real data consider three different levels of fisheye distortion, and analyse the calibrations in terms of distortion correction and a motion estimation task. From the results it is concluded that the generic calibration method achieves good performance levels at low to mid distortions, although the standard perspective method performs better for these distortions. Crucially, for higher distortion levels the accuracy of the generic method is maintained, whilst the accuracy of the standard perspective method significantly reduces. Overall, this performance characterisation clearly shows that the performance of the generic calibration method is effectively independent of the nonlinearities in the imaging sensor, and thus it outperforms the existing standard perspective method when applied to cameras incorporating significant distortion.

Generic calibration is further developed in Chapter 4, but first methods enabling convenient acquisition of input data for calibration must be addressed.

Chapter 3

Planar Grids for Non-Conventional Camera Calibration

Planar calibration targets, or calibration grids, are the de facto standard method of calibrating cameras for computer vision tasks. They are easily manufactured, for example by printing the grid pattern on a desktop printer and then mounting the pattern on a planar substrate. For these reasons calibration grids dominate over 3D calibration targets in practical camera calibration. This chapter examines two important issues that are encountered when using patterned grids for accurate calibration of cameras with large image distortion. Firstly the difficulty of extracting grid feature points in distorted images is addressed. Many existing calibration methods for wide-angle and omnidirectional cameras require calibration grid correspondences in order to perform calibration (Mei and Rives, 2007, Scaramuzza et al., 2006, Bastardly et al., 2008). The detection and ordering of calibration grid features in images from such cameras is thus an important, but often overlooked, practical problem. This chapter contributes a novel method for chessboard grid corner extraction in images from non-conventional cameras. Secondly, the benefits of using dynamic grids displayed on a flat screen monitor for acquiring dense feature sets for calibration are demonstrated. An implementation of these dynamic grids is used to evaluate their performance with respect to accuracy and robustness relative to standard chessboard grid localisation techniques.

As discussed in Section 1.2.1, chessboard grid corners are immune to both perspective bias and to distortion bias, whereas this is not the case for planar grids consisting of circular dot patterns. In addition, Mallon and Whelan (2007*b*) show that corner detection outperforms edge intersection detection when recovering the pattern features from chessboard grids. Therefore this chapter focuses on the detection and recovery of corners from images of planar chessboard calibration grids. The extraction of corners from such images can be conveniently broken into three stages - corner detection, corner localisation, and corner ordering. Chessboard corner extraction for camera calibration can be a difficult task when operating on images with significant distortion, such as those produced by fisheye, omnidirectional, and flexible mirror cameras. The distortion causes the corners to shift non-linearly in the image, and as a result the corners in the image no longer lie on a regular grid. Consequently many of the standard corner extraction methods that have been designed for distortion-free cases will fail on such images. Generally failure occurs in the ordering stage, since the corners no longer lie on straight lines joining corners on opposite sides of the grid. The CELECT method, a new semi-automatic corner detection and ordering method that is suitable for application to highly distorted images, is proposed and detailed in Section 3.1.2. By taking an edge contour tracing approach, parametric modelling and assumptions about the image distortion are not required. Simulations demonstrating the robustness of the method for synthetic data under varying conditions are presented in Section 3.3.1. Experiments with real images from four different cameras with large distortion are detailed in Section 3.3.2.

The remainder of this chapter addresses the issue of extracting dense and accurate input data from planar calibration grids. The extraction and accurate localisation of features in images of calibration grids results in a set of grid to image correspondences. Feature density in the images depends on grid spacing, grid size and camera to grid distance. However, there is a practical limit to the density that can be achieved using such static grids. In general, calibration accuracy and robustness to noise improve as the density of feature correspondences increases. As described in Chapter 2, the generic calibration method requires knowledge of these correspondences in order to determine the intersection points of camera rays with the calibration grid (i.e. the location seen on the grid by each camera pixel). Generic calibration can achieve pixel level calibration as discussed in Chapter 2, and thus it seems appropriate to

use pixel level data as input for that method. Regardless of the calibration method, dense feature data is desirable. Such data can be obtained by the use of spatio-temporally varying grids displayed on a flat screen monitor. These grids are termed active grids in this thesis, and while this method has frequently been used previously (Grossberg and Nayar, 2001, Sagawa et al., 2005, Bonfire et al., 2006, Tardif and Sturm, 2005) for acquiring dense feature point sets for calibration, no discussion has been proffered on their performance for calibration purposes. An explanation of the technique of active grids, and an example implementation, are given in Section 3.2.2. Results on the characterisation of their performance relative to standard localisation techniques are presented in Section 3.3.3. Experiments with real images are documented in Section 3.3.4. Finally, the accuracy of active grids for application to distortion correction tasks for both continuous and discontinuous cameras is examined.

3.1 Detection and Ordering of Grid Corners

The proposed method for corner detection and ordering applies low-level image processing operations in local windows in order to trace along grid edge contours. This simple concept behind the algorithm gives the method broad applicability. Edge contours are strong and robust features in images of calibration grids: by tracing a window along these contours, corner detection and ordering are achieved in this window without resorting to global image processing. Such a gradient approach overcomes the drawbacks of threshold selection associated with region based detection and ordering methods. The proposed method is termed CELECT, which stands for Corner Extraction via Local Edge Contour Tracing. The method is semi-automatic: the user is required to select the four outer corners of the grid, and to enter the number of grid corners in each row and in each column, *width* and *height* respectively. Parametric descriptions of the grid contours are not assumed, therefore the CELECT method can operate on cropped images (in which the distortion centre may not be near the image centre), and on images from non-conventional cameras that contain significant distortion.

Prior knowledge of the distortion induced by a camera allows a model-based detection and ordering algorithm to be applied (Sturm and Barreto, 2008, Mei and Rives, 2007). However, the goal of the CELECT method is to achieve

chessboard feature extraction from images with varying distortions caused by distinctly different camera types. Parametric models for fisheye distortion, omnidirectional distortion, and flexible mirror distortion do not all have significant overlap. Furthermore, flexible mirror cameras can introduce such a range of possible image distortions that the distortion is practically impossible to model a-priori. In these cases, incorporating prior knowledge of camera deformation would reduce the general applicability of any method. In contrast, the CELECT method can operate on continuous images from any camera type, without the restrictions associated with the incorporation of camera models.

Ordering of chessboard corners is the most difficult part of chessboard feature extraction for distorted images. However, the feature detection stage must first be applied.

3.1.1 Corner Detection

In the locality around a chessboard corner feature in a distorted image, corner detection can generally proceed in the same way as for distortion-free images, since sufficiently small windows around the corner feature can be considered to be distortion-free. The size of a sufficiently small window is a function of the local image distortion. Corner detection has long been recognised as fundamental to any computer vision task requiring feature points, and consequently many corner detection methods have been presented over the years. The Sobel, Roberts, Canny, Moravec and SUSAN detectors are some popular examples. Perhaps the most commonly used method is the Harris corner detector (Harris and Stephens, 1988). Chessboard grid images present high contrast between the dark and light squares, so the simplicity and efficiency of the Harris detector makes it suitable for the detection stage of the CELECT algorithm.

Harris Detector

The Harris corner detector improves on the Moravec corner detector by making the detector response isotropic and smoothly windowed, and by incorporating the variation of the filter response with the direction of shift. The detector is described by a 2×2 matrix M whose elements are partial derivative

functions. When the filter is applied to a corner, both of the eigenvalues of M are large; for an edge, one eigenvalue is large and one is small; for a homogeneous region, both eigenvalues are small. The need to explicitly evaluate the eigenvalues is circumvented by calculating their sum and product as $trace(M)$ and $det(M)$, respectively. The filter response, R, is then given by $R = det(M) - k\,(trace(M))^2$, where k is a tuning factor. This corner detector, incorporating a Gaussian smoothing window, is implemented to detect chessboard grid corners in the CELECT algorithm.

3.1.2 CELECT Algorithm

Corner detection and ordering are carried out simultaneously in the CELECT method. Beginning at the first corner, the algorithm steps along grid edge contours, where the edge contour being traced is detected by applying a Hough transform in a local neighbourhood around the current point. The next point in the trace is selected as a point further along this contour. At each step the Hough transform is reapplied, and the edge stepping process is repeated until the distance to a specified end point is reached. Corners along the traced contour are detected locally during the edge tracing and are stored for later processing.

The algorithm first traces clockwise along the four outer contours of the grid, beginning at the first user selected corner. For simplicity of explanation, and without loss of generality, it will be assumed here that the first outer corner selected is the upper left corner. Therefore when outer contour tracing is completed, the algorithm traces down the column contours of the grid. Tracing is achieved by applying a moving binary circular window to the current anchor edge contour point, $\mathbf{p}_{cur}$, and then searching for an appropriate edge contour point in the window to form the next anchor point, $\mathbf{p}_{new}$, as shown in Fig. 3.1. Within each window, lines detected by the Hough transform are filtered based on their distance from the window centre and based on how their orientations align with an estimate of the desired orientation. For a suitably small search window, curved line segments can be detected by a Hough transform. This is similar to the approach used by Devernay and Faugeras (2001) to extract distorted lines in images. Corners are detected at each iteration by applying the Harris detector to the windowed region. At the end of the tracing process for each contour all the detected corners are processed by performing merging

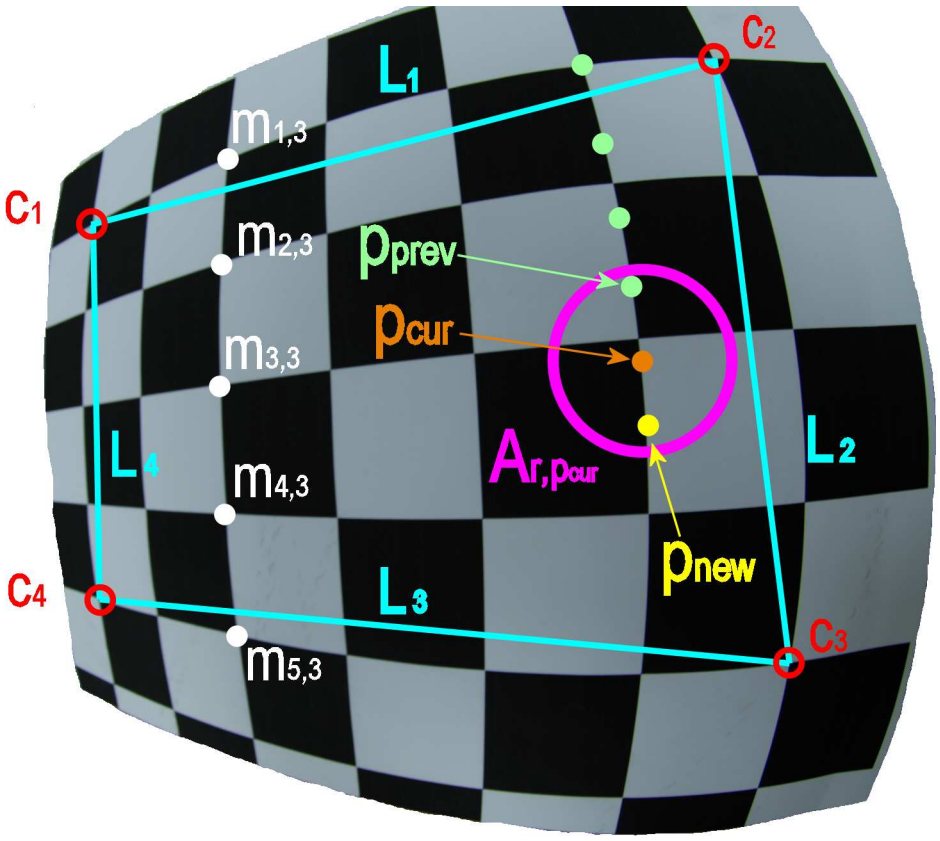


Fig. 3.1: Parameters used in the CELECT algorithm for detection and ordering of chessboard grid corners.

and discarding according to simple rules, and the resulting corners are recorded as ordered calibration grid corners.

The following sections present a detailed description of the algorithm and of the variations in it that depend on the edge contour being traced.

Outer Edge Contours

The first step in the CELECT algorithm is to detect the four corner points $\mathbf{c}_i$ $(i = 1, 2, 3, 4)$ closest to the four user selected outer grid corners in the image, I. Next assign $\mathbf{c}_1$ to $\mathbf{p}_{cur}$. Apply a window $A_{r,\mathbf{p}_{cur}}(I)$ to I, where the window radius is r and the window centre is $\mathbf{p}_{cur}$. Corners $\mathbf{n}_q$ in the windowed region are detected by a Harris corner detector and assigned to the set of candidate corners, $\mathbb{M}$. Perform Canny edge detection, $\mathcal{C}$, on the windowed region and take the Hough transform, $\mathcal{H}$, of the result. The resulting Hough points, $\mathbf{h}(R, \theta)$, are described by the distance of the corresponding image line from the window centre, R, and by the orientation of the corresponding image line, θ. Apply

non-maximal suppression to the points in the Hough image to get the set $\mathbb{S}_h$ of between two and five Hough peaks that correspond to grid edge contour lines in the windowed region

$$\mathbb{S}_h = \mathcal{H}(\mathcal{C}(A_{r,\mathbf{p}_{cur}}(I))) \tag{3.1}$$

The side lines L_i between the outer corner points are given by

$$L_i = \mathbf{c}_i \times \mathbf{c}_{(i \bmod 4)+1} \quad i = 1, 2, 3, 4 \tag{3.2}$$

Determine the set of Hough points $\mathbb{S}_d$ that are close to the window centre, since distant points may belong to an adjacent row or column

$$\mathbb{S}_d = \{\mathbf{h}(R, \theta) \in \mathbb{S}_h | R < t_3\} \tag{3.3}$$

From this extract the set $\mathbb{S}_e$ of Hough points that have angular separation greater than a threshold angle t_4 from L_4

$$\mathbb{S}_e = \{\mathbf{h}(R, \theta) \in \mathbb{S}_d | (|\theta - \angle L_4|) > t_4\} \tag{3.4}$$

and select the best candidate $\mathbf{h}_p(R, \theta)$ from $\mathbb{S}_e$ to have minimum angular separation from the side line being traced

$$\mathbf{h}_p(R, \theta) = \underset{\mathbf{h}(R,\theta) \in \mathbb{S}_e}{\operatorname{argmin}} (|\theta - \angle L_i|) \tag{3.5}$$

The linear approximation to the edge contour in the windowed region is

$$L_w = \mathcal{H}^{-1}(\mathbf{h}_p(R, \theta)) \tag{3.6}$$

The set $\mathbb{E}$ of candidate edge contour points in the windowed region is populated by those points $\mathbf{p}$ that are within t_5 pixels of the approximated edge

$$\mathbb{E} = \{\mathbf{p} \in \mathbb{S}_c | \perp \mathrm{d}(\mathbf{p}, L_w) < t_5\} \tag{3.7}$$

where

$$\mathbb{S}_c = \mathcal{C}(A_{r,\mathbf{p}_{cur}}(I)) \tag{3.8}$$

and where $\perp \mathrm{d}(\mathbf{p}, L)$ is the perpendicular distance between point $\mathbf{p}$ and line L. The best new anchor point $\mathbf{p}_{new}$ is selected as the furthest point in $\mathbb{E}$ from the window centre

$$\mathbf{p}_{new} = \underset{\mathbf{p} \in \mathbb{E}}{\operatorname{argmax}}(\mathrm{d}(\mathbf{p}, \mathbf{p}_{cur})) \tag{3.9}$$

that is in the direction of the next clockwise corner

$$(\mathbf{p}_{new} - \mathbf{p}_{cur}) \cdot (\mathbf{c}_{(i \bmod 4)+1} - \mathbf{c}_i) > 0 \tag{3.10}$$

where $\mathrm{d}(\mathbf{p}_i, \mathbf{p}_j)$ is the Euclidean distance between points $\mathbf{p}_i$ and $\mathbf{p}_j$. $\theta_{\mathbf{h}_p}$ is then assigned to θ_{prev}, $\mathbf{p}_{cur}$ is assigned to $\mathbf{p}_{prev}$ and is added to the edge contour set $\mathbb{P}$, and $\mathbf{p}_{new}$ is assigned to $\mathbf{p}_{cur}$.

The above process with some modifications is then repeated in order to trace along the edge contour until $\mathbf{p}_{cur}$ is within t_6 pixels of $\mathbf{c}_2$. When $\mathbf{p}_{cur}$ is not an outer corner point, as shown in Fig. 3.2, θ_{prev} and $\mathbf{p}_{prev}$ are available from the previous iteration, and the edge tracing is simplified. Since the direction of the edge contour being traced is available from the previous anchor points, Eqns. 3.4 and 3.5 are replaced by

$$\mathbf{h}_p(R, \theta) = \underset{\mathbf{h}(R,\theta) \in \mathbb{S}_d}{\operatorname{argmin}} (|\theta - \theta_{prev}|) \tag{3.11}$$

which minimises the change in orientation of the approximation to the edge contour being traced. Eqn. 3.9 is replaced by

$$\mathbf{p}_{new} = \underset{\mathbf{p} \in \mathbb{E}}{\operatorname{argmax}}(\mathrm{d}(\mathbf{p}, \mathbf{p}_{prev})) \tag{3.12}$$

and Eqn. 3.10 is replaced by a new direction constraint equation

$$(\mathbf{p}_{new} - \mathbf{p}_{cur}) \cdot (\mathbf{p}_{cur} - \mathbf{p}_{prev}) > 0 \tag{3.13}$$

At this stage the first side contour of the grid has been traced. The remaining three side contours are similarly traced, but with a single modification to the process. When beginning an edge trace from $\mathbf{c}_i$, $i \neq 1$, the orientation of the incorrect edge contour at $\mathbf{c}_i$ is known from the previous edge contour trace and can be filtered out. Therefore when $\mathbf{p}_{cur} = \mathbf{c}_i$, $i \neq 1$, Eqn. 3.4 is replaced by

$$\mathbb{S}_e = \mathbb{S}_d \setminus \underset{\mathbf{h}(R,\theta) \in \mathbb{S}_d}{\operatorname{argmin}} (|\theta - \theta_{prev}|) \tag{3.14}$$

Tracing continues until $\mathbf{p}_{cur}$ is within t_6 pixels of $\mathbf{c}_{(i \bmod 4)+1}$.

Column Edge Contours

The process for tracing column contours is very similar to outer edge contour tracing. For columns $k = \{2, 3, \ldots, width - 1\}$, when anchor point $\mathbf{p}_{cur}$ is the

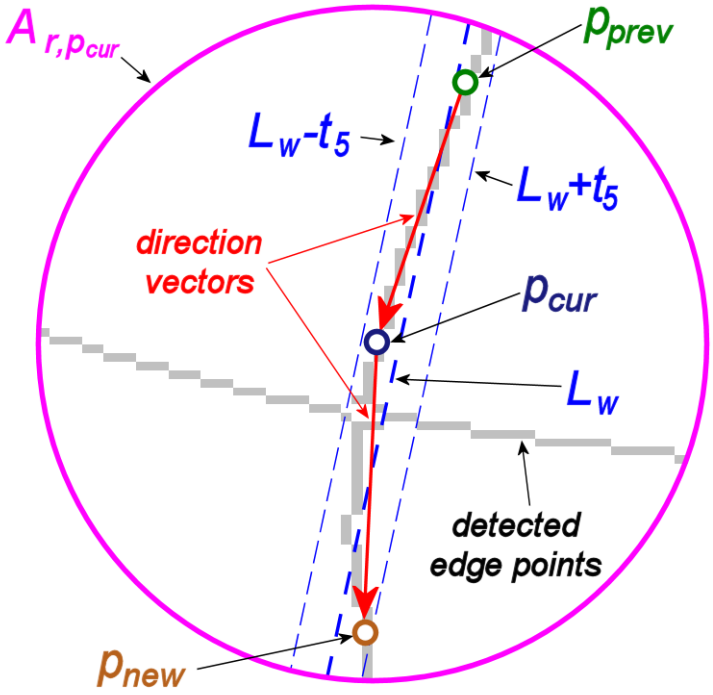


Fig. 3.2: Configuration for edge contour tracing when $\mathbf{p}_{cur} \neq \mathbf{c}_i$.

start point of a column, the orientation of the edge contour at the top of the column is approximated by θ_{top} as

$$\theta_{top} = \angle\left(\mathbf{m}_{1,k+1} - \mathbf{m}_{1,k-1}\right) \tag{3.15}$$

where $\mathbf{m}_{j,k} \in \mathbb{M}_f$ is the grid corner on the jth row and kth column of the grid, and $\mathbb{M}_f$ is the set of processed grid corners. This orientation is used to filter out the Hough point corresponding to the edge contour at the top of the column by replacing Eqn. 3.4 with

$$\mathbb{S}_e = \mathbb{S}_d \setminus \underset{\mathbf{h}(R,\theta)\in\mathbb{S}_d}{\operatorname{argmin}}\left(|\theta - \theta_{top}|\right) \tag{3.16}$$

The vector from the top corner to the bottom corner of the column contour being traced is given by

$$\mathbf{v}_c = \mathbf{m}_{height,k} - \mathbf{m}_{1,k} \tag{3.17}$$

and the orientation of this vector is used to select the best Hough point from $\mathbb{S}_e$ by replacing Eqn. 3.5 with

$$\mathbf{h}_p(R,\theta) = \underset{\mathbf{h}(R,\theta)\in\mathbb{S}_e}{\operatorname{argmin}}\left(|\theta - \angle\mathbf{v}_c|\right) \tag{3.18}$$

Eqn. 3.10 is replaced by

$$(\mathbf{p}_{new} - \mathbf{p}_{cur}) \cdot \mathbf{v}_c > 0 \tag{3.19}$$

which ensures that successive anchor points move down the columns.

When $\mathbf{p}_{cur}$ is not a column start point the contour tracing proceeds in the same way as for the outer edge contour tracing, with Eqns. 3.4 and 3.5 being replaced by Eqn. 3.11, and Eqns. 3.9 and 3.10 being replaced by Eqns. 3.12 and Eqn. 3.13, respectively. Column contour tracing continues until $\mathbf{p}_{cur}$ is within t_6 pixels of $\mathbf{m}_{height,k}$. The complete column tracing process is then repeated until the last internal column, $k = width - 1$, has been traced.

Corner Processing

After the completion of outer edge contour tracing and after the completion of each column edge contour trace the set of candidate corners $\mathbb{M}$ is processed to remove outliers and to merge candidates associated with the same corner. The set of inlier candidate corners, $\mathbb{M}_{in}$, is determined based on the distance from each corner to the line joining the two nearest edge trace points, $\mathbf{p}_\alpha$ and $\mathbf{p}_\beta$, as

$$\mathbf{p}_\alpha = \underset{\mathbf{p} \in \mathbb{P}}{\operatorname{argmin}}(\mathrm{d}(\mathbf{p}, \mathbf{n}_q)) \tag{3.20}$$

$$\mathbf{p}_\beta = \underset{\mathbf{p} \in (\mathbb{P} \setminus \mathbf{p}_\alpha)}{\operatorname{argmin}}(\mathrm{d}(\mathbf{p}, \mathbf{n}_q)) \tag{3.21}$$

$$\mathbb{M}_{in} = \{\mathbf{n}_q \in \mathbb{M} | \perp\mathrm{d}(\mathbf{n}_q, \mathbf{p}_\alpha \times \mathbf{p}_\beta) < t_7\} \tag{3.22}$$

Corners in $\mathbb{M}_{in}$ within a distance t_8 of each other are averaged to get the final set of ordered corners, $\mathbb{M}_f$.

An overview of the complete CELECT algorithm is presented in the flowchart in Fig. 3.3.

CELECT Algorithm Properties

With the exception of t_4, all the algorithm parameters are data driven. Their values are derived as functions of the user selected exterior corner points, $\mathbf{c}_i$, and of the number of row and column corners to be extracted. The functional

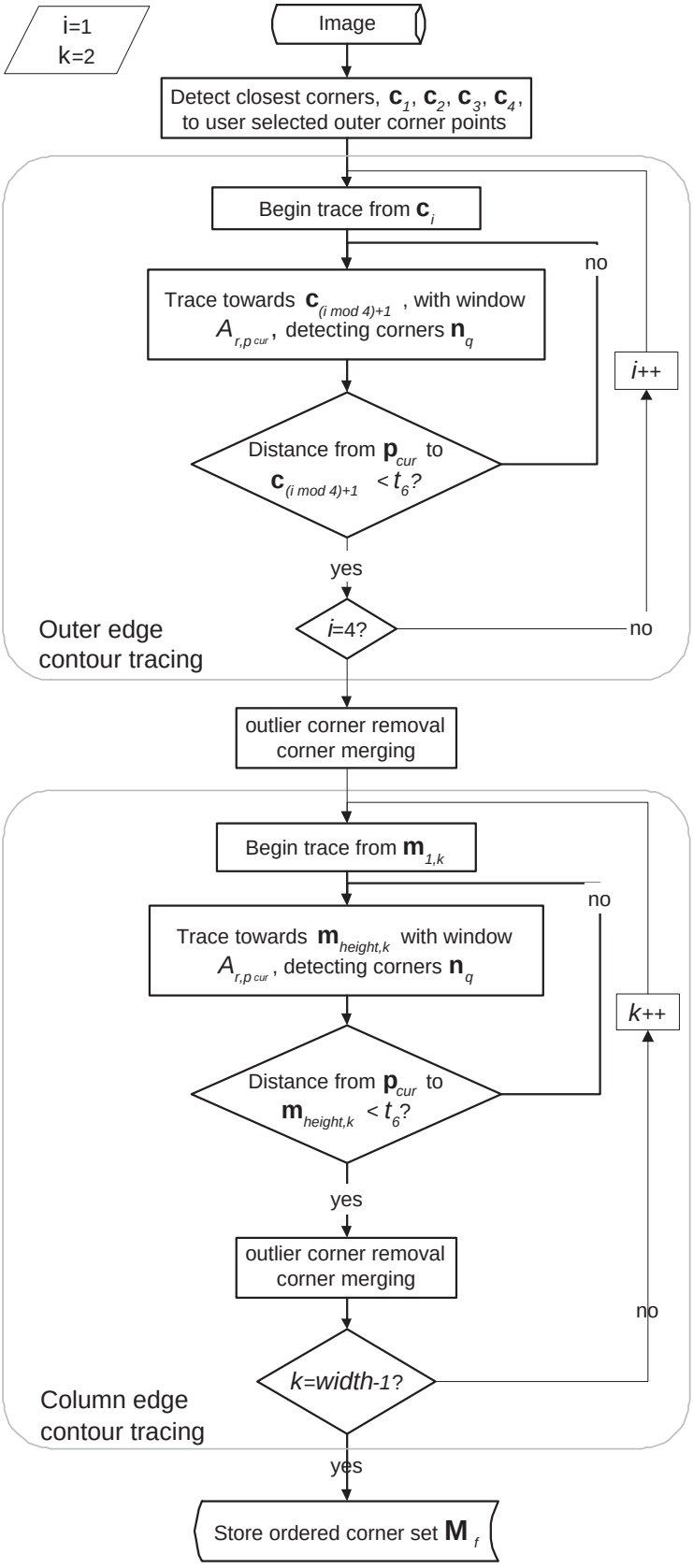


Fig. 3.3: CELECT algorithm overview.

relationships were determined heuristically by applying the algorithm to a large set of images with varying degrees and types of distortion. Parameter values are shown in Table 3.2.

The selection of the window radius parameter, r, is critical to achieving robust performance, as it controls the length of curved contour that is mapped to Hough space. If r is too large the peak in Hough space associated with the curved edge contour will be smeared and reduced in magnitude, and thus may be missed when applying peak detection and non-maximal suppression. On the other hand if r is too small, insufficient pixels are available to form strong peaks in Hough space, and the Hough space is more easily corrupted by noise and image blur. The solution when tracing the outer edge contour is to choose $r = s_i$, where s_i is a function of the size of the average square length for side i, as shown in Table 3.1. For column edge contour tracing r is chosen as 0.5 times the mean square side length of the previously traced column, for which the corners and hence square side lengths are known. Robustness to broken or noisy edge contours is boosted by allowing r to iteratively increase in size, by 20% at each iteration, when no valid $\mathbf{p}_{next}$ can be identified in the current $A_{r,\mathbf{p}_{cur}}$. As the window size increases, more pixels from the broken or noisy line segment will be mapped to Hough space, and so a valid $\mathbf{p}_{next}$ and a valid L_w are more likely to be found.

When applying non-maximal suppression in Hough space a minimum of two detected peaks is enforced. This is to ensure that both contours at a junction are considered. As a result, if there is only one contour in the windowed region the single correct Hough peak may be detected as two separated peaks after non-maximal suppression. An averaging stage is applied to reduce this error due to split peaks, where points in Hough space within t_1 bins of each other are averaged to form a single peak. Non-maximal suppression is applied to the mapped Hough space that ranges from 1^o to 180^o. Phase wrap at the $180^o/1^o$ boundary must be addressed, and this is achieved by an additional stage of non-maximal suppression that is applied to remapped Hough space for Hough points within t_2^o of 180^o. In this way Hough peaks that are detected as two peaks in the initial non-maximal suppression stage are detected correctly as a single peak in the second stage.

In the absence of noise there are four possible directed contours at each grid corner, consisting of opposite directions along each of two windowed lines, L_w.

Table 3.1: Parameter values for determining window radius r.

Parameter	s_1	s_2	s_3	s_4
Value	$\frac{\|\|\mathbf{c}_1-\mathbf{c}_2\|\|}{1.9\times width}$	$\frac{\|\|\mathbf{c}_2-\mathbf{c}_3\|\|}{1.9\times height}$	$\frac{\|\|\mathbf{c}_3-\mathbf{c}_4\|\|}{1.9\times width}$	$\frac{\|\|\mathbf{c}_4-\mathbf{c}_1\|\|}{1.9\times height}$

Table 3.2: CELECT algorithm parameter values.

Parameter	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8
Value	$0.5s_i$	$0.125s_i$	$0.4s_i$	15^o	$0.1s_i$	$0.56s_i$	$0.4s_i$	$0.8s_i$

For outer edge contour traces beginning at $\mathbf{c}_i$, $i \neq 1$, the incorrect contour is available, and this is used to filter out incorrect edges. When beginning the contour trace at $\mathbf{c}_1$, the fixed t_4 angular threshold is set to filter out the incorrect L_w from further consideration. The correct L_w is then more reliably detected, even when side line L_1 has lower angular separation from the incorrect edge contour than from the correct edge contour. A fixed threshold must be used since no information is available at this point about the orientation of the incorrect L_w.

Parameter t_8 determines the minimum distance that candidate corners must be separated by in order to be considered distinct corners. The equivalent merging parameter for merging corners detected in a column trace is 0.47 times the minimum distance between corners in the previous column. The Harris corner detector parameters are selected so as to detect too many rather than too few corners, since erroneous corners are removed during the corner removal and merging stages.

If the detection and ordering fails at any point, then the algorithm does not automatically return any corners. Graceful degradation is incorporated such that the user is prompted to manually select corners along an edge contour if the number of automatically ordered corners along that contour is different to the user specified number.

The CELECT algorithm design and the requirement of the user to select the four outer corner points $\mathbf{c}_i$ enable corners to be detected and ordered in partial grids in an image. This is an advantage for images of calibration grids in which regions of the grid are either too small to be reliably localised or are out of focus, and therefore should not be detected. Corner detection methods that require a white boundary of one square width around the grid to be extracted,

such as the OpenCV method and the method of Rufli et al. (2008), cannot be directly applied to such an image, since the white border is required to enable the grid corners to be detected automatically. Also, the CELECT algorithm is not affected by background artifacts in the image that appear similar in structure to chessboard grids. Since a white boundary is not required, the CELECT algorithm can extract both more corners from a given grid area visible in an image, and can extract grid corners closer to the image edge, than can those methods that require the white boundary. This can be important for calibration since for many non-conventional cameras the largest distortion is present at the image periphery, and this distortion should be captured in order to fully and accurately calibrate the camera. Considering a calibration grid with grid square side length δmm and with overall physical size $\lambda\delta mm \times \kappa\delta mm$, the CELECT algorithm can extract $(\lambda - 1)\times(\kappa - 1)$ grid corners, and methods that require a white border surrounding the grid can extract $(\lambda - 3) \times (\kappa - 3)$ corners. The CELECT algorithm can thus extract $2\lambda + 2\kappa - 8$ more corners for the fixed calibration grid size. $\lambda = 10$ and $\kappa = 8$ for the calibration grid used in the majority of experiments in Section 3.3.2, so for that sized grid the CELECT algorithm can extract 28 additional corners, or 80% more corners, than the methods requiring a white boundary. Of course, to get this improvement the chessboard pattern must extend to the edge of the calibration grid.

Prerequisites and Limitations

A region of at least half a window size is required around the outermost detected contour within which the only edges are grid edges. This is already guaranteed to be present if the grid corners to be detected are all interior grid corners and the whole grid is visible. A white border around the grid is not required.

Contour tracing from $\mathbf{c}_1$ will be incorrect in the CELECT algorithm when the orientation of the correct contour is within t_4 of L_4, or when the incorrect contour has orientation greater than t_4 from L_4 and has orientation closer to L_1 than the correct contour. Consequently, the user selected corner points $\mathbf{c}_i$ should be chosen such that the edge contour between $\mathbf{c}_4$ and $\mathbf{c}_1$ is the straightest edge contour of the grid.

Large grid square size in conjunction with large edge contour curvature can result in failure of the CELECT algorithm. In this scenario, the window size is large and thus the windowed portion of the grid edge contour can be significantly curved resulting in smeared peaks in Hough space. The windowed curved contour can not be accurately approximated by a straight line segment, and so contour tracing is likely to fail. Grids with small square side lengths should be used for highly distorted images to prevent this failure mode. Failure due to unsuitable window size is also possible for images with large variations in grid square size along a grid outer edge contour. This can occur in images with large perspective distortion, and in these cases the CELECT algorithm can break down due to distinct grid corners being incorrectly merged together by parameter t_8.

As with any images of grids for calibration purposes, high resolution images with minimum blur should be used in order to achieve the best corner detection and ordering performance with the CELECT algorithm.

Fig. 3.4 illustrates the contour tracing and corner detection stages of the CELECT algorithm when applied to a real chessboard grid image containing distortion.

3.1.3 Localisation

Once corners have been detected in the calibration grid image, typically their locations are determined to a subpixel level using a refinement technique.

Two common methods of localisation of grid corners are saddle point methods and derivative methods. In the saddle point method, first proposed by Lucchese and Mitra (2002), a surface is fit to the image intensity profile at a blurred and windowed corner. The surface has a characteristic saddle shape, and the subpixel corner location is given by the stationary point of inflexion of the saddle surface. The derivative method is based on edge intersections at corners. Intensity derivatives in a region of interest around the initial corner estimate are calculated, and these are non-linearly fit to the surface model of the photonic response at an edge crossing point proposed by Li and Lavest (1996).

As noted by Mei and Rives (2007), corner localisation is not significantly af-

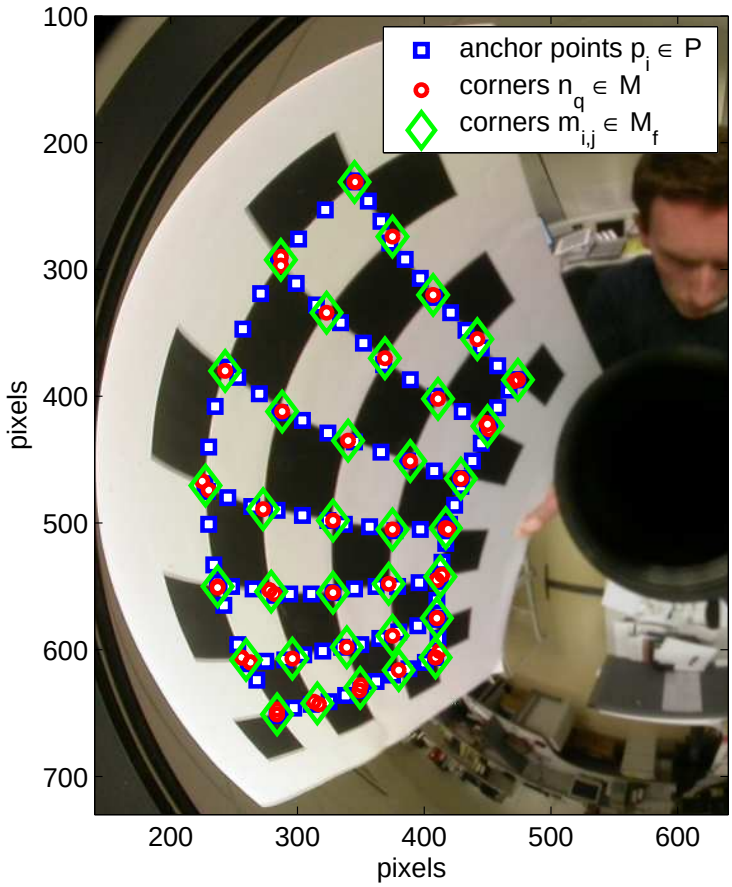


Fig. 3.4: The CELECT algorithm applied to a calibration grid image from an omnidirectional camera. Note the initial multiply detected corners and their merging in the final corner set. For clarity purposes the ordering is not illustrated.

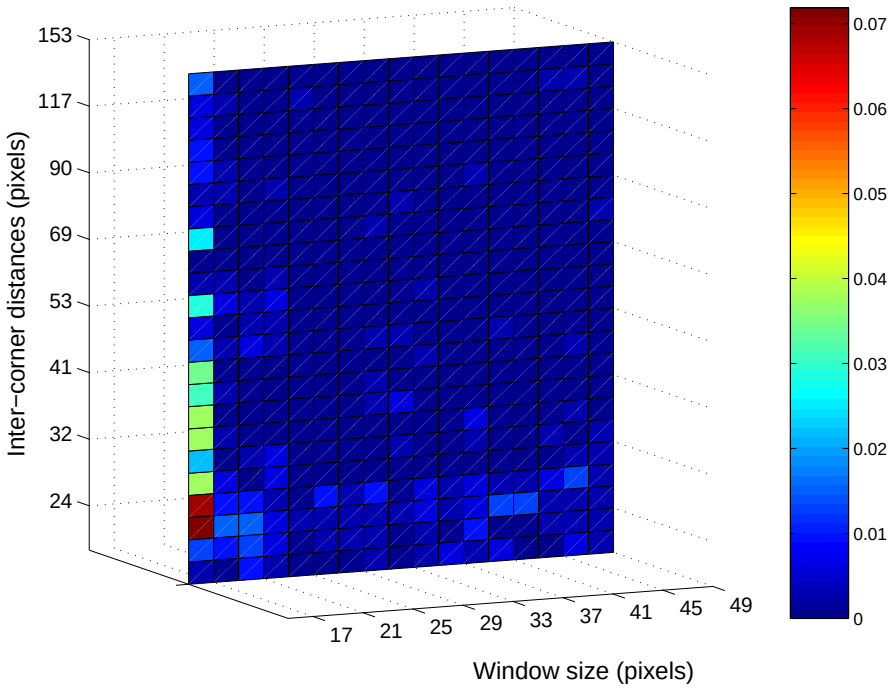
fected by image distortion since it typically operates within small regions of interest around a corner, and thus any localisation technique is applicable when the corners have been detected. Nevertheless, further investigation of the application of the localisation method is necessary so that the method's parameters can be selected to achieve maximum accuracy. The saddle point method takes two principal parameters – the window size that determines the scale of the region of interest, and the level of blur that is applied before surface fitting. The window size is the principal parameter required for the derivative method. A question arises as to the relationship between these parameter values and the distances between corners in the images. The answer to this question would provide a guide to the parameter selection that achieves optimum localisation results, and more importantly would allow the automatic selection of such parameters on a per-corner basis using the result

of the CELECT method. For distortion free images, the range of inter-corner distance values varies only with changes in perspective distortion. However, for images with large non-linear distortion, these distances can vary significantly across the image even when there is no perspective distortion present. Tests to examine the relationship were conducted by applying each localisation method to each of four 1000×1000 pixel synthetic chessboard grid images with fixed fisheye distortion, with Gaussian blur of 2 pixels SD, and with additive Gaussian noise of 4 pixels SD. The four images had chessboard grid square sizes of 50, 100, 150 and 200 pixels, respectively, before the application of fisheye distortion. Initial corner estimates for the localisations were formed by corrupting the exact corner locations with Gaussian noise of 1 pixel SD. Each localisation method was applied to each image for a range of window size and blur SD parameters in order to generate the results shown in Fig. 3.5. The results are presented in histogram format, where, for every corner, the window size producing the smallest residual was binned and accumulated for the derivative localisation, and the combination of window size and blur SD producing the smallest residual was binned and accumulated for the saddle point localisation. For derivative localisation it is seen that small windows produce the best results across all inter-corner distances, although for the smallest inter-corner distances larger window sizes also perform well. Considering the saddle point localisation results, it is seen that small window sizes are again preferred but that otherwise the window size parameter is essentially independent of the inter-corner distances. Despite evidence of a weak relationship between inter-corner distances and blur SD, a mid-range blur SD parameter produces close to optimum results for all inter-corner distances.

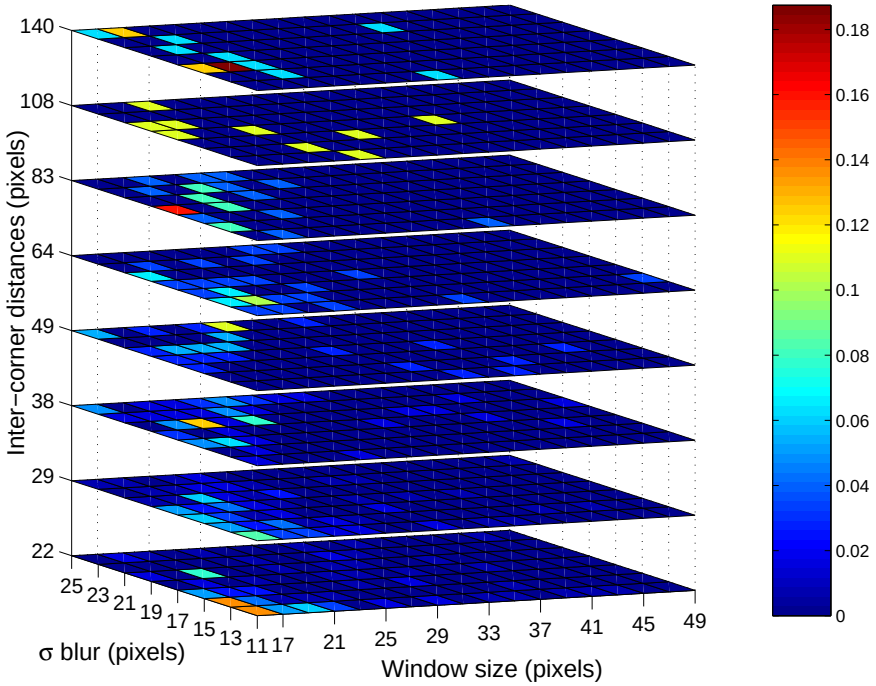
Overall, the results show that there is no meaningful relationship between variations in parameter values and variations in inter-corner distances. Consequently, there is no benefit to using the CELECT algorithm results, in the form of the inter-corner distances, to inform the selection of localisation parameters, and therefore such an approach is not further pursued.

3.2 Coded Calibration Grids

Typically camera calibration performance can be improved by increasing the density of the extracted calibration grid features. In the case of perspective



(a)



(b)

Fig. 3.5: Visualisation of the relationship between localisation parameters and inter-corner distances for (a) derivative localisation and (b) saddle point localisation. The histogram frequencies, which are indicated by colour hue, are normalised for each slice.

camera calibration using methods similar to that of Zhang (1998), improvements are achieved as a result of the increasing number of linear equations, which reduces the effect of random measurement noise on the result. Chapter 2 details an alternative calibration method, generic calibration, that is applicable to general cameras. Increasing the feature point density in generic calibration has an important advantage relating to interpolation that does not manifest itself for the perspective calibrations.

3.2.1 Calibration Grid Interpolation Bias

Step 2 of the standard generic calibration method, as outlined in Algorithm 1 in Section 2.4, requires that the intersection point of each camera ray with each calibration grid be determined. Homographic interpolation is employed by Sturm and Ramalingam (2003) to determine these intersection points based on the extracted image coordinates of the four closest grid corner points (with no three points collinear) on a standard calibration grid. However, any distortion present in the images of the calibration grids induces a bias in the results when employing this method. Fig 3.6 shows a vector plot of the error residuals after homographic interpolation is applied to an image of 500 random points on a $300mm \times 300mm$ grid. The image is formed for a simulated camera with focal length $680mm$, and principal point $[150\ 150]^T mm$. Radial distortion is incorporated using the model of Eqn. 2.4 with $k_1 = -0.8$ and $k_2 = 0.4$. The systematic bias in the plot increases with distance from the image centre, indicating that it is primarily a result of the radial distortion. The non-radial components of the residual vectors are due to the grid points used in interpolation, since for accurate interpolation near to a grid square edge, the grid corners used in the interpolation form a parallelogram rather than a square. Although interpolation bias can be reduced by decreasing the square sizes of the calibration grids, as seen in the simulation result shown in Fig. 3.7 where the camera parameters from above are used, this approach is limited by the limits of camera resolutions. Interpolation bias can also be ameliorated by applying collinearity constraints (Ramalingam et al., 2005*a*). Homographic interpolation has the additional disadvantage of requiring local image continuity. The general camera model makes no continuity assumptions, and thus it can model discontinuous cameras. By using calibration grids that require interpolation, the accuracy of generic calibration for such cameras is reduced.

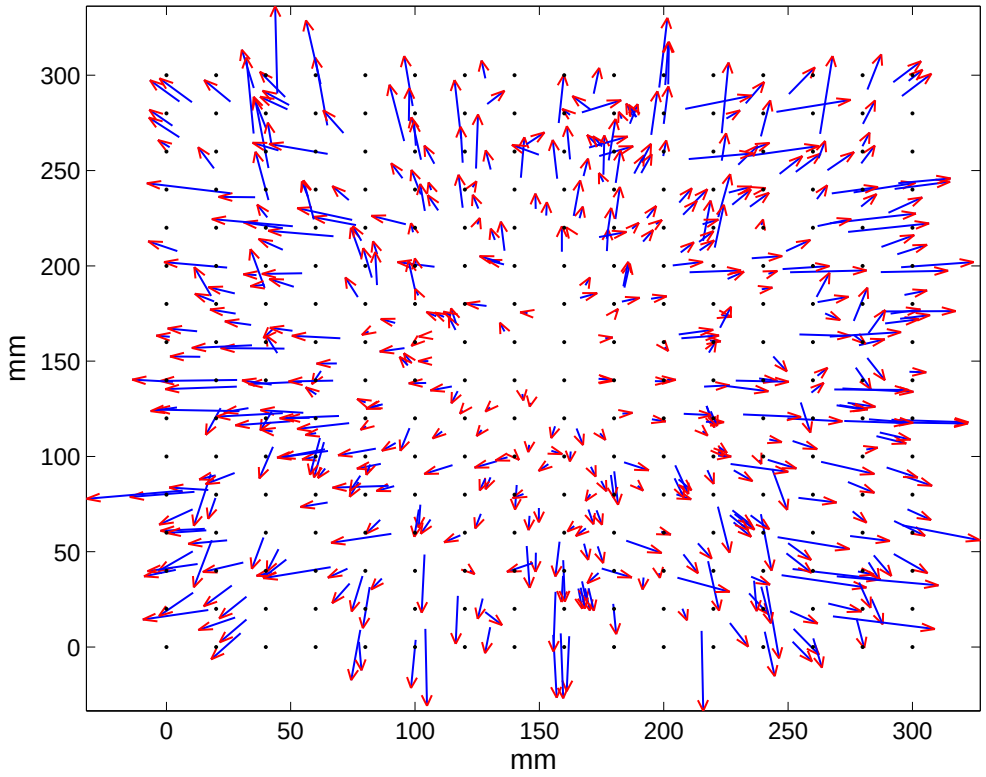


Fig. 3.6: Vector plot of error residuals for homographic interpolation (20mm grid pitch) showing interpolation bias. Vectors are scaled $\times 300$.

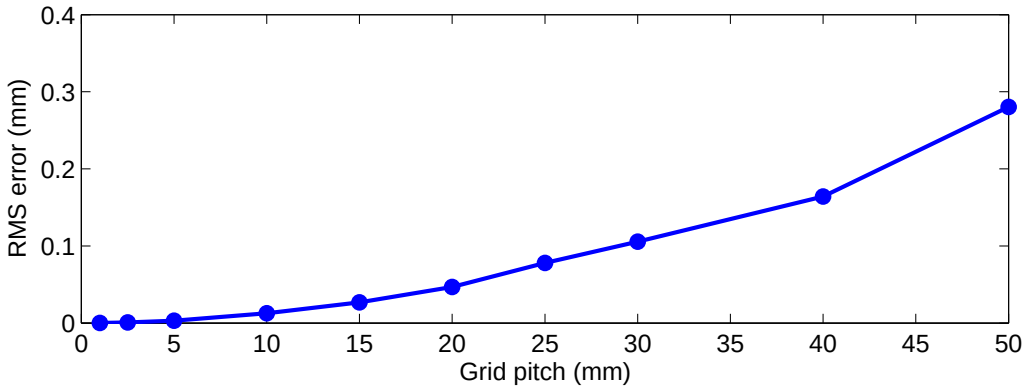


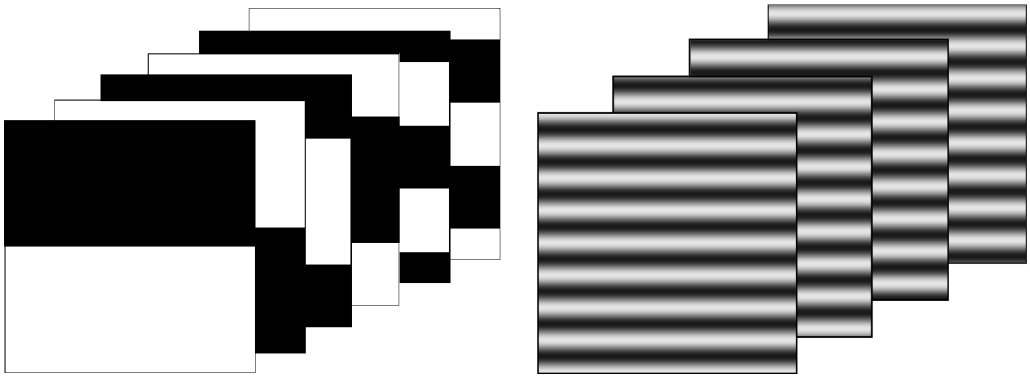
Fig. 3.7: RMS localisation error after homographic interpolation for a chessboard grid with increasing grid pitch.

A method of more accurately determining the camera ray intersections with the calibration grids is likely to lead to improved results for generic calibration.

3.2.2 Spatio-Temporal Coding – Active Grids

The use of active grids overcomes the disadvantages associated with homographic interpolation by providing a direct localisation of the point seen by *every* pixel viewing the active grid, thus enabling pixel-level generic calibration. An active grid is implemented using a flat-screen TFT monitor that displays a temporal sequence of spatially varying greyscale patterns. The location (l_x, l_y) of any point on the active grid can be decoded from the intensity displayed at that point across the sequence of patterns. Patterns from the domain of structured light are used to encode location in the active grids described in this thesis. Structured light techniques are typically used to recover shape by triangulation between a projector, which projects light patterns onto the scene, and one or more calibrated cameras, which view the illuminated scene. See Salvi et al. (2004) for a comprehensive survey of coded structured light techniques. The use of active grids in calibration has previously been employed by Sagawa et al. (2005) for distortion correction. However, due to the binary striped patterns that they use, their approach still requires interpolation in regions where the stripes appear very narrow in the images. In the method used here, each active grid requires 20 patterns to be consecutively displayed in order to fully encode the location data. 12 patterns encode location on an 8×8 grid using Gray coded binary patterns, and 8 sinusoidal greyscale patterns encode location spatially within each square of this grid. The Gray codes spatially disambiguate between the locations decoded from the sinusoidal patterns. Gray codes are suitable for encoding binary location data because the white/black boundaries occur at different positions in the different Gray coded patterns, thus limiting to a single pixel decoding errors that can occur at a boundary. Both the binary patterns and their inverses are displayed to make the decoding near white/black boundaries more robust - if the binary pattern at a pixel is lighter than its inverse, a 1 is decoded for that pixel, and vice versa. The phase of the sinusoidal patterns is shifted by 90^o between consecutive patterns. A 7×7 smoothing filter is applied to the sinusoidal images in order to interpolate the data. This process is necessary when the camera resolution is greater than the apparent active grid resolution, and for noise reduction. Fig. 3.8 shows the set of binary and sinusoidal patterns that encode vertical location.

Sublocations, d, within the 8×8 grid squares are decoded from the active grid



(a) (b)

Fig. 3.8: (a) Binary and (b) sinusoidal active grid patterns for encoding vertical location.

images (Salvi et al., 2004) as

$$d = \frac{1}{2\pi} \arctan\left(\frac{I_1 - I_3}{I_2 - I_4}\right) \tag{3.23}$$

where I_i is the intensity of sinusoidal pattern i. Eqn. 3.23 does not take account of the possibly non-linear camera and TFT monitor radiometric transfer functions, and therefore it is necessary to investigate the effect of such non-linearity on the accuracy of the decoded sublocations. An image of one of the active grid sinusoidal patterns, taken by a calibrated perspective camera, was undistorted, and the image was rectified such that the grid appeared square. The mean greylevels in each image column were determined, giving an approximately sine wave shaped greylevel intensity profile with 8 periods. A single period of this intensity profile is shown in Fig. 3.9. Although the intensity profile exhibits a greylevel offset, the offset reduces to 0 in Eqn. 3.23, and so does not affect the decoded sublocation values. However, it is clear from Fig. 3.9 that the shape of the intensity profile does not agree well with the best fit sine (RMS error of fit = 9.839 greylevels). This is due to the non-linearity of the product of the monitor and camera radiometric transfer functions. Referring to Fig. 3.9, it is seen that the intensity profile is well modelled by a Fourier series with a single harmonic (RMS error of fit = 1.309 greylevels), and that the inclusion of an additional harmonic in the series provides only a marginal improvement in the fitness of the model (RMS error of fit = 1.308 greylevels). It can be shown (see Appendix A) that Eqn. 3.23 is unaffected by odd harmonics in the model of the intensity profile, and consequently Eqn. 3.23 is valid for intensity profiles that are well modelled by the sum of the fundamental and the first harmonic. Therefore Eqn. 3.23 is robust to monitor-camera radiometric

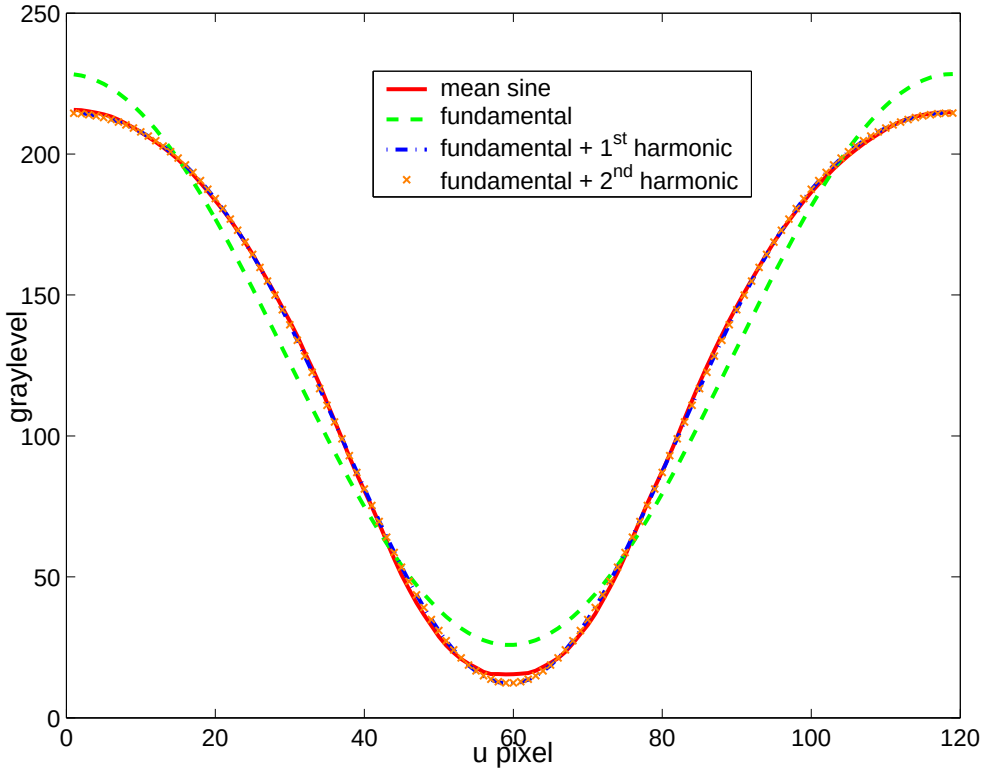


Fig. 3.9: Mean greylevel intensities for a single period of a real image of the horizontal sinusoidal active grid pattern, and best fit sinusoidal models with zero, one and two harmonics.

transfer function non-linearities of the type that are encountered in the experiments in this thesis. Sublocation decoding is also robust to camera vignetting since Eqn. 3.23 is applied independently at each pixel where vignetting has only a constant multiplicative effect on the greylevel intensity.

The raw decoded active grid horizontal location, l_x, is given by

$$l_x = (b + d)w \tag{3.24}$$

where b is the decoded Gray code value, and w is the spatial resolution of the binary patterns. The vertical location, l_y, is calculated similarly. 8-connected single pass smoothing filtering is applied to the raw location data, followed by a final infilling pass to interpolate missing pixel locations from their 8-connected neighbours. The kernels of these filters are as small as possible to minimise the requirements on image continuity and to reduce interpolation errors.

The sinusoidal patterns displayed as part of the active grid are inherently robust to errors induced by sensor blooming. Blooming occurs in camera sensor arrays for saturated regions in images when the photons from the saturated re-

gion spill into neighbouring sensor elements. Images of chessboard calibration grids contain regions of high contrast at grid edges, which, due to blooming, can cause the grid corners to be incorrectly localised. This problem is largely overcome in active grids since the highest resolution data is extracted from the sinusoidal patterns, which locally contain only low contrast. Blooming can still occur at the boundaries in the binary images, although these edges have low spatial frequencies. The effects of blooming on the Gray code decoding are reduced by performing decoding based on the comparison between the binary and inverse binary patterns. Error due to blooming can be further reduced by localising the binary pattern edges as the midpoint of the edges detected in the binary and the inverse binary images.

The implementation of active grids presented here is similar to that proposed by Bonfire et al. (2006), though they use an energy minimisation approach instead of post-filtering the data directly. The method is also similar to that of Scharstein and Szeliski (2003), where Gray coded or sinusoidal patterns are projected onto a scene in order to determine high accuracy correspondences between stereo image pairs for stereo algorithm evaluation. They propose using either Gray codes or sinusoidal patterns of different frequencies with phase shifts, although they require many more images due to variations in surface albedo of scene points.

3.3 Experiments

Initial evaluation of both the CELECT algorithm and of active grids was conducted for variations in image parameters and calibration grid configurations. The CELECT algorithm was tested on synthetically generated images of chessboard grids so that a ground truth for the corner locations was available. Simulations were conducted to determine the robustness of the algorithm to variations in image noise, image blur, grid square size, fisheye distortion and perspective distortion. Simulated data for characterising active grids is not used as errors in the configuration, for example due to non-linearity of the camera radiometric transfer function, would be difficult to emulate synthetically. Therefore, characterisation of active grids is performed on real data. The CELECT algorithm and active grids are subsequently evaluated on real images, from both central and non-central cameras, in order to verify their

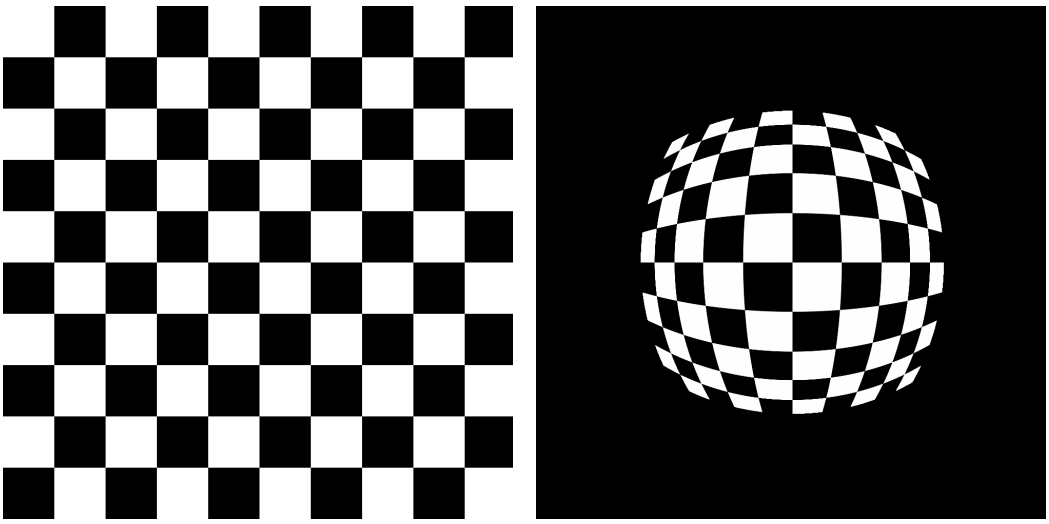
performances.

3.3.1 CELECT Algorithm Robustness

A synthetic chessboard grid image of size 1000×1000 pixels was generated, and fisheye distortion was applied subsequently using the equisolid angle projection function of Eqn. 2.7. Increasing fisheye distortion corresponds to increasing FOV, and is controlled by reducing the focal length. FOV is a more intuitive measure of distortion than focal length, and it will be used to quantify distortion here. The distortion-free synthetic image, and a corresponding fisheye distorted synthetic image with FOV=134^o, are shown in Fig. 3.10. Since the CELECT method is designed for application to distorted images, the synthetic experiments are all conducted with images containing fisheye distortion similar to that of the distorted grid in Fig. 3.10. Where not otherwise specified, Gaussian blur with 1 pixel SD was applied to the synthetic images used in the experiments. The purpose of this is to better approximate real camera images by reducing quantisation artifacts resulting from the manner in which the synthetic images are generated.

The user selected corner points $\mathbf{c}_i$ are refined by applying a Harris corner detector, as discussed in Section 3.1.2, in order to make the algorithm robust to inaccuracies in the user selection of the four outer grid corner points. However, variation in these selected points can result in small variations in the exact locations of the corresponding Harris detected corners, which causes slightly different anchor points, $\mathbf{p}_i$, to be used in the algorithm. Therefore, Gaussian noise with a standard deviation of 2 pixels, and with an upper limit of 4 pixels, was added to the true corner locations in the synthetic tests to simulate user inaccuracy in point selection. The measured corner selection error for 2880 manually selected corners used in Section 3.3.2 had Rayleigh mean and standard deviation of 0.77 pixels and 0.16 pixels respectively.

Perspective distortion is a linear distortion that results in variation of the size of a planar chessboard grid across the image when the grid is oriented in a non-fronto-parallel position. Many perspective camera calibration algorithms require that a calibration grid be imaged in varying poses to avoid singularities (Sturm and Ramalingam, 2003, Zhang, 2000), and thus it is important that the CELECT algorithm can operate correctly on such images. Perspective



(a) (b)

Fig. 3.10: (a) Distortion-free synthetic image; (b) synthetic image after applying fisheye distortion corresponding to an FOV of 134^o.

distortion is applied to the synthetic images by applying yaw rotations to a grid in $\mathbb{R}^3$, followed by projection of the grid points to $\mathbb{P}^2$ using the camera projection matrix P formed from the synthetic camera parameters $\{f, u, v\}$. Fisheye distortion is subsequently applied to the image.

The experimental results for synthetic images containing varying fisheye distortion, varying additive noise, varying blur, varying perspective distortion and varying grid square size are shown in Fig. 3.11(a)-(e). Each data point in these plots is the mean value of 50 trials. Corners detected with greater than 8 pixels error are classified as detection failures, while the CELECT algorithm as a whole is considered to have failed a trial if there is a detection failure at any one of the corners in the test image. The maximum detection errors shown are compiled over the successfully detected grids only.

The results of the synthetic experiments indicate that the CELECT algorithm is robust to variations in additive Gaussian noise, in scale, in fisheye distortion, and in perspective distortion, as evidenced by the maximum detection errors of less than 6.5 pixels for these tests. The algorithm fails for additive noise with SD of 26 greylevels, which far exceeds the level of noise that is likely to be present in most calibration images. Perspective distortion caused by out-of-plane rotations of more than 48^o results in algorithm failure due to the large compression of parts of the grid in the image. The failure of the algorithm with

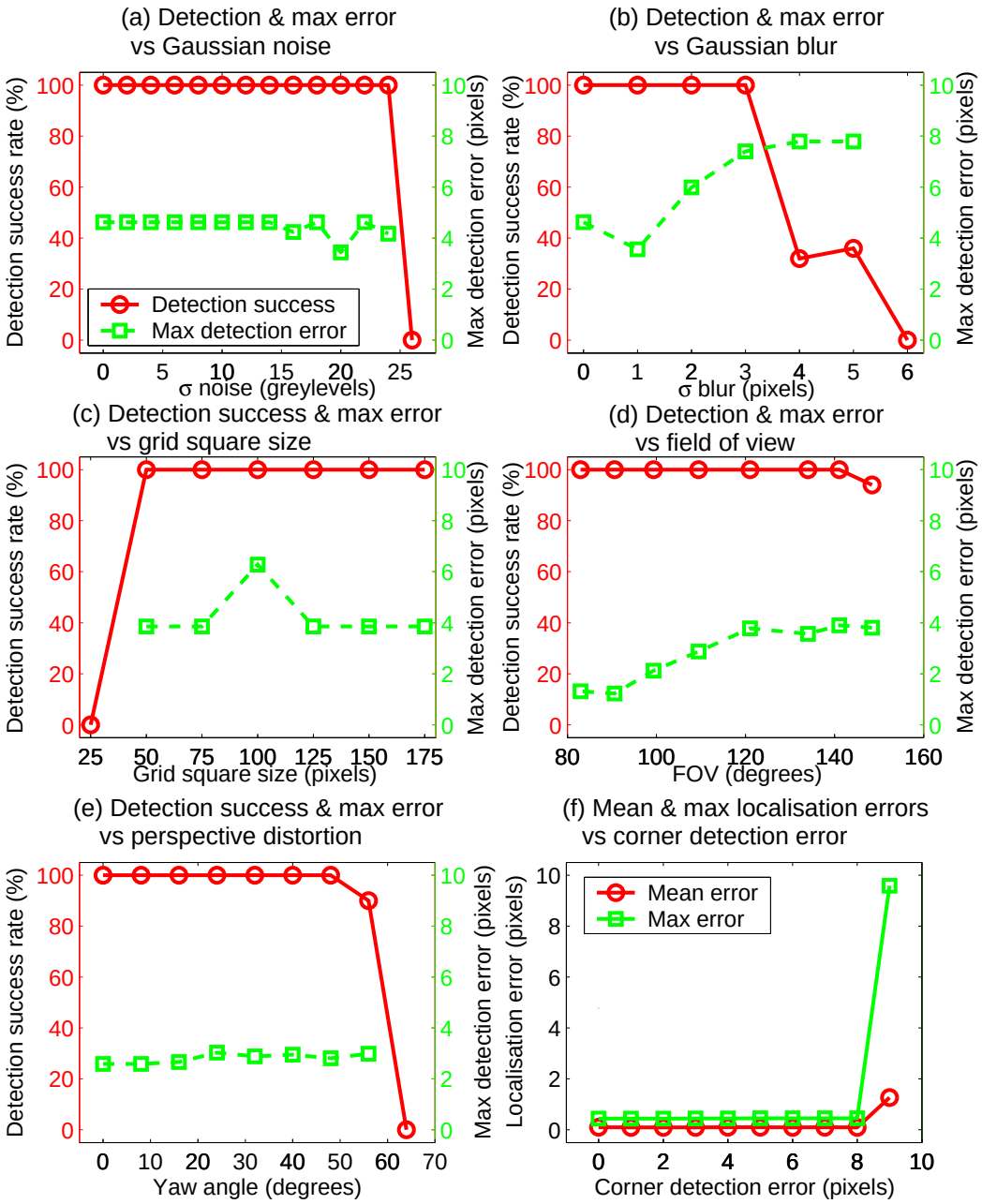


Fig. 3.11: Performance of the CELECT algorithm on synthetic images for variations in (a) additive Gaussian noise; (b) Gaussian blur; (c) grid square size; (d) fisheye distortion; and (e) perspective distortion. (f) Localisation accuracy for varying corner detection error.

decreasing grid square size is a result of inaccurate edge and corner detection due to the small size of the window and the reduced number of edge points accumulated in Hough space. Nevertheless, the CELECT algorithm can successfully detect and order all corners in the grid images with 50 pixels pitch, despite the corresponding inter-corner distances in the distorted image being as small as 12 pixels. For large fisheye distortions the algorithm fails due to the high curvature of the distorted grid edges near the grid boundary: the orientation of the grid edges varies significantly in these areas, and so the comparison of edge orientations in these adjacent windowed regions is unpredictable. The algorithm performance deteriorates when the image blur increases to 4 pixels SD. This level of blur is relatively low, although in practice for calibration image sets, motion blur is minimal and focusing is well controlled during image capture.

A corner localisation stage is generally applied after corner detection to achieve subpixel accuracy for the corner locations, as discussed in 3.1.3. For successful localisation of the detected corners, the error in the initial corner location estimates must be sufficiently small for the localisation to converge to the actual corners. Fig. 3.11(f) shows the error after localisation for initial corner estimates of decreasing accuracy, for the synthetic fisheye image of Fig. 3.10(b). Localisation was performed using the OpenCV function *cvFindCornerSubPix()* with a 9×9 search window. The maximum localisation error over 50 trials is less than 0.46 pixels for a corner detection error at each corner of up to 8 pixels. This corner detection error is the Euclidean distance *each* simulated detected corner is from its ground truth location. Referring to the test results in Fig. 3.11(a)-(e), it is seen that the maximum detection error for successfully detected grids is significantly less than 8 pixels for the majority of the synthetic tests. This shows that the CELECT algorithm is sufficiently accurate for subsequent corner localisation to converge.

3.3.2 CELECT Applied to Real Non-Perspective Images

The CELECT method was experimentally evaluated on a real image dataset against the OpenCV *cvFindChessBoardCornerGuesses()* function, hereafter referred to as the OpenCV method. Four different types of non-conventional

camera were selected, each of which exhibits large distortion. For each camera, 15 images were captured of a calibration grid in various orientations and positions with respect to the camera. Grid locations were chosen so as to obtain a challenging set of images in which the grids appear severely distorted due both to perspective and non-linear distortions. However, in order to enable the application of the OpenCV method, the entire calibration grid and a white border one square width wide around the grid were maintained within the FOV for all images. This limited the extent of grid distortion around the interior corners, since the most severe distortion typically occurs at the image periphery.

Details of the test configuration for each image set are presented in Table 3.3. The grid pitches of the 7×5 calibration grids used for the Spherical and Flexible Mirror image sets and for the Hyperboloidal image set were $30mm$ and $50mm$, respectively. The 9×7 and 12×9 grids had grid pitches of $30mm$ and $15mm$, respectively. Camera configurations for the Spherical and Flexible Mirror image sets are non-central, whereas the Hyperboloidal and Fisheye cameras are approximately central. Several additional properties of the image sets should be noted. For the Spherical image set, the reflective sphere is a gold coloured reflective Christmas decoration that, in conjunction with incandescent light sources, produces a gold hue in the images. In addition, the contrast in these images is low since the grid frequently occluded the primary light source. Several images in the Fisheye set contain weak shadows of the camera due to its proximity to the grid. The Flexible Mirror image set exhibits significant noise and chromatic aberration resulting from the low quality of the camera. Distortion in this image set was induced by directly warping the flexible mirror.

The CELECT method and the OpenCV method were applied to every test image in each image set. The location accuracy of the resulting detected corners was evaluated with respect to manually selected corners that were subpixelly localised using the OpenCV *cvFindCornerSubPix()* function. Ordering success of the detected corners was evaluated qualitatively by visual inspection. For each image set, the mean, maximum and SD of the errors in the corners detected with each detection and ordering method were determined, both before and after localisation with OpenCV's *cvFindCornerSubPix()*. The complete results are presented in Tables 3.4, 3.5, 3.6 and 3.7. Error statistics are compiled over successfully detected and ordered grids only.

Table 3.3: Experimental configurations for the CELECT algorithm evaluation with real images.

Image set	Non-conventional Camera	Resolution (pixels)	Grid size (int. corners)
Spherical	Canon PowerShot SD500 + spherical bauble	1600×1200	7×5
Hyperboloidal	Nikon CoolPix 4500 + 360 OneVR hyperboloidal mirror	1280×960	7×5
Fisheye	Nikon CoolPix 4500 + FC-E8 fisheye converter	1024×768	9×7 12×9
Flexible Mirror	Creative Live! Cam webcam + flexible mirror	640×480	7×5

Table 3.4: Experimental results for the Spherical image set.

Method	# correctly detected and ordered grids	Error before localisation (pixels)			Error after localisation (pixels)		
		Mean	Max	SD	Mean	Max	SD
OpenCV	**3**/15	0.31	1.35	0.24	0.0069	0.0752	0.0096
CELECT	**12**/15	1.54	7.33	0.92	0.0137	0.1985	0.0219

Table 3.5: Experimental results for the Hyperboloidal image set.

Method	# correctly detected and ordered grids	Error before localisation (pixels)			Error after localisation (pixels)		
		Mean	Max	SD	Mean	Max	SD
OpenCV	**6**/15	0.39	1.35	0.24	0.0088	0.0457	0.0093
CELECT	**12**/15	1.44	6.07	0.80	0.0108	0.1584	0.0145

Table 3.6: Experimental results for the Fisheye image set.

Method	# correctly detected and ordered grids	Error before localisation (pixels)			Error after localisation (pixels)		
		Mean	Max	SD	Mean	Max	SD
OpenCV	**8**/15	0.39	1.60	0.26	0.0086	0.0942	0.0111
CELECT	**12**/15	1.27	4.31	0.76	0.0125	0.2160	0.0162

Table 3.7: Experimental results for the Flexible Mirror image set.

Method	# correctly detected and ordered grids	Error before localisation (pixels)			Error after localisation (pixels)		
		Mean	Max	SD	Mean	Max	SD
OpenCV	**9**/15	0.56	1.86	0.33	0.0084	0.0520	0.0110
CELECT	**15**/15	2.17	5.25	0.96	0.0114	0.1341	0.0152

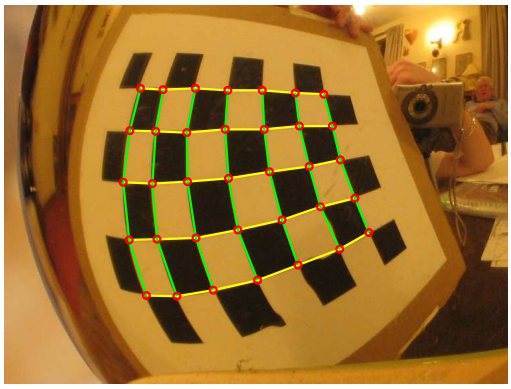
The results clearly demonstrate the superior ability of the CELECT method with respect to the OpenCV method to correctly detect and order grid corners in images containing significant distortion. Out of the total of 60 grid images used in the evaluation, only 2 grids were detected and ordered correctly by the OpenCV method but not by the CELECT method. In contrast, the CELECT method successfully detected and ordered 27 grids for which the OpenCV method failed. The overall success rate of the CELECT method for detecting and ordering the grids for all tested image sets is 85.0%, which is significantly greater than the equivalent rate of 43.3% for the OpenCV method.

The statistics on detection accuracy reveal that the corners detected using the CELECT method are less accurate than for the OpenCV method. However, after the application of localisation to the corners, detection accuracy is within 0.12 pixels of the localisation function's set precision level of 0.1 pixels for all detected corners for each method. Thus the greater corner detection inaccuracy of the CELECT method does not significantly impact the final result after localisation is applied.

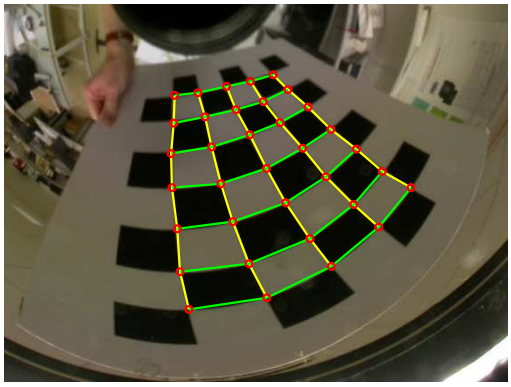
Figs. 3.12 and 3.13 show the results of applying the CELECT method to images from each image set. Images are presented for which both the CELECT and OpenCV methods operate successfully, for which only the CELECT method operates successfully, and for which both methods fail.

For the complete set of 60 grid images, grid contour tracing only fails for 2 grid images. Both of these result from the contours associated with orthogonal grid edges having very similar orientations in the image due to severe fisheye and hyperboloidal distortion. False positive and false negative corner detections are the cause of the remainder of the failures for the CELECT method.

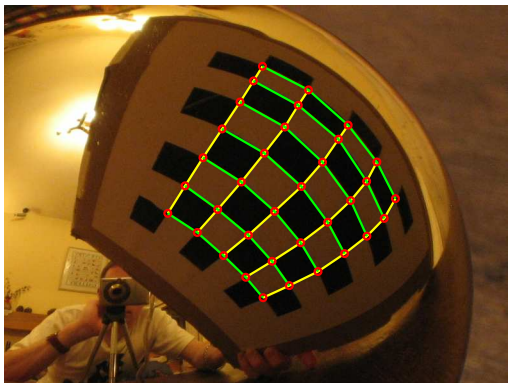
Considering the Spherical image set, two of the failures of the CELECT method are due to image noise resulting from aberrations in the surface of the low-



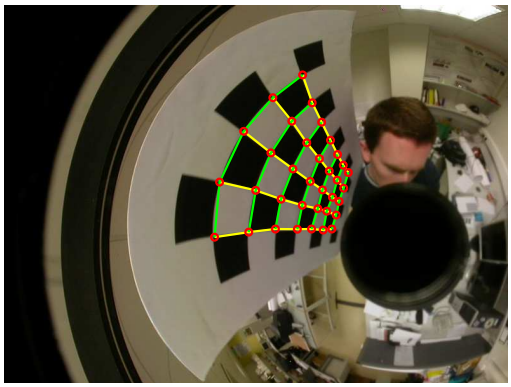
(a) Spherical – CELECT succeeds, OpenCV succeeds



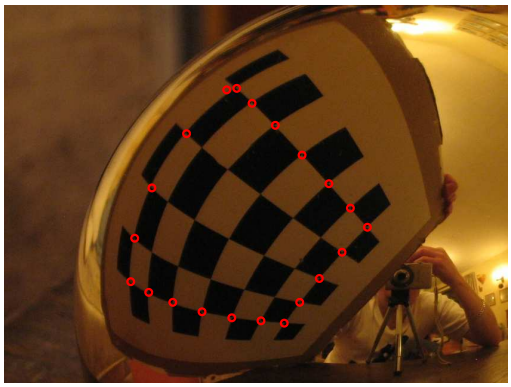
(b) Hyperboloidal – CELECT succeeds, OpenCV succeeds



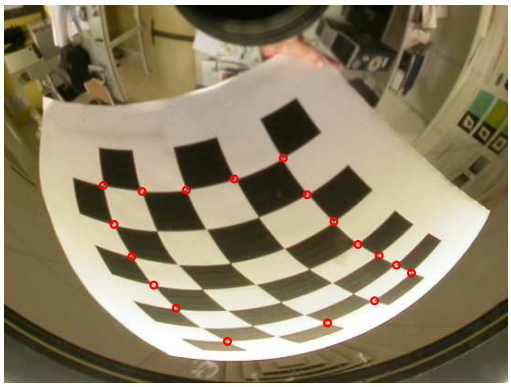
(c) Spherical – CELECT succeeds, OpenCV fails



(d) Hyperboloidal – CELECT succeeds, OpenCV fails

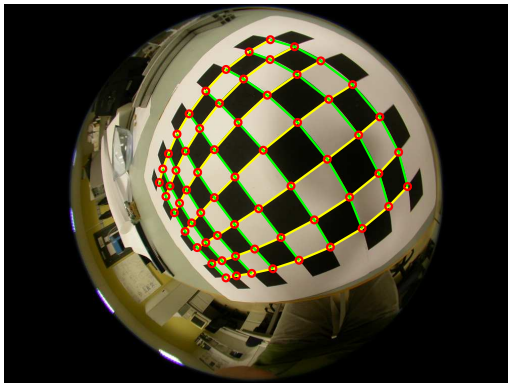


(e) Spherical – CELECT fails, OpenCV fails



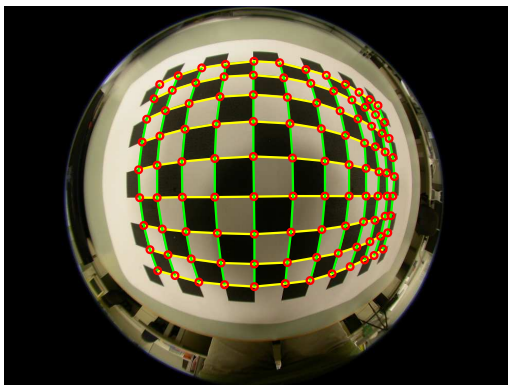
(f) Hyperboloidal – CELECT fails, OpenCV fails

Fig. 3.12: Images from the Spherical and the Hyperboloidal sets for which grids are successfully detected and ordered by the CELECT method and the OpenCV method, by the CELECT method only, and by neither method. Corners detected by the CELECT method are shown with red circles, and corners ordered by the CELECT method are connected by yellow and green lines.

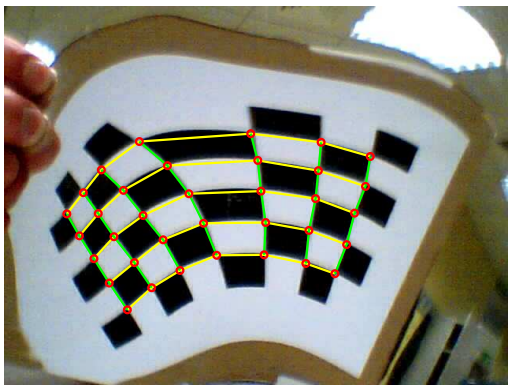


(a) Fisheye – CELECT succeeds, OpenCV succeeds

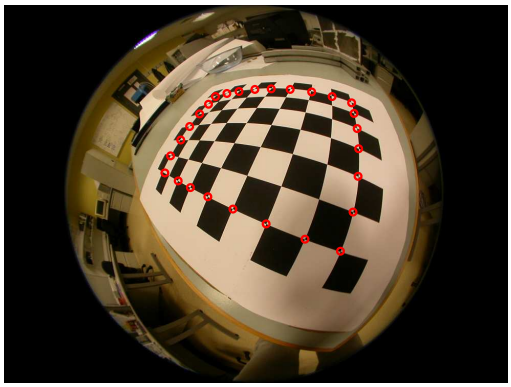
(b) Flexible Mirror – CELECT succeeds, OpenCV succeeds



(c) Fisheye – CELECT succeeds, OpenCV fails



(d) Flexible Mirror – CELECT succeeds, OpenCV fails



(e) Fisheye – CELECT fails, OpenCV fails

Fig. 3.13: Images from the Fisheye and the Flexible Mirror sets for which grids are successfully detected and ordered by the CELECT method and the OpenCV method, and by the CELECT method only. An image for which both methods fail is only shown for the Fisheye image set, as the CELECT method was successful for every image in the Flexible Mirror set. Corners detected by the CELECT method are shown with red circles, and corners ordered by the CELECT method are connected by yellow and green lines.

quality mirror, where the noise is falsely detected as a corner. The remaining failure is the result of a single missed corner. The high sensitivity of the Harris corner detector in the CELECT method resulted in the false detection of lens dirt as a corner in one of the Hyperboloidal images. A further image in the Hyperboloidal set, shown in Fig. 3.12(f), failed when the corner detector missed several corners due to image glare. Images in the Fisheye image set exhibit large variation in grid square size along single edge contours. This weakens the corner merging criteria, since that is based on the mean square side length along a contour. Two Fisheye image failures can be attributed to merging failure. The first failure occurs due to adjacent corners being incorrectly merged together, and the second is the result of a single distorted and blurred corner being detected as two separate corners that are not subsequently merged. Shadow in the Fisheye images does not affect the performance of the CELECT method, but it does cause failure due to missed corners for the OpenCV method.

Successful detection and ordering with the CELECT method was seen to vary slightly depending on the direction of contour tracing resulting from manual selection of the $\mathbf{c}_i$s. An informed selection of the outer corners can aid effectiveness of the CELECT method. For example, the $\mathbf{c}_i$ that is adjacent to the smallest grid square, where image compression is greatest, should be chosen as the first corner. This reduces the likelihood of corners beyond the outer corner being detected during contour tracing.

Failure of the OpenCV method to correctly detect and order the corners of a grid is generally found to be a result of missed corners. When that method detects less than the user defined number of internal corners, even if only a single corner is missed, then the resulting ordering is chaotic and unusable. This highlights the importance of ordering, since even if the subset of correctly detected corners is returned, the information is useless for calibration applications without their associated ordering.

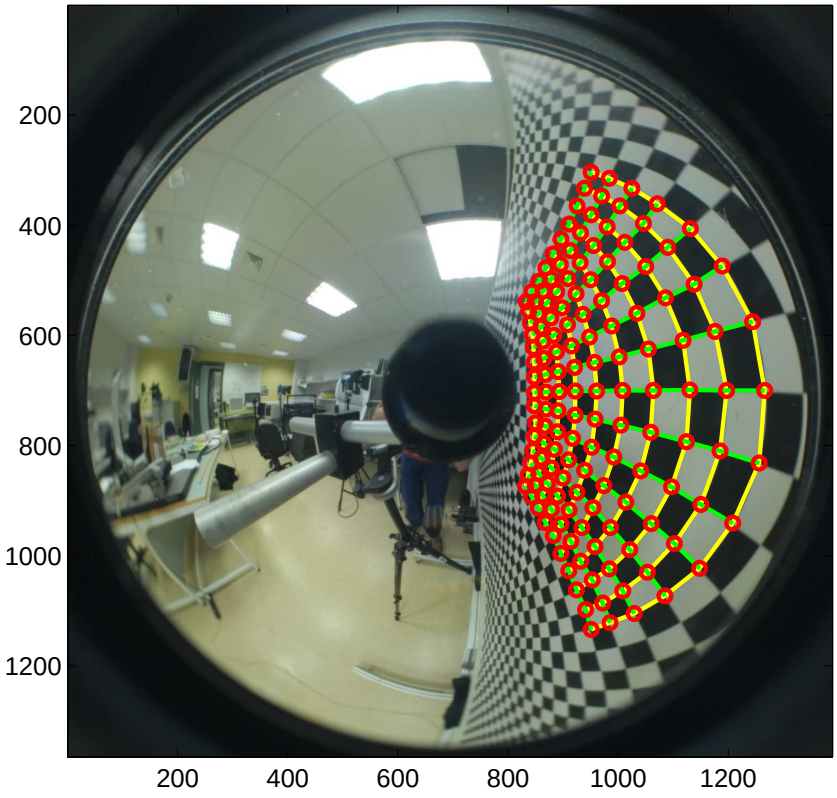
Fig. 3.14 shows images of a large calibration grid taken with a hyperboloidal catadioptric camera and a fisheye camera that each exhibit blurred corners and small grid squares. These unusable corners are easily excluded using the CELECT method by selecting only the grid region that contains localisable corners, as seen in the figure. Such detection and ordering of corners from a partial grid is not possible with either the OpenCV method or the method of

Rufli et al. (2008) since the partial grid is not surrounded by a white border in the image. The images in Fig. 3.14 were modified by using desktop graphics software to insert a white border of increasing size around each of the grids until the OpenCV method returned a correctly detected and ordered corner set. The largest such sets returned by the OpenCV method were of sizes 17×5 and 20×6 for the hyperboloidal image and the fisheye image, respectively. These compare with grids of size 15×10 and 19×11 detected and ordered by the CELECT method without requiring any modification to the images. The CELECT method detected and ordered corners in the fisheye image that are too close to the image boundary to be detected by the OpenCV method.

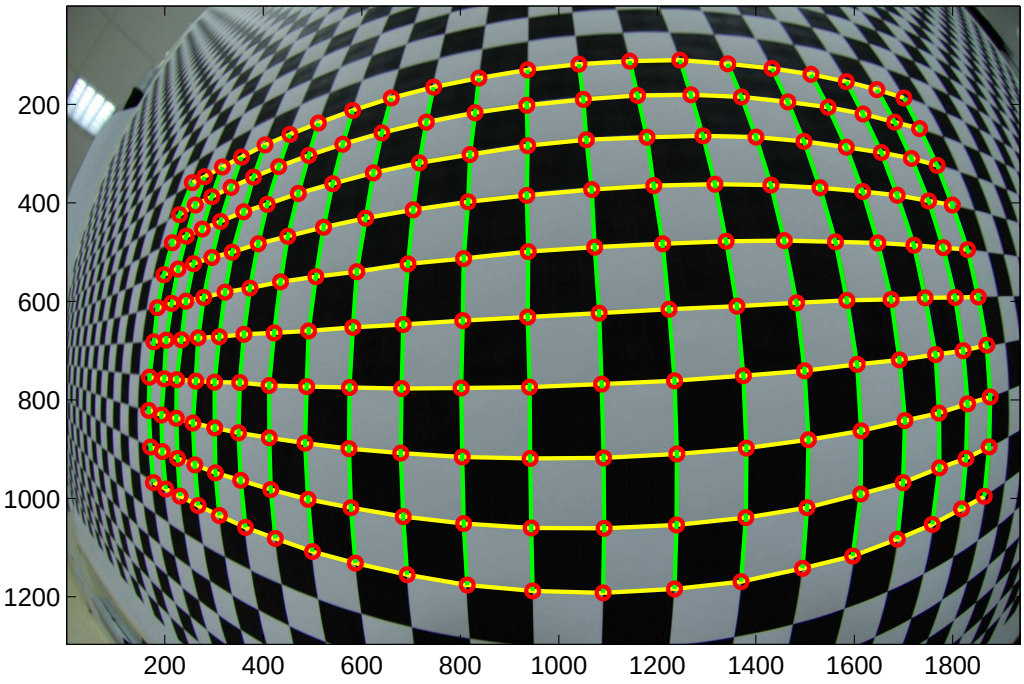
3.3.3 Active Grids Robustness

Active grids overcome the distortion bias associated with homographic interpolation and consequently are ideal for use in the calibration process, and in particular for use in generic calibration. An efficient generic calibration method is presented in Chapter 4 that employs active grids to acquire input data, and therefore a performance evaluation of these grids was conducted. Both the derivative based and the saddle-point based chessboard corner localisation techniques that are discussed in Section 3.1.3 are used for the benchmarking process. When using a standard, static chessboard calibration grid, these methods are typical of the localisation techniques that would be used to subpixelly determine the grid corners. The comparison between the active grids method and these two standard methods is shown in Fig. 3.15. Note that these tests determine the relative performance of the methods for detecting the grid corner locations, and that they do not address the errors resulting from localising points that are not grid corners (which would require bias inducing interpolation for the static grid methods and thus would reduce the performance of the static grid methods relative to the active grid method).

The experiments were conducted with real data by subpixelly localising corners in the image of a chessboard grid, displayed on a TFT monitor, using the two standard methods. An active grid was then displayed on the monitor and decoded. The subpixel corner locations for the active grid method were determined by searching this decoded location data with the known metric grid dimensions. A second active grid was then placed in front of the camera, and the locations on this grid seen by the corner subpixels estimated by each



(a)



(b)

Fig. 3.14: Corner detection and ordering using the CELECT method applied to large chessboard grids without white grid borders: (a) image from a hyperboloidal catadioptric camera; (b) image from a fisheye camera. All axes are in units of pixels.

method were decoded directly. By mapping these locations to the known metric chessboard structure via homographies, the RMS error residuals for each method were determined. It is seen in Fig. 3.15 that the robustness of active grids to variations in camera-grid distance, grid orientation, Gaussian image blur and additive Gaussian noise is superior to that of the standard methods under almost all conditions. The grid pattern used has 15×15 corners and is $301.5 \times 301.5mm$ in size, and, where not otherwise specified, the camera-grid distance is $200mm$, the orientation is 0^o, and there is no blurring or additive noise. The excellent performance of the active grids in the comparison is partly due to their robustness to image sensor blooming, as discussed in Section 3.2.2. The robustness of active grids to image blur as seen in Fig. 3.15 is an important benefit for the calibration process as it means that camera focus can be fixed during calibration.

It can be concluded from these results that active grids have localisation accuracy and robustness equal to or exceeding those of the standard methods for the detection of corners from chessboard calibration grids. Combined with the lack of interpolation bias, active grids are thus ideal for use in generic calibration, where pixel-level localisation is a key requirement of the calibration process. However, one drawback to using active grids during calibration is that the viewing angle for TFT monitors is limited, resulting in a falloff in intensity with increased viewing angle. For catadioptric and other wide angle lenses the usable area of the active grids for certain grid locations is consequently reduced.

The effects on accuracy when the active grids patterns use a grid size other than 8×8 are not examined here, and may merit further investigation. However, the suitability of active grids as an alternative to static calibration grids has been firmly established in this chapter, and any accuracy improvements achieved as a result of such investigation would only confirm this suitability.

3.3.4 Active Grids with Real Images

The localisation performance of active grids was compared to the derivative and saddle point methods discussed in Section 3.1.3, as well as the OpenCV library's *cvFindCornerSubPix()* function, which is based on locating radial saddle points. A Nikon SLR D200 digital camera was calibrated for the ex-

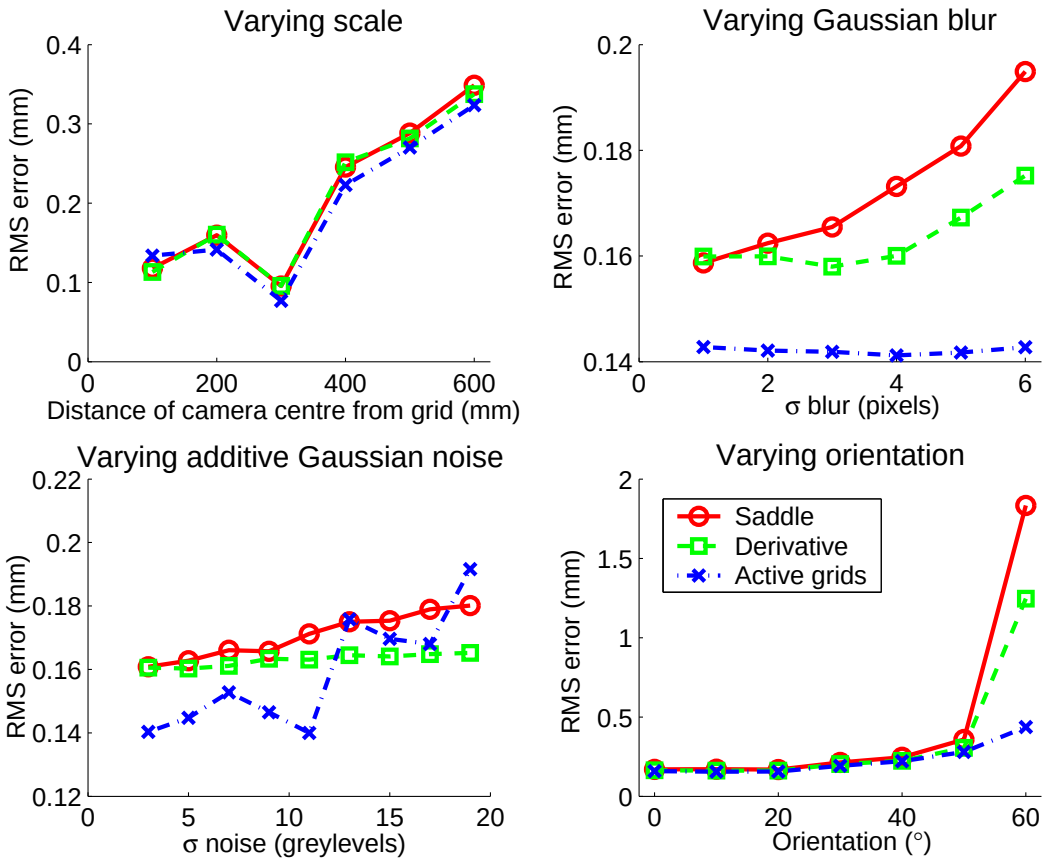


Fig. 3.15: Performance plots for saddle point localisation, derivative localisation and active grids localisation for varying parameters. Grid orientation is measured between the grid normal and the camera principal axis in the horizontal plane.

periment using the method of Zhang (1998). This camera was used to image a monitor displaying a chessboard grid in an approximately fronto-parallel position, and subsequently to image an active grid displayed on the monitor in the same position. All the captured images were undistorted to remove lens distortion (Zhang, 1998). Each of the three localisation methods were used to localise the grid corners in these images, whilst the grid corners locations were extracted from the active grids data by searching for their known metric locations and interpolating between the four nearest available pixels (with no three collinear). The residuals obtained after mapping the metric grid to the localised corners for each method are given in Table 3.8. It is seen that the derivative method's accuracy is superior to that of each of the other methods, although its performance is only marginally better than the performances of the OpenCV and active grids methods. Note that the termination criterion for the OpenCV method was an accuracy of 0.1 pixels. Fig. 3.16 shows the vector

residuals for each method. Whilst it is clear from these plots that some distortion, principally radial, remains in the image despite the distortion removal, the relative performances of the methods are still apparent. The OpenCV and active grids localisation results are similar and only marginally worse than the derivative method result.

The parameter values used in the experiment for the derivative, saddle point and OpenCV methods are also presented in Table 3.8. The image corner points were localised for a wide range of window sizes and blur SDs in order to determine these optimum parameter values. It was found that the localisation performance of the derivative method improves with increasing window size up to the point at which windowed pixels fall outside the image boundary causing the method to fail. The most suitable localisation parameters depend on the apparent grid square sizes in the image. In contrast, the active grids method takes no parameters and is therefore tolerant to changes in the apparent grid square size. Note that the optimum parameters presented in Table 3.8 are not directly comparable to the parameter value results of Section 3.1.3, particularly in the case of derivative localisation, because here the grid images are distortion free rather than containing significant fisheye distortion.

Recall that the test image used to generate the results in Table 3.8 had its radial distortion removed. Therefore additional smoothing and interpolation could be applied in the active grids method to remove the outliers evident in the residuals plot. However, for images with high distortion local smoothing is undesirable. Such an image, taken with a Nikon D200 and a $10.5mm$ Nikkor DX ED fisheye lens, is shown in Fig. 3.17(a). Distortion was removed from the image by using the data from an active grid imaged in an approximately fronto-parallel position. The undistorted image is shown in Fig. 3.17(b), in which it is seen that all the real world straight lines appear straight. The vertical world lines are not vertical in the undistorted image due to the imaged active grid having a negative pitch angle from the camera's principal axis.

One of the benefits of using active grids over conventional static grids is the ability to calibrate discontinuous cameras. Fig. 3.18(a) shows an image of an approximately spherical 'disco' mirror constructed from multiple planar mirror facets. The static chessboard grid reflected in the disco mirror is significantly distorted due to the discontinuity of the mirror surface normals. By replacing the static grid with an active grid, location data was acquired that enabled the

Table 3.8: Chessboard grid localisation errors and method parameters for derivative, OpenCV, saddle point and active grids methods.

Method	Mean (pix)	SD (pix)	Win size	Blur size
Derivative	0.0934	0.0672	49	-
OpenCV	0.1058	0.0775	35	-
Saddle point	0.1536	0.0993	15	9
Active grid	0.1014	0.0713	-	-

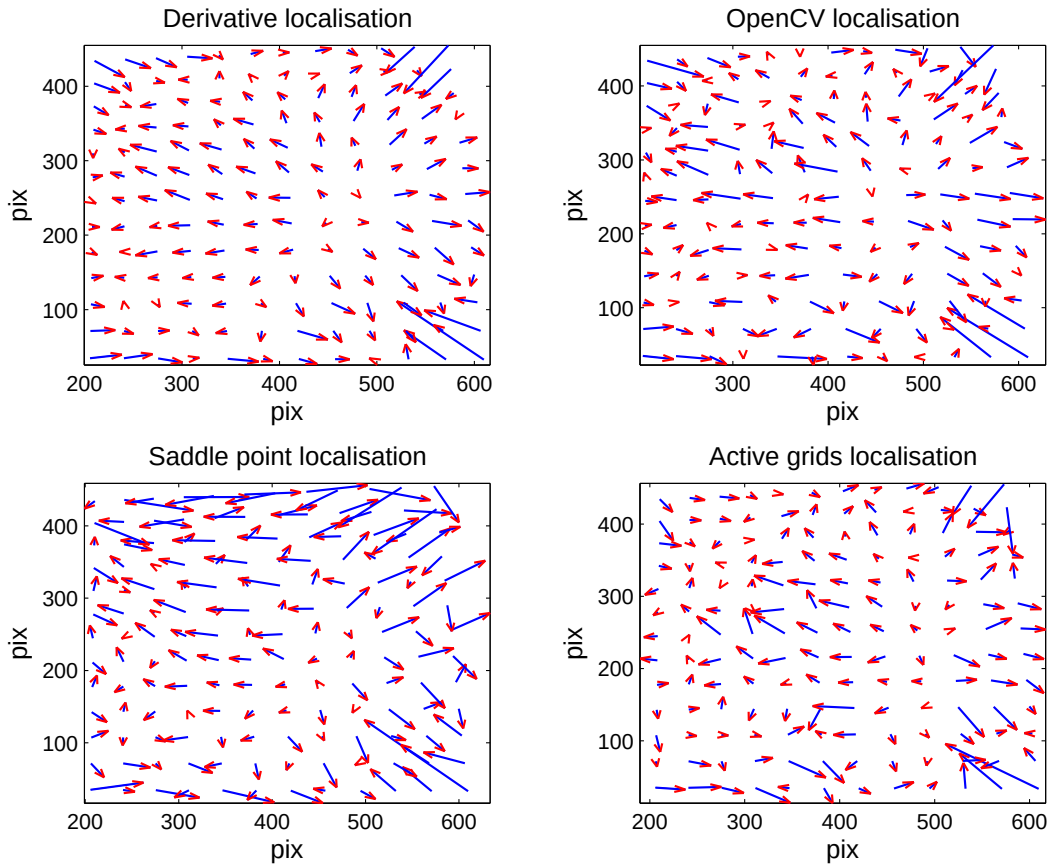


Fig. 3.16: Error residuals after mapping the metric grid to the localised grid corners in an image of a chessboard grid for the derivative, OpenCV, saddle point and active grids methods. Vectors are scaled ×200.

chessboard grid image to be approximately perspectively reconstructed. The corrected image is shown in Fig. 3.18(b) superimposed on the actual chessboard grid, from which it is seen that the grid corners and edges closely match the underlying grid. Noise in the corrected image is due to non-specular inter-facet reflections that can not be decoded.

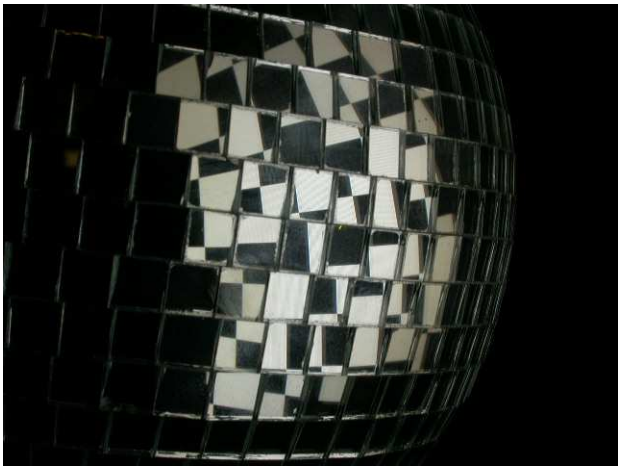
(a) (b)

Fig. 3.17: (a) Original fisheye image; (b) Fisheye image after distortion correction using active grids input data.

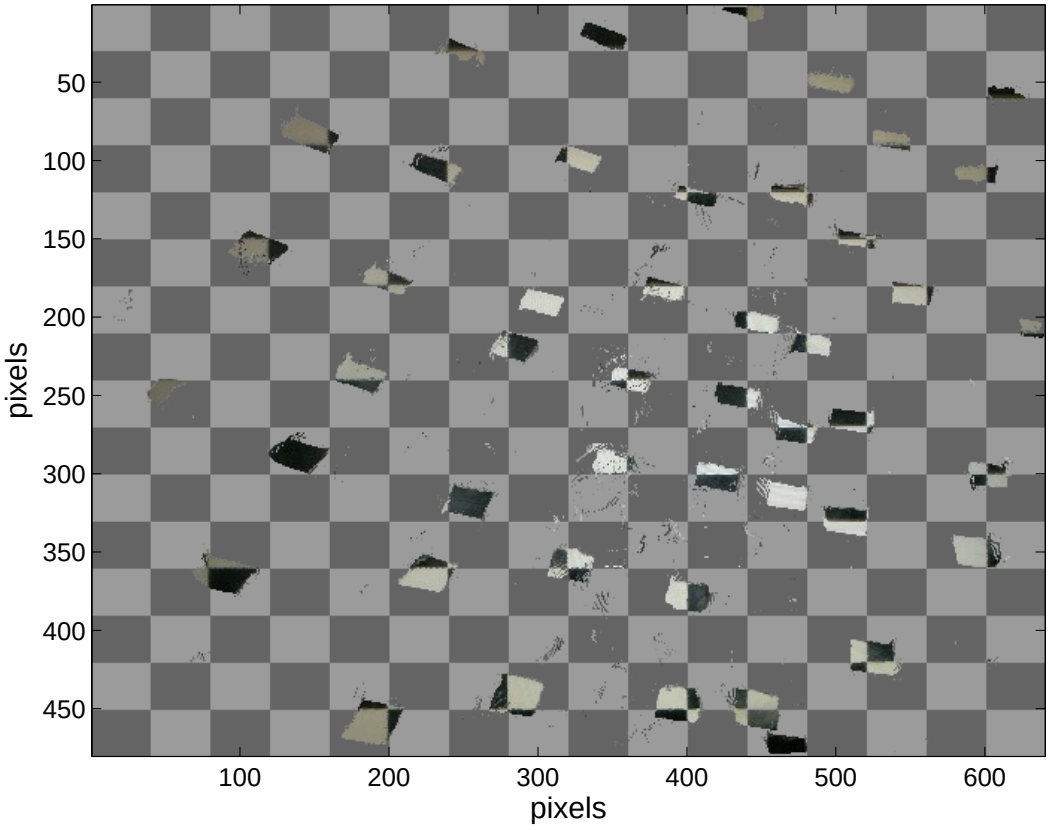
3.4 Discussion

This chapter addresses two practical aspects of calibration using planar calibration grids. These aspects require particular attention when calibrating non-conventional cameras whose images contain significant distortion.

Firstly the task of chessboard corner extraction from images with large distortions is examined, and a solution to this task, the CELECT method, is proposed and evaluated. Corner extraction is necessary in all planar calibration methods that employ chessboard grids, but until recently no non-manual selection methods existed that had wide applicability and that had been shown to operate successfully on severely non-linearly distorted images. Extraction is shown to consist of the subtasks of corner detection, corner ordering and corner localisation. Methods of corner detection are reviewed in Section 3.1.1, whilst localisation is discussed in 3.1.3. The CELECT method, whose primary contribution is the ability to determine correct corner ordering in conjunction with corner detection, is presented in Section 3.1.2. The method is semi-automatic, requiring only the locations of the four outer grid corners, and the number of rows and columns in the grid. Given this information, the grid edge contours are systematically traced, and corner detection is applied locally. The CELECT method does not assume any underlying parametric arrangement of the grid feature points in the image, thus rendering it applicable to calibration grid images with severe distortion. The only assumptions are that the image is continuous, and that the portion of the grid to be extracted has a fixed num-



(a)



(b)

Fig. 3.18: (a) Image of a disco mirror reflecting a chessboard grid; (b) remapped disco mirror chessboard grid image after distortion correction using active grids input data. The correct grid is shown in shades of grey.

ber of grid squares in each row and in each column. The CELECT method solves a similar problem to that of Rufli et al. (2008), although the methods were proposed independently and have different implementations. Due both to the contour tracing approach and to its semi-automatic nature, the CELECT method can detect more corners and can detect corners closer to the image edge than existing automatic methods.

The robustness of the CELECT method to noise, scale, fisheye distortion and perspective distortion is shown in Section 3.3.1 through experiments with synthetic data to be good, although performance is shown to degrade with relatively small levels of blur. Accuracy of the method is also shown to be sufficient for subsequent subpixel localisation to converge. Results for real images from four different types of non-conventional camera are presented in Section 3.3.2, showing that the CELECT method has the capability to correctly detect and order corners in highly distorted images from both central and non-central cameras. Performance is evaluated with respect to the established method for grid corner extraction available in the OpenCV image processing library. The CELECT method achieves improved performance over the OpenCV method for each camera type, and correctly detects and orders the corners in 85.0% of the test images, compared to the success rate of only 43.3% for OpenCV. The vastly superior performance of the CELECT method over the OpenCV method for images from the spherical catadioptric camera demonstrates the effectiveness of the CELECT method under low contrast conditions caused by poor illumination. Weak shadows in the fisheye image set do not affect the CELECT method, but do cause the OpenCV method to miss corners. The high success rate for real images indicates that the CELECT method is a significant practical aid for the calibration of non-conventional cameras when applying calibration methods that require images of planar chessboard grids.

The second key contribution of this chapter is a discussion of spatio-temporally coded grids – active grids – and a characterisation of their performance. Accurate, high-density feature points are an important property of calibration grids that are to be used for generic calibration. High density data can be extracted from static calibration grids by interpolation, but this is shown in Section 3.2.1 to induce a bias in the extracted data points that is due to image distortion. By displaying a sequence of spatially varying patterns on a TFT monitor, active grids can overcome this induced bias. Section 3.2.2 presents an implementation of active grids, in which a combination of course location data, decoded from

binary patterns, and fine location data, decoded from phase shifted sinusoidal patterns, is used to determine the location on the active grid that is seen by each pixel.

Section 3.3.3 provides a performance evaluation of active grids under variations in scale, Gaussian blur, Gaussian noise, and orientation. These experiments conclusively show that active grids can outperform standard localisation methods in terms of their robustness. Results presented in Section 3.3.4 for additional experiments with real data show that active grids have a comparable localisation accuracy to the best performing localisation methods optimised for the test image. The benefits of active grids for the purpose of non-conventional camera calibration are demonstrated for the cases of a fisheye image and an image from a discontinuous camera. This characterisation of active grids shows that they can provide accurate and dense calibration data, which is an important benefit for the calibration method that is proposed in the next chapter.

Chapter 4

Efficient Central Generic Calibration

Standard generic camera calibration, which is discussed in Chapter 2, is a non-parametric calibration technique that is applicable to any type of vision sensor. The goal of generic calibration is to determine the mapping between image pixels and 3-space camera rays. Different variants of the standard generic method for calibration using planar targets have been proposed for central and non-central cameras (Sturm and Ramalingam, 2003, Ramalingam et al., 2005*a*). Nevertheless, the generic calibration framework is the same for both camera types, that is, a collinearity constraint is applied to the intersection points of each camera ray with the grids involved in the linear estimation stage. The solution for the grid poses, and for the camera centre in the case of central cameras, is obtained by solving equations that are formed by enforcing this collinearity constraint (see Section 2.4.1 for details). The central and non-central models only vary in the steps required to solve these equations.

The ability of generic calibration to calibrate any camera type removes the requirement for the practitioner to determine the most appropriate model for any particular camera to be calibrated. There is a trade-off between ease of use and performance. It can be difficult to distinguish between camera types that require specific models. However, some camera categories can be easily identified, and consequently a decision on the camera calibration model can be taken. Central cameras have a single centre of projection, and they include all cameras that fit the pinhole model. Although fisheye cameras are not exactly

central since they have a small locus of projection centres called a diacaustic, they are approximately central for many fisheye lenses (Ying and Hu, 2004, Swaminathan, Grossberg and Nayar, 2003). The complete family of central *catadioptric* cameras has been determined by Baker and Navar (1998). The family comprises (1) a perspective camera and a planar mirror, (2) a perspective camera located at the focal point of a hyperboloidal mirror, (3) an orthographic camera located on the axis of a parabolic mirror, and (4) a perspective camera located at the focal point of an elliptic mirror. All other catadioptric systems are non-central. The identification of approximately central cameras is thus made by reference to this list of central cameras. Consequently, the presentation of a generic calibration method specifically for central cameras is a significant contribution in that such cameras are in wide use and are readily identified as being approximately central.

Standard central generic calibration is a specialisation of the general generic calibration framework, and so a link between the pinhole camera model and the geometric constraints of central cameras is not considered in the standard central generic method. Additionally, since the polynomial based pose estimation method of standard generic calibration is the same for both central and non-central cameras, it does not account for known centrality. This chapter proposes an improved generic calibration method for cameras with a single centre of projection called the CGSP method – Central Generic Synthetic Pinhole calibration method. Alternatives to steps 3 and 5 of Algorithm 1 of the standard generic method are presented that improve the accuracy and robustness of the calibration. Improvements are achieved by using a synthetic pinhole image plane that takes advantage of the geometric constraints resulting from a single centre of projection. Input data for the CGSP algorithm is acquired using the active grids approach that is discussed in Chapter 3, as they can provide the dense input data that is ideal for the new linear estimation stage. The first improvement results from applying a novel method for the linear estimation of the camera centre. The estimation of the camera centre and the initial grid poses in the standard generic method, as detailed in Chapter 2, is complicated, partly due to coupling of the variables to be extracted. In Section 4.1 active grids are shown to facilitate other, more intuitive and more accurate methods of determining the camera centre for central cameras. The second improvement is achieved by way of an alternative pose estimation stage that is proposed for use in generic calibration. The pose estimation

stage proposed by Sturm and Ramalingam (2003) for standard generic calibration is a 3-point polynomial technique that does not lend itself well to large scale single-shot pose estimation. The alternative pose estimation algorithm is derived and evaluated in Section 4.2. Together, the above modifications serve to make the CGSP method for central cameras both more robust and more accurate than the standard method. Simulations and experiments with real data are presented in Section 4.3 that demonstrate the improved performance. Comparative distortion correction experiments are conducted for both an omnidirectional camera and a fisheye camera using the standard and CGSP methods. Motion reconstruction experiments are undertaken for the omnidirectional camera to validate the calibration approach, and the results of these are detailed. The effects of the proposed modifications and the accuracy of the complete CGSP calibration are clearly shown and discussed.

4.1 Linear Estimation

The purpose of the linear estimation stage in central generic calibration is to determine both the grid poses and the position of the camera centre in the camera coordinate system attached to the base (usually first) grid. The camera centre is the single point through which all camera rays would pass if no reflection or refraction occurred, thus the accuracy of the entire calibration is directly dependent on the accuracy of the centre estimate. The linear estimation stage of the standard generic method is discussed in Chapter 2, in which the method of determining both the camera centre and the plane positions and orientations is detailed. Eqns. 2.14 through 2.25 indicate the level of involvement required to decouple and solve for the unknowns. The standard generic method's complexity is further indicated by the fact that Eqn. 2.25, which is non-linear in the unknowns, is solved by observing through simulation that certain variables are always 0.

An alternative linear estimation stage for central generic calibration is proposed, and is used in the CGSP method. This linear estimation stage is less complicated than, and is shown to be more accurate than, the method of standard generic calibration.

4.1.1 Synthetic Pinhole Calibration

The proposed linear estimation stage results from a novel interpretation of existing methods for the calibration of pinhole cameras. As known, pinhole calibration techniques are not directly suitable for wide FOV cameras due to the existence of severe non-linear image distortion that invalidates the pinhole projection model. For cameras with FOVs equal to and exceeding 180^o the pinhole model itself is invalid, since there is no image plane location and positive focal length for which all 3-space points linearly projected through the camera centre can intersect. However, pinhole calibration methods are well established and it would be beneficial to utilise this established theory. The linear estimation stage of the CGSP method enables such utilisation. The key idea is that an additional calibration grid, referred to as the base grid, is used as a synthetic image plane in the calibration process, thus forming a synthetic camera that is exactly pinhole. By placing the base grid in front of the general camera so as to intersect the camera rays on the scene side of the camera optics, as shown in Fig. 4.1, an image is formed on the synthetic image plane. The synthetic image points are the points of intersection of the rays with the inserted synthetic image plane. If this plane is an active grid, the intersection locations can be determined directly as described in Chapter 3.

Consider the camera pixels that view the object in Fig. 4.1. The camera rays associated with these pixels are coincident at the centre of projection of the camera. Therefore the intersections of these rays with an additional grid, the synthetic image plane, will be a perspective projection of the object corners. By accurately determining the intersection points with the synthetic image plane of all the camera rays that intersect the object, a pinhole image of the object is formed on the synthetic image plane. Since the projection from the object through the synthetic image plane to the centre of projection preserves point collinearity, the synthetic image is free of all distortion. The synthetic camera can then be calibrated from at least two such synthetic images of a calibration grid in different positions. Any standard pinhole calibration method can be used to achieve this calibration. Since the centres of the synthetic and general cameras coincide, the desired estimate of the general camera centre is available directly from the synthetic camera calibration as $[p_x \ p_y \ f]^T$, where p_x and p_y are the principal point offsets of the synthetic camera, and f is the synthetic camera's focal length. Furthermore, the pose of grids two and

three can be extracted from the synthetic pinhole calibration using well known techniques (Sturm, 2000). Note that there is no constraint on the placement or the pose of the base grid acting as the synthetic image plane, once it is located externally to the general camera. The only prerequisites for the linear estimation stage of the CGSP method are that the general camera is central and that the calibration targets are planar with known calibration patterns.

The non-linear calibration problem is thus linearised by moving the calibration from a point at which the optics are non-linear to a point at which they are linear. This new approach provides a link between the established theory of pinhole calibration and the generic calibration of central cameras that has not been expounded previously. Within the CGSP calibration framework, this link enables the generic calibration of non-pinhole central cameras using pinhole calibration techniques, so that *any* pinhole calibration technique can be used to calibrate *any* central camera. The minimum number of grids required for the linear estimation stage of the CGSP method is three - two for the pinhole calibration plus one for the synthetic image plane - which is the same number as required for the standard generic method's linear estimation stage. The benefits of active grids as discussed in Chapter 3, in particular their ability to provide ray-grid intersection points directly and accurately, make them ideal for use as synthetic image planes in this method. Standard chessboards, in conjunction with homographic interpolation, could also be used to form the synthetic image plane. However, in that case interpolation would introduce bias into the synthetic camera and so the camera could no longer be precisely modelled with a pinhole model. Note that the use of active grids in prior work was for providing high density feature points, and not as a means of forming a synthetic image plane as proposed here.

A question arises as to which pinhole calibration technique should be used for the linear estimation stage of the CGSP method. To answer this, two well known pinhole calibration techniques, those of Sturm and Maybank (1999) and Wang and Liu (2006), were incorporated into separate implementations of the CGSP method's linear estimation stage. Both of these techniques are based on the same underlying constraints on the Image of the Absolute Conic, IAC, but they take different approaches to determining the solutions. The relative performance of these two implementations was evaluated, resulting in the conclusion that the method of Sturm and Maybank (1999) is more accurate and robust than that of Wang and Liu (2006) for this application.

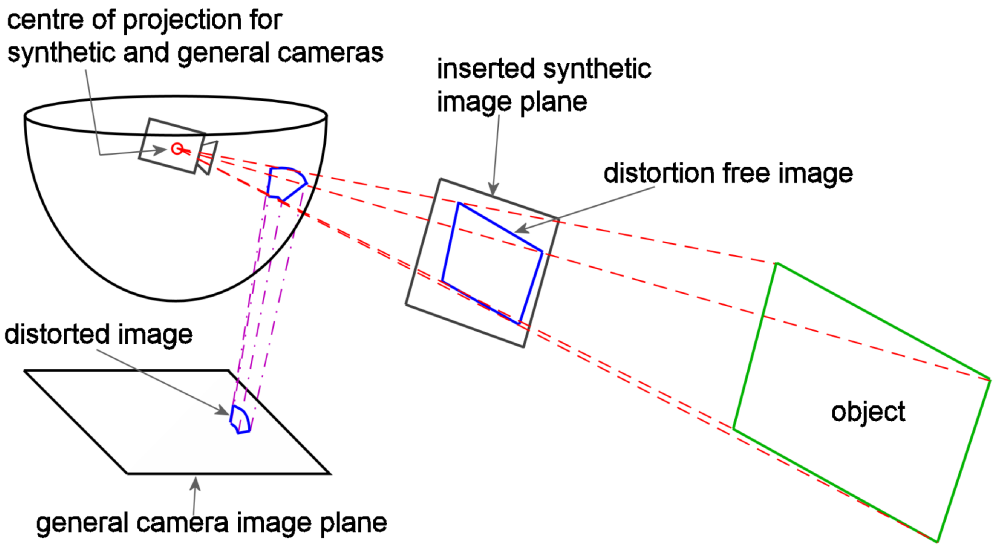


Fig. 4.1: Linear estimation of the camera centre in the CGSP method. The synthetic image plane allows the use of pinhole calibration techniques for determining the camera centre.

Therefore the pinhole calibration technique of Sturm and Maybank (1999) is applied in the CGSP method. This technique is based on the IAC, ω, and the corresponding relationship between ω and the camera calibration matrix given by

$$\omega \simeq K^{-T}K^{-1} \tag{4.1}$$

The application of the technique to the linear estimation stage of the CGSP method is detailed next.

4.1.2 CGSP Linear Estimation Procedure

The CGSP method's linear estimation stage for n calibration grids proceeds as follows. Let $\mathbf{Q}_{ij}$ be the intersection point of ray i with calibration grid j in a coordinate frame attached to grid j, and let the pose of calibration grid j in the world coordinate frame be given by T^j composed of rotation R^j and translation $\mathbf{t}^j$. The perspectivity, H^j, due to central projection gives

$$\mathbf{Q}_{i1[1\ 2\ 4]} \simeq H^j \mathbf{Q}_{ij[1\ 2\ 4]}, \quad \forall i, \, and \, j = 2 \ldots n \tag{4.2}$$

The standard camera projection matrix that maps scene points to image points is given by (Hartley and Zisserman, 2003)

$$P = KR_c[I_{3\times3} \mid -\mathbf{t}_c] \tag{4.3}$$

where R_c and $\mathbf{t}_c$ are the camera rotation and translation, respectively, in the world coordinate system. By placing the synthetic camera such that its image plane is on the world $Z = 0$ plane, and by taking the case of planar scene points, Eqns. 4.2 and 4.3 can be combined to give

$$\begin{aligned} H^j &\simeq K\left[I_{3\times3} \mid -\mathbf{t}_c\right] \begin{bmatrix} R^j_{[1\ 2]} & \mathbf{t}^j \\ \mathbf{0}_{1\times2} & 1 \end{bmatrix} \\ &\simeq K\left[R^j_{[1\ 2]} \mid \mathbf{t}^{j*}\right] \end{aligned} \tag{4.4}$$

where $\mathbf{t}^{j*} = \mathbf{t}^j - \mathbf{t}_c$.

The camera calibration matrix, whose general form is given by Eqn. 2.1, takes on a simpler form for the synthetic camera. The synthetic camera's aspect ratio and skew are determined by the properties of the grid that acts as the synthetic camera's image plane. By using an active grid implemented on a TFT monitor that has zero pixel skew, the skew factor s of the synthetic camera is 0. The synthetic camera's aspect ratio, α, is 1, since correction for the active grid TFT monitor pixel aspect ratio is incorporated in the active grid decoding stage. The simplified camera calibration matrix is

$$K = \begin{bmatrix} f & 0 & p_x \\ 0 & f & p_y \\ 0 & 0 & 1 \end{bmatrix} \tag{4.5}$$

Substituting Eqn. 4.5 into Eqn. 4.1, multiplying and scaling gives an expression for ω in terms of the unknown internal camera parameters

$$\omega \simeq \begin{bmatrix} 1 & 0 & -p_x \\ 0 & 1 & -p_y \\ -p_x & -p_y & f^2 + p_x^2 + p_y^2 \end{bmatrix} \tag{4.6}$$

Combining Eqn. 4.1 with Eqn. 4.4 allows the constraints inherent to the homographies to be used to provide constraints on ω

$$\begin{aligned} H^{jT}\omega H^j &\simeq H^{jT}K^{-T}K^{-1}H^j \\ &\simeq \begin{bmatrix} 1 & 0 & \mathbf{r}_1^{jT}\mathbf{t}^{j*} \\ 0 & 1 & \mathbf{r}_2^{jT}\mathbf{t}^{j*} \\ \mathbf{r}_1^{jT}\mathbf{t}^{j*} & \mathbf{r}_2^{jT}\mathbf{t}^{j*} & \mathbf{t}^{j*T}\mathbf{t}^{j*} \end{bmatrix} \end{aligned} \tag{4.7}$$

where $\mathbf{r}_i^j$ is the ith column of R^j. Using this equation, and given that R^j and $\mathbf{t}^{j*}$ are unknown and that the equation is up to an undetermined scale, two homogeneous linear constraints on ω can be derived

$$\mathbf{h}_1^{jT}\omega\mathbf{h}_1^j - \mathbf{h}_2^{jT}\omega\mathbf{h}_2^j = 0 \tag{4.8}$$

$$\mathbf{h}_1^{jT}\omega\mathbf{h}_2^j = 0 \tag{4.9}$$

where $\mathbf{h}_i^j$ is the ith column of H^j. The unknown elements of ω are then $\{\omega_{11}, \omega_{13}, \omega_{23}, \omega_{33}\}$. There are three camera parameters to estimate, and in the minimal case there are four independent equations resulting from H^2 and H^3 (two constraints per homography), so the system to be solved is overconstrained. The inhomogeneous camera centre, $\tilde{\mathbf{C}}$, after solving linearly for ω using least squares, is given by

$$\tilde{\mathbf{C}} = \begin{bmatrix} \frac{-\omega_{13}}{\omega_{11}} \\ \frac{-\omega_{23}}{\omega_{11}} \\ -\frac{\sqrt{\omega_{11}\omega_{33}-\omega_{13}^2-\omega_{23}^2}}{\omega_{11}} \end{bmatrix} \tag{4.10}$$

The camera calibration matrix of the synthetic camera is

$$K = \begin{bmatrix} \tilde{C}_3 & 0 & \tilde{C}_1 \\ 0 & \tilde{C}_3 & \tilde{C}_2 \\ 0 & 0 & 1 \end{bmatrix} \tag{4.11}$$

Poses of the grids used in the linear estimation stage with respect to the base grid acting as the synthetic image plane can be extracted from the homographies H^j by factorisation. Letting $G^j = K^{-1}H^j$, and with $\mathbf{g}_i^j$ the ith column of G^j,

$$R^j = s\begin{pmatrix} \mathbf{g}_1^j & \mathbf{g}_2^j & \mathbf{g}_1^j \times \mathbf{g}_2^j \end{pmatrix} \tag{4.12}$$

$$\mathbf{t}^j = s\mathbf{g}_3^j + \tilde{\mathbf{C}} \tag{4.13}$$

$$s = \frac{1}{mean\begin{pmatrix} \|\mathbf{g}_1^j\| & \|\mathbf{g}_2^j\| \end{pmatrix}} \tag{4.14}$$

where s is a scale factor. The sign of s is chosen so that the planes are located on the same side of $\mathbf{C}$ as the synthetic image plane. Orthonormal rotation matrices can be obtained via the SVD as in Zhang (1998).

Interestingly, the 15 independent parameters to be computed – 3 for each of $\mathbf{C}$, R^2, $\mathbf{t}^2$, R^3, $\mathbf{t}^3$ – can be determined from a minimum of 4 $(\mathbf{Q}_{i1}, \mathbf{Q}_{i2}, \mathbf{Q}_{i3})$ triplets. Therefore the CGSP method's linear estimation stage provides a

minimal solution to the calibration problem, something that has only recently been proposed for the linear estimation stage of the standard generic method (Ramalingam and Sturm, 2008).

Homographies are determined within a RANSAC framework that selects inliers as the point sets that are in homographic correspondence with one another. By calculating homographies in this way, the CGSP method's linear estimation stage conveniently rejects outliers as part of the estimation process. The H^js are calculated using the standard DLT with normalisation, and the RANSAC parameters are selected based on the experimental results in Chapter 3. Outliers can exist in the decoded data due to image shot noise, non-linearities in the camera radiometric transfer function, and possible incorrect decoding of active grids due to sharp local discontinuities. The implementation of the CGSP method's linear estimation stage requires the computation of homographies between large corresponding datasets (the size of the datasets is given by the number of pixels that see both grids, typically in the region of $20,000$ point pairs). The homographies are calculated using the standard DLT algorithm (Hartley and Zisserman, 2003), with the SVD calculated using the Java JAMA matrix package[1], which was found to be more accurate than the OpenCV implementation. For very large corresponding datasets, a random subset of the point pairs can be used to decrease computation time.

4.1.3 Ray Directions

Once the linear estimation stage is complete, estimates of the camera centre and the calibration grid poses are available. This information is sufficient to calculate the rays associated with each camera pixel that views at least one of the grids. Each ray is defined by the camera centre, through which it must pass, and by the intersection points with each of the calibration grids in the world coordinate frame. The weighted centroid of these intersection points, in conjunction with the centre, are used to determine the Plücker matrices for each ray. The intersection points of rays with the grids involved in the linear estimation stage are weighted more heavily than all other intersection points. This weighting takes account of the greater uncertainty in the intersection point locations associated with the grids whose poses are themselves estimated

[1] http://math.nist.gov/javanumerics/jama/ (Accessed December 2007)

from data resulting from the pose estimation stage (Section 4.2).

It is important to have a computationally efficient method of determining the intersection points of rays, represented by their Plücker matrices, with planes. Due to the skew-symmetry of Plücker matrices, the intersection points of multiple rays with a single plane can be determined from a matrix equation requiring only a single transpose operation and one matrix multiplication. Given Plücker matrices, M_i, for each of n rays i, stored in a matrix M such that $M = [M_1\ M_2\ M_3\ \ldots\ M_n]$, and given a plane, S, the intersection points $\mathbf{X}_i$ of the rays with S are given by

$$\mathbf{X} \simeq -M^T S \tag{4.15}$$

where $\mathbf{X} = \left[\mathbf{X}_1^T\ \mathbf{X}_2^T\ \mathbf{X}_3^T\ \ldots\ \mathbf{X}_n^T\right]^T$.

4.1.4 Bundle Adjustment

In the standard generic calibration method, bundle adjustment, BA, is applied to the result of the linear estimation stage in order to improve the quality of the calibration. The calibration is the look-up table that maps camera pixels to ray directions in 3-space. Clearly, for the general camera model, it is the ray directions that should be updated in any bundle adjustment scheme, but there are several approaches to doing this. The rays are calculated as the join of the camera centre, $\mathbf{C}$, and the 3-space ray-grid intersection points of ray i with plane j, $\mathbf{P}_{ij}$, where $\mathbf{P}_{ij} = T_j\mathbf{Q}_{ij}$. Thus the parameters that determine the calibration of each camera pixel are $(\mathbf{C}, T_j)$. Consequently the ray directions can be adjusted directly, or updated by bundle adjusting T_j or both of $(\mathbf{C}, T_j)$ for $j = 2 \ldots n$.

The first possible scheme is to adjust the ray directions directly, based on a ray-point distance metric. The standard generic calibration method of Ramalingam et al. (2005*a*) uses this scheme. In this case all the rays are forced to pass through the linearly estimated $\mathbf{C}$. The ray directions and the T_js, $j = 2 \ldots n$, are then adjusted to minimise the ray-point distances between rays and $\mathbf{P}_{ij}$s. A second possible approach is to adjust the ray directions indirectly by optimising the T_js, and again forcing the rays to pass through $\mathbf{C}$ and as close as possible to the updated $\mathbf{P}_{ij}$s. For this approach only the T_js, $j = 2 \ldots n$, are directly adjusted. In a third possible scheme, a best-fit line

in 3-space can be determined for the $\mathbf{P}_{ij}$ associated with each ray i, so that adjustments to the location of the camera centre can be made in addition to adjustments to the T_{ij}s, $j = 2 \dots n$. This is a centre adjusting bundle adjustment scheme. For the case of only three points a linear least-squares solution can be determined by fixing a coordinate frame to the plane defined by the join of the $\mathbf{P}_{ij}$s for each i, given by the null vector of $[\mathbf{P}_{i1}\ \mathbf{P}_{i2}\ \mathbf{P}_{i3}]^T$, and then by applying 2D line fitting on this plane. When more than three points are available, the least-squares 3-space line fitting method of Barreto et al. (2008) can be used. This bundle adjustment scheme attempts to find a configuration for the plane poses for which the resulting rays are as close as possible to concurrent.

The centre adjusting bundle adjustment scheme was implemented but was found to perform poorly in tests. Stable convergence was generally not achieved, and many local minima were observed. Due to noise and camera physical setup, the camera rays will not exactly pass through a single point. This poor centrality may be the cause of the instability in centre adjusting bundle adjustment.

In either of the first two approaches a ray-point distance measure can be used as the error metric. This function measures the perpendicular distance between the ith ray and the $\mathbf{P}_{ij}$ intersection points, and can be determined based on the dot product. Given the ray direction unit vector $\mathbf{r}_i$, the closest point along $\mathbf{r}_i$ to $\tilde{\mathbf{P}}_{ij}$ is parameterised by λ_{ij} resulting in the constraint

$$\mathbf{r}_i \cdot (\lambda_{ij}\mathbf{r}_i - (\tilde{\mathbf{P}}_{ij} - \tilde{\mathbf{C}})) = 0 \tag{4.16}$$

The solution for λ_{ij} is then

$$\lambda_{ij} = \mathbf{r}_i \cdot (\tilde{\mathbf{P}}_{ij} - \tilde{\mathbf{C}}) \tag{4.17}$$

The ray-point distance, d_{ij}, can then be determined conveniently as

$$\begin{aligned} d_{ij} &= \|\lambda_{ij}\mathbf{r}_i - (\tilde{\mathbf{P}}_{ij} - \tilde{\mathbf{C}})\| \\ &= \|(\mathbf{r}_i \cdot (\tilde{\mathbf{P}}_{ij} - \tilde{\mathbf{C}}))\mathbf{r}_i - (\tilde{\mathbf{P}}_{ij} - \tilde{\mathbf{C}})\| \end{aligned} \tag{4.18}$$

The bundle adjustment method that is applied throughout this thesis for both the standard generic and CGSP calibration methods is the second approach described above, where $\mathbf{C}$ is fixed and only the T_js, $j = 2 \dots n$, are directly adjusted in order to minimise the geometric ray-point distance. After each

bundle adjustment iteration, the ray directions are recalculated as the join of $\mathbf{C}$ and the centroid of the new $\mathbf{P}_{ij}$s determined by the updated T_js. This is a relaxation optimisation, in which the ray directions are indirectly updated at each iteration. It involves the minimisation of $6(n-1)$ parameters – 6 parameters for each T_j, $j = 2 \dots n$, using the Rodrigues representation. The alternative approach described above, in which the ray directions are directly updated in the bundle adjustment, requires two additional parameters to be minimised for each camera ray involved in the bundle adjustment (typically thousands).

4.1.5 Simulated Experiments

A comparison, using synthetic data, of the robustness to Gaussian noise of the linear estimation stages of the standard generic method, and of the CGSP method, is shown in Fig. 4.2. Errors in the estimation of the camera centre, and in the translation and rotation of the second and third grids involved in the linear estimation stage, are presented (averaged over 50 trials). The ray-point error is the perpendicular distance between each estimated ray and its known point of intersection with each calibration grid (see Section 4.1.4). These results are for a simulated camera with camera centre $[0\ 0\ 600]^T$ (in coordinate frame of base grid), and with focal length and distortion parameters chosen to simulate a wide angle camera with FOV of 100^o. Results are shown for the standard generic method both with and without bundle adjustment of the grid transformations; bundle adjustment was not applied in the CGSP method's linear estimation stage in these experiments.

The results clearly show that the CGSP method's linear estimation stage without bundle adjustment outperforms that of the standard generic method with bundle adjustment across all levels of noise tested. The results also indicate that bundle adjustment does not significantly improve the calibration result for the standard generic method, although the ray-point error is reduced. Possibly this is due to error in the linear estimate of the camera centre, $\mathbf{C}$, which is not updated in the bundle adjustment. The translation and rotation values are coupled in the transformations, and so their errors can not be considered independently and are only indicative of the accuracy of the calibrations. In addition, translation direction is not considered in the translation error. Examination of the mean errors in the intersection point locations after applying

the transformations estimated by the standard generic method, and by the standard generic method with bundle adjustment, reveal that the application of bundle adjustment does indeed reduce this error, despite the larger rotation errors for the standard method with bundle adjustment shown in Fig. 4.2.

4.2 Pose Estimation

Pose estimation is required during generic calibration in order to increase the number of calibrated camera rays. When the pose of an additional grid is estimated, the camera ray associated with each pixel that sees this additional grid can be included in the calibration. Exact solutions to the general pose estimation problem can be found for either three or four non-collinear point-image pairs by solving a fourth or higher degree polynomial (Fischler and Bolles, 1981, Haralick et al., 1994). The most common approach to pose estimation using more than four points is to minimise either the image space error or the object space error by using standard non-linear minimisation techniques. Quan and Lan (1999) presented linear methods for determining pose from n points, and (Lu et al., 2000) proposed an iterative technique. As described in Section 2.4.2, a geometric three point algorithm that operates for both central and non-central cameras is used to estimate pose in the standard generic method. However, when included in a RANSAC framework, re-estimation of the pose using all inliers (typically the final step in RANSAC) is not possible using this algorithm. The algorithm is also very sensitive to additive noise (although a guided selection of sufficiently separated points can alleviate this problem). Additionally, the required polynomial solving can be memory intensive – in tests the polynomial solver often did not produce a solution, resulting in non-linear minimisation being implemented to solve the equations.

A linear least-squares solution to the pose estimation problem for central generic calibration is therefore described to overcome these issues. Although the method does not minimise geometric error, it is linear, fast, always gives a solution, and can conveniently be incorporated within a RANSAC framework.

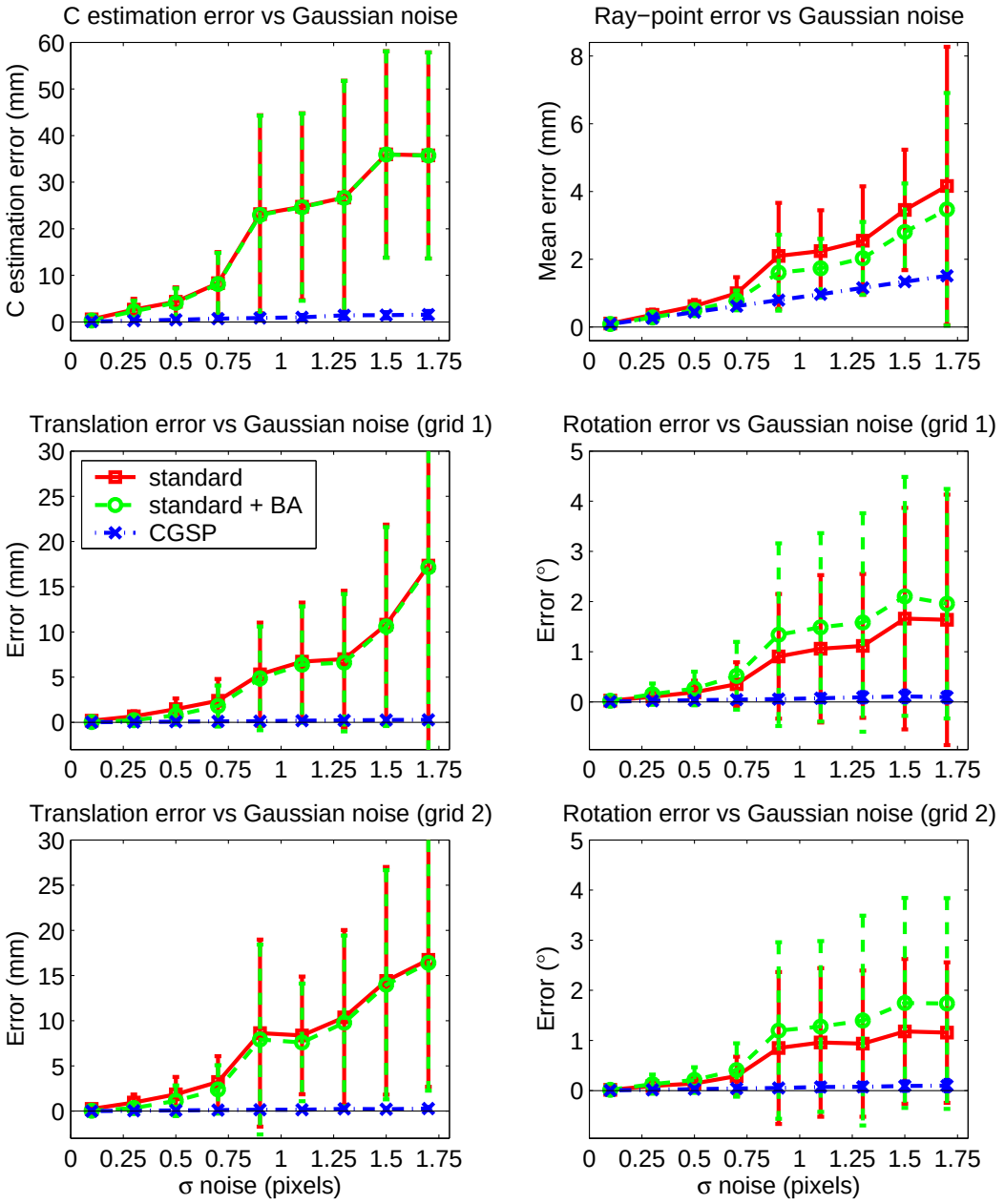


Fig. 4.2: Centre and transformation estimation errors versus Gaussian noise for the standard generic method, the standard generic method with bundle adjustment, and the CGSP method. The rotation error is defined as the magnitude of the Rodrigues angle of $R_{actual}^{-1} R_{estimated}$. Note that the SDs for the CGSP method in these results are non-zero, but are significantly smaller in magnitude than the SDs of the standard method.

4.2.1 CGSP Pose Estimation Procedure

The method allows pinhole pose estimation to be applied to central generic cameras through the use of a synthetic image plane. By intersecting the previously calibrated camera rays that see points on the unknown grid with the synthetic plane, the distortion free pinhole projection of the unknown grid is formed on that plane. Consider this synthetic plane as the image plane of a synthetic pinhole camera, where the pinhole camera's centre is coincident with the camera centre of the general camera. With reference to Fig. 4.3, given a grid in the base position with world coordinate points $\mathbf{X}_i$, and given a grid with an unknown pose T relative to the base grid containing the corresponding unknown world points $\mathbf{X}'_i$, the goal is to determine the unknown pose T. Although general cameras are usually not pinhole, a solution is possible via the insertion of the synthetic image plane in a known orientation (the orientation selection is discussed later) between the camera centre and the grid with unknown pose, as shown in Fig. 4.3. Since the pose of the synthetic image plane is chosen, all the intrinsic and extrinsic parameters of the synthetic pinhole camera are known. They are the camera projection matrix P, camera calibration matrix K, camera rotation R_s, and camera centre $\mathbf{C}$. Pose estimation can therefore proceed using the established pose estimation method for pinhole cameras described by Sturm (2000). Note that the synthetic image plane is only a mathematical construct, and is not physically realised. The general pose estimation problem for central cameras is therefore cast as a pinhole pose estimation problem for which an established solution is available.

Points $\mathbf{X}'_i$ can be projected onto the synthetic image plane by intersecting the previously calibrated rays with this plane. The corresponding projection is according to the pinhole model

$$\mathbf{x}'_i \simeq P\mathbf{X}'_i \tag{4.19}$$

where $\mathbf{x}'_i$ are the imaged points on the synthetic image plane. Also, since the base grid is on the world $Z = 0$ plane

$$\mathbf{x}'_i \simeq H\mathbf{X}_{i[1\ 2\ 4]} \tag{4.20}$$

and

$$\mathbf{X}'_i = T\mathbf{X}_i \tag{4.21}$$

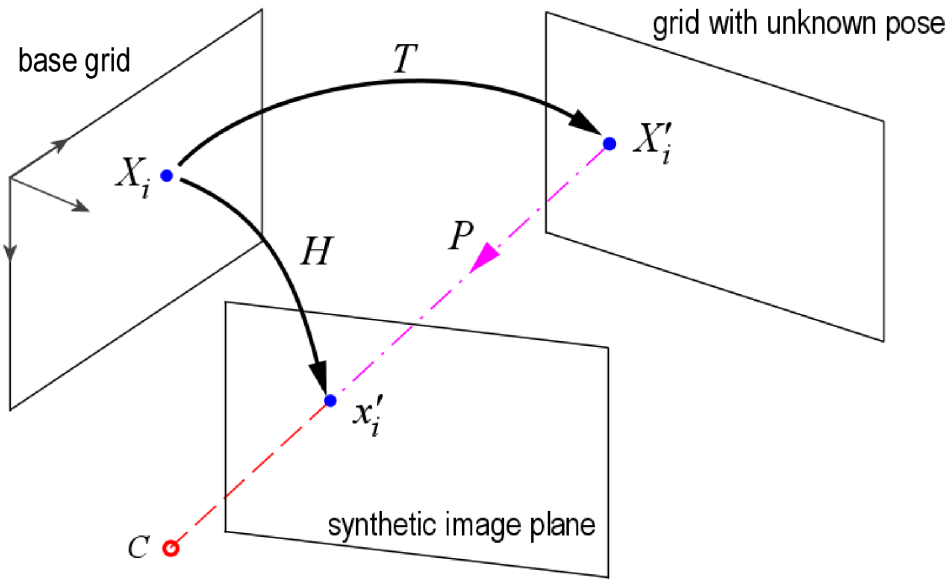


Fig. 4.3: The CGSP method's linear pose estimation scheme, employing a synthetic image plane.

where H is a homography. Therefore

$$\begin{aligned} H\mathbf{X}_{i[1\ 2\ 4]} &\simeq PT_{[1\ 2\ 4]}\mathbf{X}_{i[1\ 2\ 4]} \\ &\simeq KR_s[I_{3\times 3}|-\tilde{\mathbf{C}}]T_{[1\ 2\ 4]}\mathbf{X}_{i[1\ 2\ 4]} \end{aligned} \tag{4.22}$$

giving

$$(KR_s)^{-1}H \simeq \begin{pmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \mathbf{t}_4 - \tilde{\mathbf{C}} \end{pmatrix} \tag{4.23}$$

Letting $G = (KR_s)^{-1}H$, and applying Eqns. 4.12, 4.13 and 4.14, a solution for the rotation, $\hat{R}$, and the translation, $\mathbf{t}$, of the grid with unknown pose can be determined. An orthonormal rotation R is obtained from $\hat{R}$ via the SVD. Non-linear minimisation can subsequently be applied to the linearly estimated pose using the ray-point error metric described in Section 4.1.4.

One is free to choose the orientation of the synthetic image plane, after which the synthetic camera parameters are directly determined. Ideally the synthetic image plane should be as close as possible to perpendicular to the known rays involved in the pose estimation process. This orientation can be determined in a least-squares sense by minimising the sum of the angles between the calibrated rays and the normal of the synthetic image plane, in a similar way to Ramalingam et al. (2006). The unit plane normal, $\mathbf{n}$, can be found as the solution to

$$\underset{\mathbf{n}}{\operatorname{argmin}} \sum_{i=1}^{m} \|[\mathbf{r}_i]_\times \mathbf{n}\|_2 \quad \textit{subject to } \|\mathbf{n}\| = 1 \tag{4.24}$$

where $\mathbf{r}_i$ is the unit vector representing ray i, and $[\mathbf{r}_i]_\times$ is its corresponding skew-symmetric matrix.

4.2.2 Evaluation

The robustness to additive Gaussian noise of the pose estimation stage of the standard generic method, and of the pose estimation stage of the CGSP method, both embedded in RANSAC frameworks, was evaluated for synthetic data. The simulated camera centre was fixed at $[0\ 0\ 600]^T$, and the translations and Euler rotations of the grid, whose pose was to be estimated, were randomly chosen from $[-150mm\ 150mm]$ and $[-30^o\ 30^o]$ respectively. The mean rotational and mean percentage translational errors over 50 trials are shown in Fig. 4.4. It is seen that the CGSP method's pose estimation consistently outperforms that of the standard generic method over all simulated levels of noise. Embedding the linear method in a RANSAC phase does not improve the CGSP method's performance because the synthetic image data in the experiments does not contain any outliers.

The two pose estimation methods, incorporated in RANSAC frameworks, were also evaluated against each other using real data so that their robustness to outliers could be determined. The experiment consisted of a Kodak MegaPlus 1.4i perspective camera imaging an active grid in two orientations, related by a transformation consisting of a randomly selected rotation and translation. The camera was pre-calibrated using the plane-based method of Zhang (2000), allowing the ray-pixel look-up table to be determined directly. A 3D laser scanner (depth resolution $< 0.1mm$) was used to gather ground truth data for the scene, from which the pose of the active grid in the second position relative to the first position was accurately determined. Using the pose estimation methods of the standard generic and CGSP methods, the poses of the active grid in each position were determined, allowing the relative pose estimates to be calculated and compared to the ground truth. The experimental results in Table 4.1 show the angular and translational errors for each pose estimation method. The CGSP method's pose estimates are again seen to be more accurate than the those of the standard generic method. The standard generic method still performs relatively well due to the small percentage of outliers in the data ($< 0.3\%$). The results also show the importance of using RANSAC for the linear pose estimation, since without RANSAC the linear pose estimates

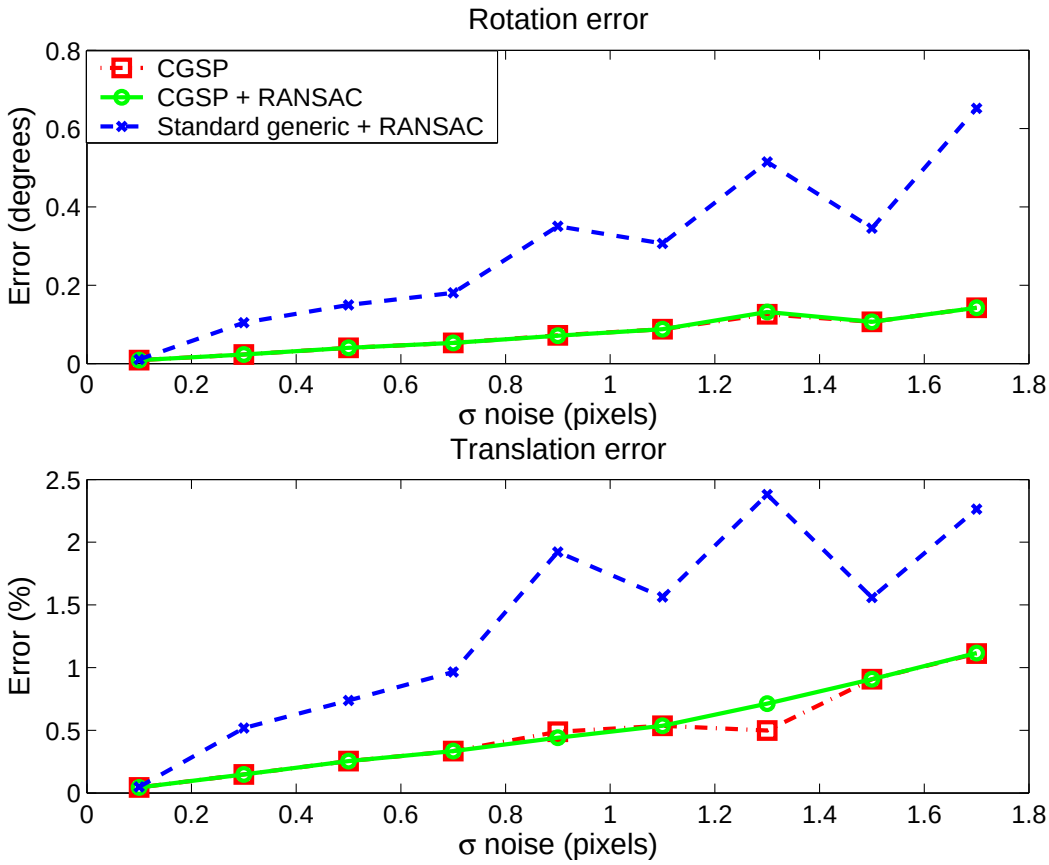


Fig. 4.4: Performance comparison of the pose estimation stages of the standard generic method with RANSAC, the CGSP method, and the CGSP method with RANSAC, for synthetic data with increasing levels of additive Gaussian noise. The rotation error is defined as the magnitude of the Rodrigues angle of $R_{actual}^{-1} R_{estimated}$.

are seen to degrade significantly due to outliers. Error magnitudes of the real experimental results are larger than the error magnitudes of the results for synthetic data due to the smaller grids and larger camera-grid distances used in the real experiments.

The implementation of the standard generic method's pose estimation is incorporated in a RANSAC framework for all experiments in this section. Corresponding re-estimation using all inliers is performed by non-linearly minimising ray-point distances. The CGSP method's pose estimation is also included in a RANSAC phase unless otherwise indicated. However, for that method re-estimation is performed linearly using the method described in Section 4.2.1, and non-linear refinement is not applied.

Table 4.1: Performance comparison of the pose estimation stages of the standard generic method with RANSAC, the CGSP method, and the CGSP method with RANSAC, for real data. The rotation error is defined as the magnitude of the Rodrigues angle of $R^{-1}_{ground\ truth}R_{estimated}$.

	Standard + RANSAC	CGSP	CGSP + RANSAC
Rotation error (^{o})	1.49	1.53	0.75
Translation error (mm)	9.27	17.77	7.63

4.3 Experiments

Both the standard generic and CGSP methods are analysed for real data with respect to linear estimation calibration parameters, a ray-point error metric, camera centrality, distortion correction, and separate motion reconstruction tasks.

4.3.1 Experimental Setup

An omnidirectional catadioptric camera and a camera with fisheye lens were used to capture the images for the experiments. The omnidirectional camera consists of a 360 OneVR hyperboloidal omnidirectional mirror[2] mounted on a Nikon D70 SLR digital camera. With the correct positioning and alignment, this catadioptric configuration has a single centre of projection. However, the mirror could not be mounted directly onto the camera due to limitations on the minimum focusing distance of the camera lens, and thus an external bracket was used to fix the configuration. The complete omnidirectional camera configuration is shown in Fig. 4.5. Mirror alignment was performed based on the known position of the closest focus of the hyperboloid, and by alignment of the mirror outline to the centre of the image. However, accurate alignment could not be guaranteed (an evaluation of the centrality of the experimental configuration is presented in Section 4.3.2). The second camera used was a Nikon CoolPix 4500 digital camera attached to a Nikon FC-E8 fisheye converter lens, which has a 183^o FOV. For each calibration method approximately 207^o of the horizontal FOV and approximately 82^o of the vertical FOV of the omnidirec-

[2]Kaidan Inc., Feasterville, PA

tional camera was calibrated; for the fisheye camera approximately 94% of the entire FOV was calibrated. Three grids were used in each calibration for the linear estimation stage, and a further three grids for the omnidirectional and seven grids for the fisheye cameras to extend the calibrated regions to include additional pixels. A minimal number of grids were used in the linear estimation stage of calibration in these experiments in order to examine the relative performance of the standard generic method and the CGSP method in the most challenging case. For high fidelity calibration more than the minimum number of grids should be employed. Lens movement in the Nikon D70 for upright and inverted camera positions exceeded 5 pixels for a 3008×2000 image, and exceeded 9 pixels for the CoolPix for a 1600×1200 image, and so in order to eliminate this error, camera movement was kept to a minimum during calibration. Active grids were used for all grids during calibration, and the same images were used as input to both calibration methods. This ensured that direct comparisons between the standard generic and CGSP methods are not influenced by the type of input data. A RANSAC stage is applied to the locations decoded from the active grids in order to remove outliers from the decoded location data. Normalisation is applied in the computation of all homographies.

4.3.2 Experimental Results

For both cameras, the estimates of the camera centre, and the Rodrigues rotation magnitudes and translation magnitudes for the poses of the second and third grids used in the linear estimation stages for each calibration method are shown in Table 4.2. The difference between the estimated values of the camera centre z coordinate for each method for the omnidirectional camera is significant. In contrast, all the estimated parameters for the fisheye camera using both calibration methods are within 2.19% of each other, showing closer agreement than the omnidirectional parameter estimates.

The ray-point error metric, described in Section 4.1.4, can be applied to each calibration dataset to give an indication of the relative errors in the calibrations (ground truths for the camera centres and the second and third grid positions are not known). Table 4.3 shows the mean and standard deviation of the ray-point errors for each method and for each camera, both before and after bundle adjustment. Bundle adjustment is applied to the CGSP method here

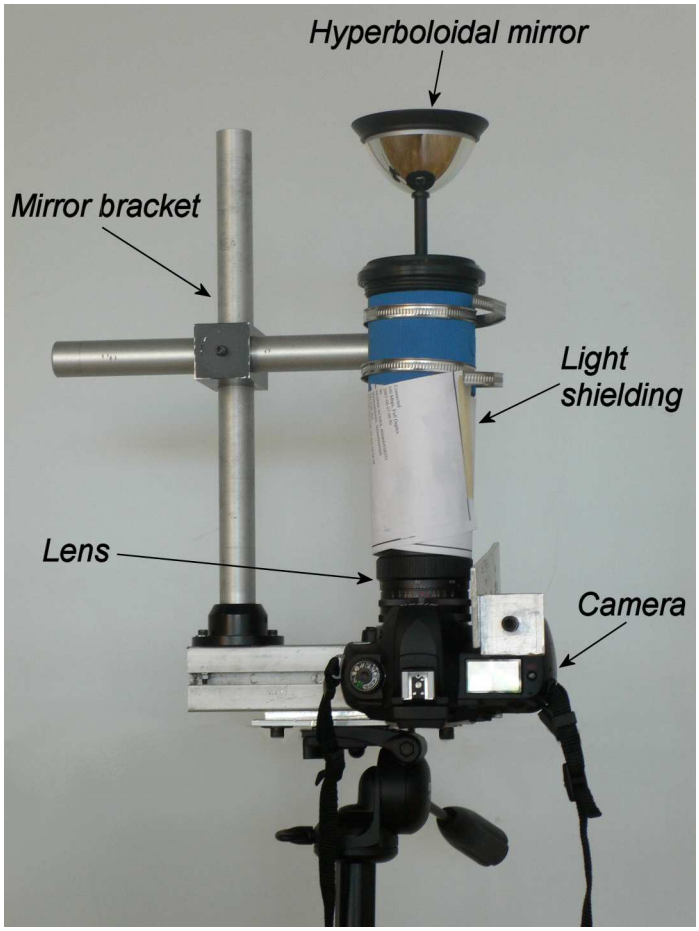


Fig. 4.5: The omnidirectional camera configuration used in the experiments.

for comparative purposes only. For the linear estimation stage, the non-bundle-adjusted parameters are used in the remainder of the calibrations with the CGSP method, whereas the bundle adjusted results are used for calibration with the standard generic method (as per Ramalingam et al. (2005*a*)). For the omnidirectional camera the ray-point error results clearly show that the non-bundle-adjusted configuration of the camera centre and the calibration grids is in greater geometric agreement for the CGSP method than for the standard generic method. The ray-point error and its standard deviation after calibration with the CGSP method are approximately an order of magnitude smaller than the results for calibration with the standard method, both before and after bundle adjustment. The difference in the magnitude of the errors for the calibration methods is less for the calibration of the fisheye camera, but the mean errors after applying the CGSP method are smaller than for the standard method. For both cameras, bundle adjustment reduces the ray-point errors of the standard generic method calibrations significantly. The relatively small ray-point error of the omnidirectional camera calibrated using the CGSP

Table 4.2: The camera centre and grid transformation estimates for the omnidirectional and the fisheye camera calibrations. The centre and the translations are measured in mm, the rotations are measured in degrees.

Camera	Parameter	Standard method	CGSP method
Omnidirectional	$\tilde{\mathbf{C}}$	169.29	167.90
		152.06	159.82
		-106.48	-116.21
	R_2	34.78	36.83
	$\mathbf{t}_2$	130.58	139.66
	R_3	36.55	36.83
	$\mathbf{t}_3$	182.84	192.13
Fisheye	$\tilde{\mathbf{C}}$	104.71	107.00
		161.89	162.29
		-125.69	-124.01
	R_2	43.34	42.59
	$\mathbf{t}_2$	122.18	120.22
	R_3	37.95	37.28
	$\mathbf{t}_3$	287.72	282.39

Table 4.3: Ray-point errors (mm) for all rays involved in the linear estimation stage for each calibration method and for each camera.

		Omnidirectional		Fisheye	
Method	Error type	Error	Error after BA	Error	Error after BA
Standard method	Mean	3.22	1.16	0.32	0.09
	SD	1.49	0.66	0.16	0.06
CGSP method	Mean	0.19	0.13	0.14	0.08
	SD	0.09	0.07	0.09	0.07

method indicates that the misalignment of the omnidirectional mirror with the camera is not significant. Overall, the difference in performance between the standard and CGSP calibration methods is less for the fisheye camera than for the omnidirectional camera. Additionally, the ray-point error is similar across both camera types after applying the CGSP calibration, but is an order of magnitude different when standard generic calibration is applied. These results agree with the simulated results in Section 4.1.

Centrality of the camera configurations after calibration was examined to get further insight into the quality of the calibrations. This is done by determining the best-fit 3-space ray, $\hat{i}$, based on the world intersection points $\mathbf{P}_{ij}$ of ray i with grid j for $j = 1, 2, 3$, which are estimated in the calibration. Least squares ray fitting for three points in 3-space is performed as described in Section 4.1.4. Then the rays $\hat{i}$ are intersected with a plane passing through the estimated camera centre $\mathbf{C}$ that is closest to perpendicular to the $\hat{i}$s (see Section 4.2 for a method of calculating this plane). The distribution of the intersection points on this plane indicates the extent of centrality of the camera calibration. Since only the rays that intersect the first three grids are used, the FOV of these rays is relatively small. Consequently the use of an intersecting plane that is close to perpendicular to all these rays is important, since in this orientation the distribution of the intersection points on the plane is most compact (under the assumption that the rays are affected by Gaussian noise only). Fig. 4.6 shows plots of these distributions for each calibration method for each camera. These plots are for the estimated camera centre and grid poses after application of the linear estimation stage only and before bundle adjustment is applied. It can be seen that the distribution is highly non-Gaussian and non-isotropic for the omnidirectional camera calibrated with the standard generic method. In contrast, the distribution is both more compact and closer to Gaussian for the same camera calibrated using the CGSP method. The wide distribution for the standard generic method may partly be a result of inexact camera centrality, but importantly the distribution for the CGSP method indicates that it achieves a solution that is geometrically consistent with an approximately central configuration. Therefore misalignment between the camera and the mirror in the experimental configuration is not significant. For the fisheye camera, the results using the two calibration methods are very similar, although the distribution is marginally closer to Gaussian for the CGSP method. Note that a lower bound exists on the distribution of the intersection points for the fisheye camera due to the fisheye diacaustic. An isotropic Gaussian distribution of the errors indicates that the camera centre estimate and grid poses are in geometric agreement, with the error resulting solely from the Gaussian noise in the intersection point locations. The fisheye camera centre estimated in the CGSP method's linear estimation stage is within the convex hull of the intersection points. However this is not the case for the camera centre linearly estimated using the standard method, indicating greater inconsistency between the estimated camera centre and the grid poses for that method.

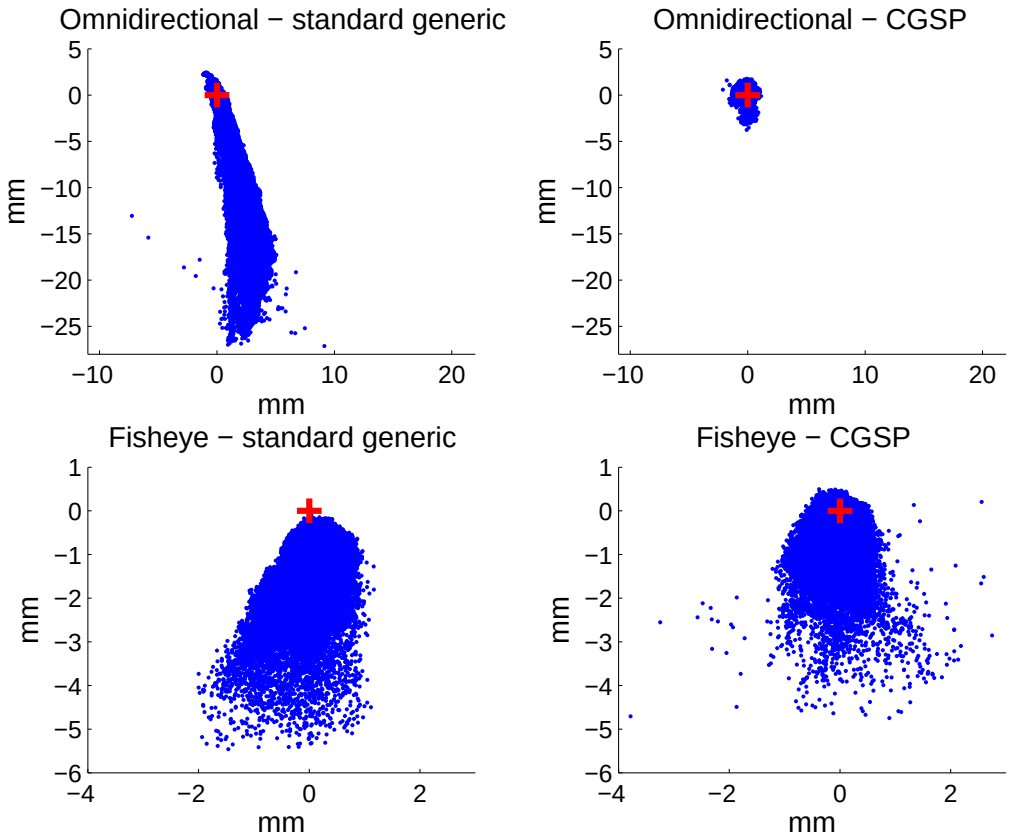


Fig. 4.6: Distribution of the intersection points of the best-fit rays for grids 1-3 with a plane that is most perpendicular to these rays and that passes through the estimated camera centre, for each camera and for each calibration method. The camera centre is shown with a red +. Note that the scales of the axes are smaller for the fisheye plots.

Distortion Correction

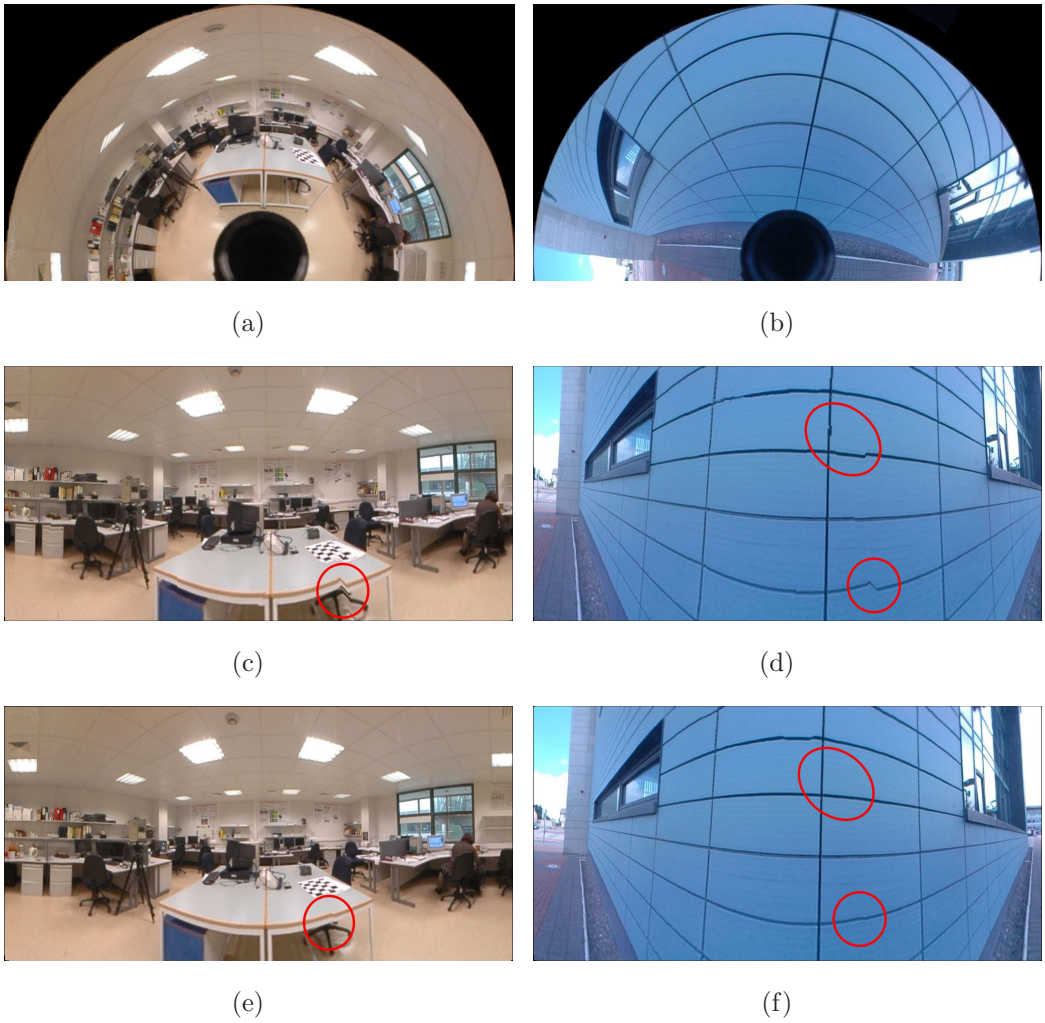
Two distortion correction experiments are conducted in order to both qualitatively and quantitatively evaluate each of the calibration methods.

In the first experiment, the calibration data is used to remove the inherent non-linear distortion from the calibrated regions of images of real scenes. For the omnidirectional camera, a portion of a cylindrical image was formed by intersecting the calibrated rays with a unit cylinder, the axis of which was coincident with the camera centre, and then by unwrapping the cylinder to form a planar image. Fig. 4.7 shows the original images and the cylindrically unwarped images calculated using the calibration data from both the standard generic and CGSP methods. As expected, real world straight lines that are parallel to the mirror axis (vertical) are mapped to straight lines in the images

corrected using either method. However, some abberations are visible in the images corrected using the standard generic method (highlighted by ellipses). In contrast, the corrected images formed using the CGSP method have significantly less aberration. The FOV of the cylindrically unwarped images in Figs. 4.7(c) and 4.7(d) is less than that of Figs. 4.7(e) and 4.7(f) due to the difference in magnitude of the camera centre z coordinate estimated using the two methods.

Distortion correction for the fisheye camera is best demonstrated by generating perspectively corrected images from the originals. This is readily achieved by intersecting the calibrated camera rays with a plane whose orientation is determined as in Section 4.2. The results after perspective correction using each calibration dataset, and for the same image pixels, are shown in Fig. 4.8. Again, real world straight lines properly appear as straight lines in both sets of corrected images. Similarly to the distortion corrected omnidirectional images, there are some visible aberrations in the images corresponding to the meeting points of mis-estimated grids. However, their magnitudes are so small that they are only noticeable at close inspection. The similarity of the results in Fig. 4.8 is to be expected given the correlation of the linearly estimated values for the camera centre and the grid orientations determined by each generic calibration method, as given in Table 4.2.

Quantitative evaluation of the calibrations was carried out by generating perspectively corrected images of planar chessboard grids. The plane onto which the corrected images were projected was selected as described in Section 4.2. Distortion residuals were measured for each image after applying a homography between the distortion corrected image of the grid and the known metric grid structure. Fig. 4.9 shows the distortion residuals for both the standard and CGSP methods for both cameras. In the case of the omnidirectional camera, no radial distortion bias is visible in either vector plot, but the plot for the standard generic method displays large divergences along roughly vertical lines at the left and right of Fig. 4.9(a), indicated by dashed lines. These coincide with areas where two active grids with mis-estimated pose meet, and correspond to the aberrations seen in Figs. 4.7(c) and 4.7(d). The divergences for the omnidirectional camera seen in Fig. 4.9 and the distortion correction residuals given in Table 4.4 are smaller for the vector plot using the CGSP method than for the vector plot using the standard generic method, indicating a better calibration. For the fisheye camera the distortion residual plots are



(a) (b)

(c) (d)

(e) (f)

Fig. 4.7: Omnidirectional camera distortion correction results: (a),(b) original omnidirectional images; (c),(d) cylindrically unwarped images after standard generic calibration; (e),(f) cylindrically unwarped images after CGSP calibration.

almost identical across the two calibration methods. The mean and standard deviation of the residuals for the fisheye camera using both methods are also very similar, which is as expected given the similarity in the linear estimated parameters for each method, and since the distortion correction is applied to the centre region of the image where the rays are determined by the linear estimation stage alone. Note that the errors towards the left and right of the vector plots are amplified due to the increased skew of the rays relative to the perspective plane in these regions.

(a) (b)

(c) (d)

(e) (f)

Fig. 4.8: Fisheye camera distortion correction results: (a),(b) original fisheye images; (c),(d) perspectively corrected images after standard generic calibration; (e),(f) perspectively corrected images after CGSP calibration.

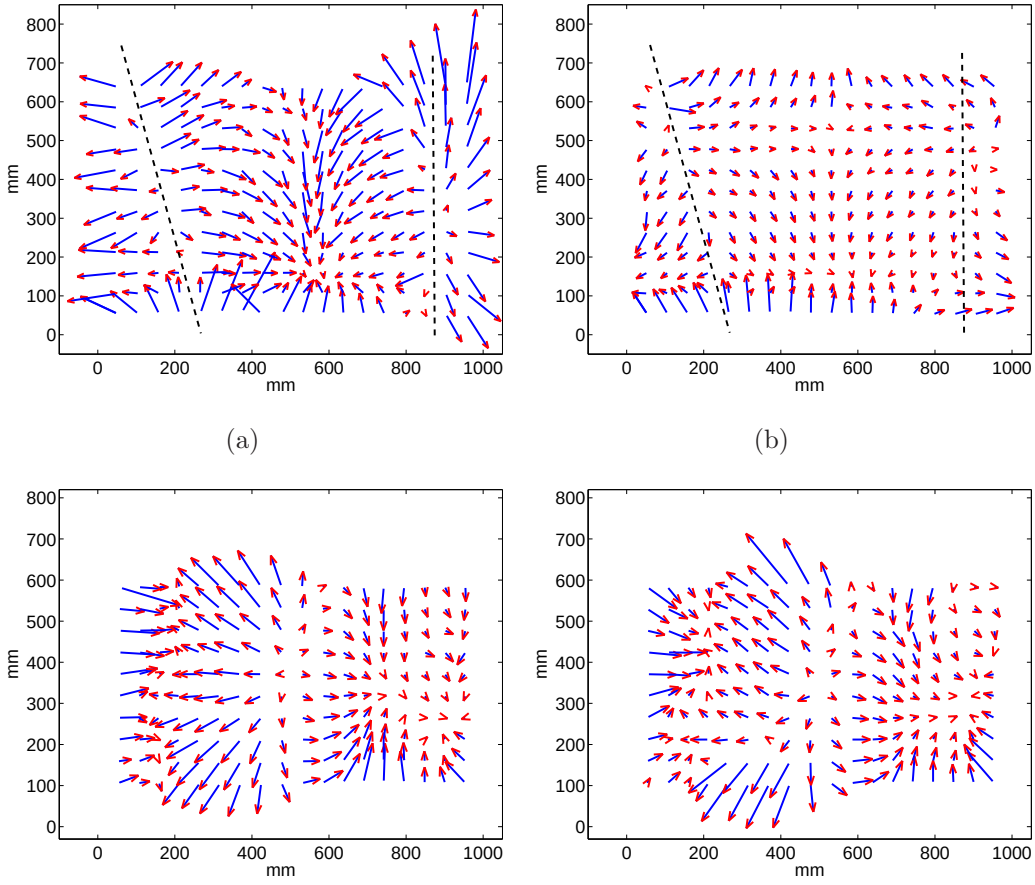


Fig. 4.9: Vector plots of the residuals after perspective correction of a chessboard grid: omnidirectional camera using (a) standard generic method calibration data, and (b) CGSP method calibration data; fisheye camera using (c) standard generic method calibration data, and (d) CGSP method calibration data. Distortion corrected points are mapped to the same size metric grid for both cameras so as to enable direct comparison of the residuals. Vectors are scaled $\times 15$.

Table 4.4: Residuals (mm) after distortion correction for the omnidirectional and the fisheye cameras using the standard generic and the CGSP methods.

Method	Error type	Omnidirectional	Fisheye
Standard method	Mean	4.54	2.93
	SD	1.96	1.69
CGSP method	Mean	2.23	2.72
	SD	1.06	1.68

Motion Reconstruction

Motion reconstruction experiments were conducted with the omnidirectional camera for the cases of pure translation and pure rotation. The similarity of the results for the fisheye camera calibration using the two calibration methods mean that any motion reconstruction experiments would likely be too similar for the purpose of comparison.

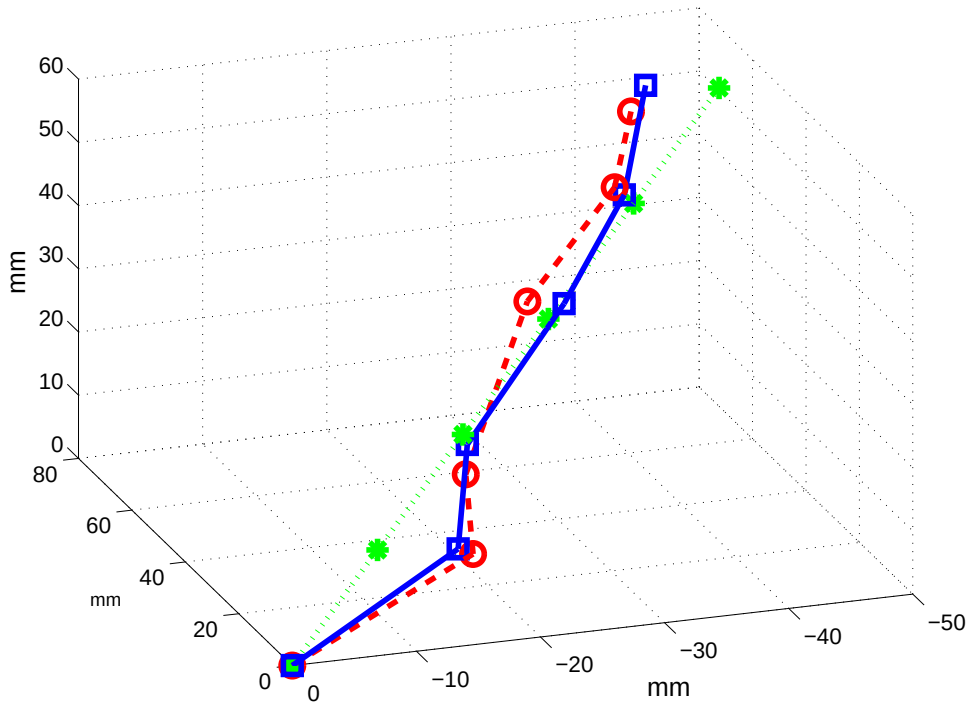
The experimental setup consisted of a 3D calibration target (two orthogonal planar chessboard grids) rigidly mounted on a stage capable of controlled rotation and translation. For the translation experiment, the object was translated $100mm$ in steps of $20mm$, and for the rotation experiment it was rotated through 110^o in steps of 10^o. Point matches were extracted from the images using the CELECT corner detection and ordering method proposed in Chapter 3. These were used to index the Plücker matrix lookup tables for each calibration method to get the corresponding ray direction information. The essential matrix, E, between each image pair was linearly estimated using the ray-based epipolar constraint (Ramalingam et al., 2006)

$$L_2^T E L_1 = 0 \tag{4.25}$$

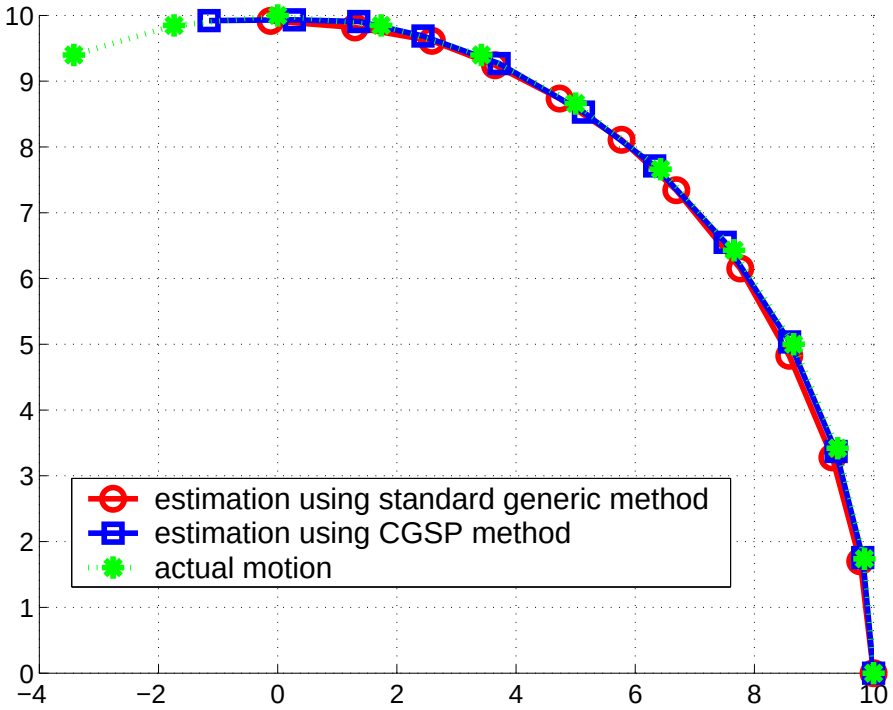
where L_1 and L_2 are the first 3 components of the Plücker vectors derived from the Plücker matrices. Rotations and translations are extracted from the essential matrices using the method of Nister (2003). The motion reconstruction results are shown in Fig. 4.10. It can be seen that the motion estimated with the CGSP method is closer to linear in the case of translation, and closer to the ground truth value of 110^o in the case of rotation, than for the standard generic method. For visualisation purposes the differences between the average translation vector and the estimated translation vectors are scaled by 4.

4.4 Discussion

This chapter addresses generic camera calibration for cameras that have a single centre of projection, and proposes an efficient method of generic calibration, the CGSP method, for such cameras. The standard generic method of Sturm and Ramalingam (2003) was developed such that both central and non-central cameras can be calibrated within the same framework. Consequently, estab-



(a)



(b)

Fig. 4.10: (a) Translation and (b) rotation reconstruction using the calibration data from the standard generic method and the CGSP method.

lished parametric calibration techniques are not utilised for the central case. An improved central generic calibration is proposed that takes advantage of the geometric constraints resulting from a single centre of projection in order to enable the utilisation of such techniques. As a result the CGSP calibration method can obtain greater calibration accuracy and robustness than the standard generic method.

Firstly, Section 4.1 proposes a novel linear estimation stage for use in the CGSP calibration method. The linear estimation method enables simple calibration algorithms to be used instead of the more complicated method that has previously been promoted, allowing the use of any pinhole calibration technique for determining the camera centre for any type of central camera. Active grids play an important role in the proposed linear estimation. Section 4.1.1 describes how an active grid can be used as a synthetic image plane during calibration. Exact pinhole images are formed on this synthetic plane by intersecting camera rays with it, and then by decoding the intersection locations from the active grid data. The task of determining the camera centre is thus converted to a simple pinhole calibration problem, for which many solutions exist. Section 4.1.2 details the CGSP method's linear estimation process when applying the pinhole calibration method of Sturm and Maybank (1999). Bundle adjustment is a standard treatment for refining parameters in any calibration process, and Section 4.1.4 discusses the application of bundle adjustment to the linearly estimated camera centre and pose estimates. Comparative results for the standard generic and CGSP linear estimation stages are presented in Section 4.1.5, where it is shown that the camera centre and initial grid poses estimated using the CGSP method without bundle adjustment are more accurate in the presence of noise than the equivalent parameters estimated using the standard method.

Secondly, an alternative pose estimation stage is proposed in Section 4.2 for application in the CGSP calibration method. Pose estimation of additional grids is required to expand the calibrated FOV. Section 4.2.1 details how perspectivities can be applied to data from a synthetic, unrealised, image plane in order to determine a linear solution for pose that utilises a standard pinhole pose estimation technique. Such a solution enables linear re-estimation of pose in a RANSAC framework, which is not possible using the pose algorithm of the standard generic method. The CGSP and standard generic pose estimation methods, both incorporated in RANSAC frameworks, are evaluated

against each other for both simulated and real data in Section 4.2.2, with the conclusion that the CGSP method's pose estimation performs better in terms of both accuracy and robustness to noise.

The complete CGSP method is evaluated against the standard generic method in Section 4.3.2, where each method is used to calibrate both an omnidirectional camera and a fisheye camera. When a ray-point error metric is applied to the calibrated cameras rays involved in the linear estimation stage, it is seen that the mean residual error, both before and after the application of bundle adjustment, is always less for the CGSP calibration than for the standard calibration, which agrees with the simulated results. The mean residual error before bundle adjustment is approximately 16 times larger for the standard method than for the CGSP method when applied to the omnidirectional camera, and approximately 2.2 times larger when applied to the fisheye camera. Qualitative and quantitative improvement in accuracy is particulary noticeable for the omnidirectional camera. The distortion correction experiments demonstrate the superior performance of the CGSP method for the omnidirectional camera, whereas the results for the fisheye camera are approximately equivalent for both methods. Motion reconstruction experiments are also undertaken for the omnidirectional camera, with the results again showing the accuracy of the CGSP method to be superior to that of the standard method.

It is concluded from the experimental results, both for synthetic and real data, that the CGSP method should be used instead of the standard generic method in order to achieve the best results for the generic calibration of central cameras.

A prerequisite to the application of the CGSP calibration method is knowledge of camera centrality. For completely non-central cameras, such as stereo rigs, the CGSP method is certainly not applicable. However, for near-central cameras, for example catadioptric configurations with misalignment between the camera and mirror, the approach taken by Ramalingam et al. (2005*a*) can be adopted, whereby the camera is initially treated as central. Therefore the CGSP method can be applied first, followed by a non-central bundle adjustment scheme to relax the centrality constraint.

Chapter 5

Calibration of Flexible Mirror Imaging Systems

There is currently a trend towards new camera designs for use in such areas as surveillance, mobile robot navigation and virtual reality. These new camera designs achieve desirable image benefits over perspective cameras. Chapter 4 deals with non-conventional central camera systems, which can provide significant benefits over conventional cameras in terms of large FOVs that allow greater persistence of vision, and specific resolution properties such as equiresolution. However, the area of flexible catadioptrics, in which the mirrored surface used is flexible, is only recently receiving attention. Flexible mirror imaging systems consisting of a perspective camera viewing a scene reflected in a flexible mirror provide unique camera characteristics that enable a dynamic FOV and alterable allocation of the camera sensor area resources to the imaged scene. Thus flexible mirror cameras can form vision systems that have not previously been realisable. However, the benefits that they provide are balanced by an important disadvantage of such systems, which is that they are difficult to calibrate due to the vast range of possible mirror shapes and due to the flexible nature of the system. This chapter proposes the fundamentals of a dynamic calibration approach for flexible mirror imaging systems, leading to the development of the SPFC^2 method – Scene Point based Flexible mirror Camera Calibration method.

A prerequisite for calibration is a suitable model of the flexible mirror camera. With the exception of the general model, none of the camera models

discussed in Chapter 2 is applicable in this case due to the non-centrality and to the unknown, generally non-quadric, mirror shape of flexible mirror imagers. Considering the general model, the direct mapping between every image pixel and associated camera ray would require to be updated after each mirror deflection if applied to a flexible mirror camera. A model with a small number of parameters and that can cope with a continuum of mirror surface shapes would be more suitable. This is achieved by directly modelling the flexed mirror surface using some parametric surface description. Section 5.1 presents two such camera models, one of which models the mirror as a curve in the plane for constrained deflections, and the second of which models the mirror surface using a B-spline surface. Therefore the complete camera model is parametric, where the parameters are the perspective camera parameters and the parameters describing the mirror surface shape.

The focus of this chapter is on a method for effectively calibrating the above models. The principles of an alternative calibration method to Kuthirummal and Nayar (2007) are established, where the aim is to dynamically update the calibration rather than attempt to recalibrate completely after each mirror deflection. Key to the proposed SPFC^2 method is the use of scene points, rather than points on the mirror, for calibration. The SPFC^2 method takes its inspiration in part from the field of mirror design for catadioptric systems. Swaminathan, Nayar and Grossberg (2003) presented a linear method for determining the catadioptric mirror surface shape necessary to implement a desired scene to image map. That method employs constraints on the incident and reflected ray directions in order to determine the parameters of a B-spline surface model of the mirror. Based on the method of Halstead et al. (1996), the surface is determined directly from the set of surface normals. Section 5.1.2 describes how this method is applied to flexible mirror camera calibration. There are scale ambiguities in this surface estimation, and so an additional constraint based on the fixed points of the mirror is required in order to completely determine the mirror surface.

Calibration in the SPFC^2 method is achieved by separately calibrating the perspective camera, determining the flexible mirror shape, and determining the relative orientations of the perspective camera and the flexible mirror. The SPFC^2 calibration process, described in Section 5.2, consists of an initial primary calibration stage followed by in-service dynamic calibration. Perspective camera calibration and the determination of mirror position are performed in

the primary stage. Then dynamic calibration implements a linear surface approximation to initialise a non-linear minimisation step, the result of which is the estimate of the mirror surface shape. Scene information required for dynamic calibration is obtained by imaging a calibration grid both before and after the mirror deflection.

In comparison to the only existing flexible mirror camera calibration method (Kuthirummal and Nayar, 2007), the SPFC^2 method has the significant advantage of a much simpler calibration setup. It does not require a stereo rig for surface shape estimation, nor is there a large and laborious data collection procedure required. The SPFC^2 method is adaptive where that of Kuthirummal and Nayar (2007) is exhaustive. In addition, the mirror shapes that it can calibrate are not limited to the set of pre-calibrated deformations nor is there a requirement for the entire mirror boundary to be visible in the image since calibration is performed using information from reflected scene points. On the other hand, the SPFC^2 method requires the use of a calibration grid, and calibration is not available in real time unlike the existing method (Kuthirummal and Nayar, 2007).

The SPFC^2 method is most similar in spirit to the corneal reconstruction method of Halstead et al. (1996), although the methods differ in their implementations and calibration configurations, and they have different application domains. In Halstead et al. (1996) the 3-space scene points, whose locations are fixed and known, lie on the inner surface of a special purpose conical calibration object that forms part of the videokeratograph configuration for corneal measurement. In contrast, for the SPFC^2 method, the scene point locations in 3-space are not initially known and are not fixed relative to the perspective camera, and so they must be estimated based on the previous calibrated mirror surface. In addition, the locations of the scene points vary between applications of dynamic calibration in the SPFC^2 method. Initialisation of the surface refinement in Halstead et al. (1996) is performed by guessing a corneal shape, which is acceptable since the corneal shape is approximately standard. For the purpose of flexible mirror imager calibration, there is no approximate shape for the surface to be refined, and so an alternative method is required for initialisation. Also, for the application in Halstead et al. (1996), the primary measurement to be recovered is surface shape, and so the accurate recovery of surface scale is not discussed. In contrast, scale information is important for calibration and its determination is detailed in the SPFC^2 method.

Results for a simulated camera based on the reduced model, presented and discussed in Section 5.3.1, demonstrate the robustness of the SPFC^2 method to variations in camera and scene configurations. The applicability of both the reduced model and the full model for a real flexible mirror camera is examined in Sections 5.3.2 and 5.3.3, in which extensive experimental evaluation is conducted. Motion and scene reconstruction experiments are additionally conducted for the full model. Finally, the application of the SPFC^2 method, with modifications, to the task of flexible mirror self-calibration is investigated in Section 5.4.

5.1 Flexible Mirror Camera Model

The flexible mirror imaging system dealt with in this chapter consists of a perspective camera imaging a scene reflected in a thin, flexible, developable mirror surface. A developable surface is a surface that can be flattened to a planar surface without any stretching or tearing occurring (Liang et al., 2008). Experimentation is restricted to such surfaces since non-developable flexible mirrors are difficult to realise in practice, although the theory of flexible mirror imaging calibration that follows is equally applicable to non-developable mirror surfaces. Furthermore, one edge of the flexible mirror is assumed to be fixed in position and in first derivative relative to the perspective camera. This is achieved by clamping the mirror at one edge. Apart from providing a convenient method of holding the mirror, it will be seen that this constraint is sufficient to remove the ambiguity in the location of the mirror in the camera coordinate system.

Reflection of light rays from specular surfaces is governed by the law of reflection, which states that the angle of incidence is equal to the angle of reflection. Knowledge of the surface normal at any point on the surface allows the reflected ray at that point to be calculated for any incident ray. Referring to Fig. 5.1, reflected light rays entering the perspective camera are termed viewing rays and are represented by unit vectors $\mathbf{V}_l$. The corresponding light rays that fall on the mirror from the scene are called scene rays and are represented by unit vectors $\mathbf{V}_r$. For any given mirror surface shape, the flexible imager can be modelled by the general camera model, which maps image pixels (u, v) to scene rays $\mathbf{V}_r$ using a look up table. However, for efficiency reasons a para-

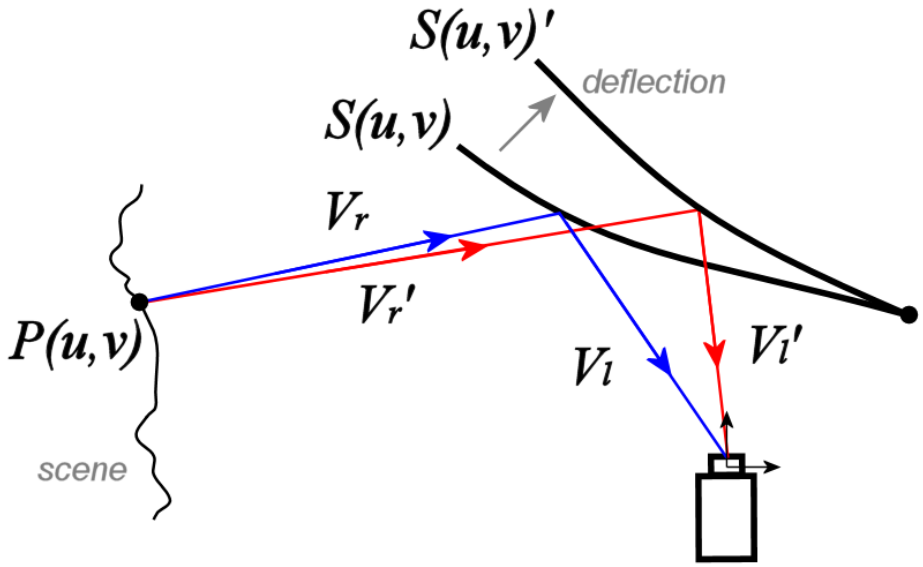


Fig. 5.1: Schematic of a flexible mirror imaging system.

metric camera model is desirable, and so the look-up table is replaced by a model of the flexible mirror surface. Similarly to Micusik and Pajdla (2004), rays anchored to the mirror surface rather than rays on the mirror caustic are used to represent the camera. By using the mirror surface rather than the caustic the computation of the Jacobian is not necessary, and high-order surface parameterisations required to describe caustics resulting from slightly non-central flexible mirror imager configurations are avoided.

5.1.1 Mirror Shape Model

Two different mirror surface shape models are examined. The reduced model approximates the mirror surface as a curve in the plane. This imposes restrictions on the mirror configuration and deflection, but it is beneficial for simulation purposes since synthetic mirror shapes can be formed easily. The full model describes the mirror as a general surface allowing unconstrained mirror deflections, and so it is more useful in practice. For this reason the mirror shape estimation, the application of the positional constraint, and the method of calibration are described with respect to the full mirror model. However, the theory and the calibration are equally applicable to the reduced model.

Reduced Model

The flexible mirror can be approximated by a thin plate deflection model. Assume the fixed edge of the mirror is vertical with respect to the perspective camera viewing it, and is perpendicular to the camera's principal axis. Then if deflection of the mirror occurs in a plane perpendicular to the fixed mirror edge, a reduced model of the mirror shape consisting of a 2D planar curve can be considered. From structural mechanics beam deflection theory, a cantilever beam with isotropic material properties, fixed in position and first derivative at one end and undergoing small deflections, can be modelled by

$$v = \frac{Px^2}{6EI}(3m - x) \tag{5.1}$$

where the deflection v is determined by the applied force P, the material properties E and I, the distance m from the fixed edge at which the force is applied, and the distance x from the fixed edge, where $x \leq m$. Clearly then the mirror should be modelled by a curve of at least $3rd$ order. The reduced mirror shape model is

$$S(u) = au^3 + bu^2 + cu + d \tag{5.2}$$

where $S(u)$ is the mirror shape curve, u is the horizontal image pixel coordinate, and $\{a,\ b,\ c,\ d\}$ are the model parameters. This model and the beam deflection equation are physically accurate for small deflections of the mirror for which $\left(\frac{dS(u)}{du}\right)^2$ is much less than unity (Gere and Timoshenko, 1991). Larger deflections have more complicated shapes that are described by elastica theory.

Full Model

Whilst the reduced model indicates the type of shape that the mirror model can assume in a planar scenario, it is limited in that it can be only applied to single dimensional flexing resulting from a single-point deflection applied at the free edge of the mirror. In order to allow for greater generality a B-spline surface model is used to represent the mirror, similar to Kuthirummal and Nayar (2007) and to Halstead et al. (1996). The form of the reduced model suggests that splines of at least third order are required, but because multiple B-splines are used in the full model, the order of the splines can be reduced.

Lower order B-splines have fewer spline coefficients and so the number of image points required for calibration is less. In this thesis quadratic B-spline basis functions are chosen to model the mirror surface across open uniform knot vectors.

As in Swaminathan, Nayar and Grossberg (2003) and Kuthirummal and Nayar (2007) the complete mirror surface is described parametrically by the distances, $D(u, v)$, of the mirror points from the perspective camera centre, described by a B-spline model, as measured along the viewing rays $\mathbf{V}_l$ of the primary optics. The perspective camera centre is assumed without loss of generality to be at the world coordinate origin. The full mirror model, $S(u, v)$, is consequently given by

$$S(u, v) = D(u, v)\mathbf{V}_l(u, v) \tag{5.3}$$

were

$$D(u, v) = \sum_{i=1}^{K_f} \sum_{j=1}^{K_g} c_{ij} f_i(u) g_j(v) \tag{5.4}$$

$$= \mathbf{f}(u)^T C \mathbf{g}(v) \tag{5.5}$$

and

$$\mathbf{f}(u) = \begin{bmatrix} f_1(u) & f_2(u) & \dots & f_{K_f}(u) \end{bmatrix}^T \tag{5.6}$$

$$\mathbf{g}(v) = \begin{bmatrix} g_1(v) & g_2(v) & \dots & g_{K_g}(v) \end{bmatrix}^T \tag{5.7}$$

u and v are the horizontal and vertical image pixel coordinates, respectively, c_{ij} are the B-spline surface coefficients, $f_i(u)$ and $g_j(v)$ are the corresponding spline basis functions, and K_f and K_g are the number of basis functions in the u and v directions, respectively. K_f and K_g are determined by the elements of the knot vectors. In the full mirror model knot vectors are regenerated after every calibrated change in the mirror shape so as to span the area of the mirror visible in the image. In this way the complete mirror surface is always modelled by the same number of basis functions independently of changes in its apparent size due to deflections. As a consequence of its bi-dimensionality the full mirror model does not place any restrictions on the orientation of the fixed mirror edge relative to the camera principal axis.

5.1.2 Mirror Shape Estimation

The mirror surface is estimated from a set of linear equations as in Swaminathan, Nayar and Grossberg (2003) and in Halstead et al. (1996). For known $\mathbf{V}_l s$ and $\mathbf{V}_r s$, the unit surface normals, $N(u,v)$, are given by

$$N(u,v) = \frac{\mathbf{V}_l(u,v) - \mathbf{V}_r(u,v)}{||\mathbf{V}_l(u,v) - \mathbf{V}_r(u,v)||} \tag{5.8}$$

The tangents to the surface are given by the first derivative of the surface. Enforcing orthogonality between normals and tangents leads to the following

$$\frac{\partial S(u,v)}{\partial u} \cdot N(u,v) = 0 \tag{5.9}$$

$$\frac{\partial S(u,v)}{\partial v} \cdot N(u,v) = 0 \tag{5.10}$$

Combining Eqns. 5.3, 5.9 and 5.10 gives

$$\left(D(u,v)\frac{\partial \mathbf{V}_l}{\partial u} + \mathbf{V}_l \frac{\partial D(u,v)}{\partial u} \right) \cdot N(u,v) = 0 \tag{5.11}$$

$$\left(D(u,v)\frac{\partial \mathbf{V}_l}{\partial v} + \mathbf{V}_l \frac{\partial D(u,v)}{\partial v} \right) \cdot N(u,v) = 0 \tag{5.12}$$

Expanding out with the B-spline basis functions and coefficients gives

$$\frac{\partial \mathbf{V}_l}{\partial u} \cdot N(u,v) \left(\mathbf{f}(u)^T C \mathbf{g}(v) \right) + \mathbf{V}_l \cdot N(u,v) \left(\frac{d\mathbf{f}(u)^T}{du} C \mathbf{g}(v) \right) = 0 \tag{5.13}$$

$$\frac{\partial \mathbf{V}_l}{\partial v} \cdot N(u,v) \left(\mathbf{f}(u)^T C \mathbf{g}(v) \right) + \mathbf{V}_l \cdot N(u,v) \left(\mathbf{f}(u)^T C \frac{d\mathbf{g}(v)}{dv} \right) = 0 \tag{5.14}$$

which can be rewritten as

$$\left[\left(\frac{\partial \mathbf{V}_l}{\partial u} \cdot N(u,v) \right) \mathrm{vec}\left(\mathbf{f}(u)\mathbf{g}(v)^T \right)^T + (\mathbf{V}_l \cdot N(u,v)) \, \mathrm{vec}\left(\frac{d\mathbf{f}(u)}{du} \mathbf{g}(v)^T \right)^T \right] \mathrm{vec}(C) = 0 \tag{5.15}$$

$$\left[\left(\frac{\partial \mathbf{V}_l}{\partial u} \cdot N(u,v) \right) \mathrm{vec}\left(\mathbf{f}(u)\mathbf{g}(v)^T \right)^T + (\mathbf{V}_l \cdot N(u,v)) \, \mathrm{vec}\left(\mathbf{f}(u) \frac{d\mathbf{g}(v)^T}{dv} \right)^T \right] \mathrm{vec}(C) = 0 \tag{5.16}$$

where vec() is the vectorise operator that converts a matrix to a vector by stacking the matrix columns. From the pinhole model for the perspective

camera $\mathbf{V}_l$ is given by $\frac{[u\ v\ f]^T}{|||u\ v\ f]|||}$, where f is the camera focal length, and thus

$$\frac{\partial \mathbf{V}_l}{\partial u} = \frac{1}{\sqrt{u^2+v^2+f^2}^3} \begin{bmatrix} v^2+f^2 \\ -uv \\ -uf \end{bmatrix} \tag{5.17}$$

$$\frac{\partial \mathbf{V}_l}{\partial v} = \frac{1}{\sqrt{u^2+v^2+f^2}^3} \begin{bmatrix} -uv \\ u^2+f^2 \\ -vf \end{bmatrix} \tag{5.18}$$

Each image point and corresponding mirror surface normal allow one of each of Eqns. 5.15–5.16 to be formed. By stacking such equations into a matrix a homogeneous equation of the form $Ax = 0$ is obtained. This can then be solved using the SVD with equilibration to determine the B-spline coefficients. Each $\{\mathbf{V}_l, \mathbf{V}_r\}$ pair contributes two independent equations to the data matrix A, and provides a linear constraint on $(k+1)^2$ B-spline blending functions, where k is the degree of the blending functions. The rank of A depends on the number of $\{\mathbf{V}_l, \mathbf{V}_r\}$ pairs, n, for which data is available. A has dimension $2n \times K_f K_g$, and must have rank $K_f K_g - 1$ in order to enable a non-trivial solution for the blending functions. The minimum number of $\{\mathbf{V}_l, \mathbf{V}_r\}$ pairs necessary to determine the mirror surface shape is therefore $\frac{K_f K_g}{2}$. In addition, $\{\mathbf{V}_l, \mathbf{V}_r\}$ pairs for a minimum of $ceil\left(\frac{K_f}{k+1}\right) \times ceil\left(\frac{K_g}{k+1}\right)$ distinct knot regions are required, and in this case the $\{\mathbf{V}_l, \mathbf{V}_r\}$ pairs must be maximally spread across the B-spline surface. However, in order to improve estimation accuracy in the presence of noise, $\{\mathbf{V}_l, \mathbf{V}_r\}$ pairs for every (non-zero) knot region should be included in matrix A.

5.1.3 Incorporating Mirror Positional Constraint

The above solution for the B-spline coefficients shows that a surface solution can be obtained once the surface normals are known. However, the solution is only up to scale as the mirror position has not been fixed. In fact two ambiguities can arise during surface estimation depending on the inputs to the estimation - mirror depth ambiguity and mirror shape ambiguity.

Firstly, ambiguity arises in the linear equations of the previous section from the fact that the position of the mirror in the world coordinate frame is not fixed. The method of mirror shape estimation presented in Section 5.1.2 is solely

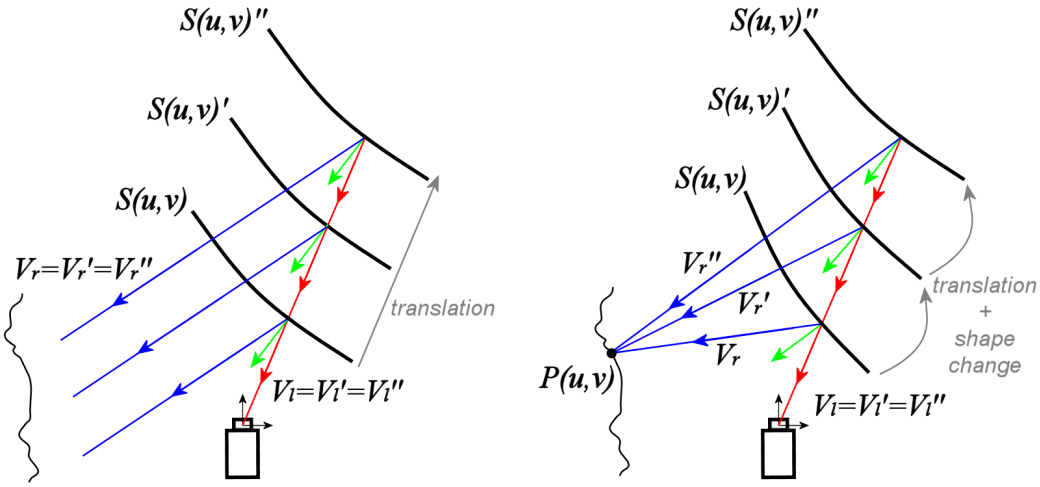


Fig. 5.2: (a) Mirror depth ambiguity; (b) mirror shape ambiguity.

based on viewing and scene rays, $\mathbf{V}_l$ and $\mathbf{V}_r$, from which mirror surface normals and finally B-spline coefficients are determined. Referring to Fig. 5.2(a), without a fix on the 3-space position of the scene points the mirror can translate along the $\mathbf{V}_l$s whilst the surface normals, and consequently the surface shape, remain fixed. When either a scene point or a mirror surface point is known in the world coordinate system, the ambiguity in the distance of the mirror surface from the camera, given by $D(u, v)$, can be removed through a direct scaling of C.

The second ambiguity relates to the situation where the $\mathbf{V}_l$s and scene points are known but the $\mathbf{V}_r$s are unknown. This occurs during dynamic calibration, described in Section 5.2.2. In this case, the mirror surface can translate along the viewing rays and continue to reflect the scene points along the $\mathbf{V}_l$s by simultaneously changing shape, as illustrated in Fig. 5.2(b). As the distance of the mirror from the perspective camera increases, the $\mathbf{V}_r$ directions must alter to keep the fixed scene points visible along the $\mathbf{V}_l$s. This ambiguity can not be removed by scaling C since the mirror surface shape alters for different mirror positions. The two ambiguities can be seen to be related by considering the $\mathbf{V}_r$s in the mirror depth ambiguity to be meeting in a point at infinity.

Both ambiguities can be removed by incorporating the constraint on the mirror position described in Section 5.1 into the mirror shape estimation equations. Note that in Halstead et al. (1996) an interpolation constraint is imposed

in the estimation equations for the purpose of constraining the solution set, but its importance and its connection to the depth and shape ambiguities described above are not discussed there. Assume that a solution for the mirror surface and scale is available for some position and deflection of the mirror (see Section 5.2). In this case, for an image point (u_e, v_e) viewing a fixed point on the mirror the spline value $D(u_e, v_e)$ is known. This value is constant for all mirror deflections so that after a deflection of the mirror the spline equation associated with (u_e, v_e) is

$$D(u_e, v_e) = \mathbf{f}'(u_e)^T C' \mathbf{g}'(v_e) \tag{5.19}$$

$$= \text{vec}\left(\mathbf{f}'(u_e)\mathbf{g}'(v_e)^T\right)^T \text{vec}(C') \tag{5.20}$$

where the prime indicates new spline coefficients and basis functions. A least squares solution for C' can thus be extracted in a subspace spanned by $(K_f K_g - 1)$ basis vectors as

$$\text{vec}(C') = \mathbf{c}_P + C_B \, \mathbf{\Phi} \tag{5.21}$$

where $\mathbf{c}_P$ is the particular solution, C_B is a matrix of basis vectors, and $\mathbf{\Phi}$ is the new vector of unknowns. Finally, C' is incorporated into Eqns. 5.15–5.16, and after some rearrangement the new equations to be solved for the flexed mirror surface that incorporate the positional constraint are

$$E_u C_B \, \mathbf{\Phi} = -E_u \mathbf{c}_P \tag{5.22}$$

$$E_v C_B \, \mathbf{\Phi} = -E_v \mathbf{c}_P \tag{5.23}$$

where

$$E_u = \left(\frac{\partial \mathbf{V}_l}{\partial u} \cdot N(u,v)\right) \text{vec}\left(\mathbf{f}'(u)\mathbf{g}'(v)^T\right)^T + \left(\mathbf{V}_l \cdot N(u,v)\right) \text{vec}\left(\frac{d\mathbf{f}'(u)}{du}\mathbf{g}'(v)^T\right)^T \tag{5.24}$$

$$E_v = \left(\frac{\partial \mathbf{V}_l}{\partial u} \cdot N(u,v)\right) \text{vec}\left(\mathbf{f}'(u)\mathbf{g}'(v)^T\right)^T + \left(\mathbf{V}_l \cdot N(u,v)\right) \text{vec}\left(\frac{d\mathbf{f}'(u)}{du}\mathbf{g}'(v)^T\right)^T \tag{5.25}$$

A solution for $\mathbf{\Phi}$ can be determined by stacking Eqns. 5.22–5.23 into a matrix and solving using standard least squares techniques, as in Section 5.1.2. The B-spline coefficients are then available from Eqn. 5.21. A single fixed point on the mirror is sufficient to solve for scale, although for the mirror model in this thesis the fixed mirror points are the points along the fixed edge of the mirror.

For the experiments in Section 5.3 the fixed imaged points along the clamped edge of the mirror are manually selected, and subsequently refined subpixelly by line-fitting. Refinement is necessary in order to prevent small errors in the selected fixed points inducing large errors in the estimated surface.

5.2 Flexible Mirror Camera Calibration – the SPFC^2 Method

The goal of calibration is to determine the mapping between viewing rays $\mathbf{V}_l$ and scene rays $\mathbf{V}_r$. This mapping is derived from the mirror surface by applying the law of reflection, so that calibration of the flexible mirror imager reduces to the estimation of the mirror surface. The SPFC^2 method consists of two stages: (1) calibration of primary optics and determination of initial mirror surface, and (2) dynamic updating of calibration after the mirror is flexed. The primary calibration stage is only executed once, whereas the dynamic stage is applied after each mirror deflection.

5.2.1 Primary Calibration

Primary calibration aims to determine a complete calibration of the imager for some initial mirror position and shape. It employs two basic calibration techniques. Firstly the perspective camera is calibrated using a standard method such as that of Zhang (2000). This information is used to remove radial distortion from all subsequent perspective camera images so as to enable the $\mathbf{V}_l$s, which remain fixed, to be calculated as described in Section 5.1.2. The complete flexible mirror imager is then calibrated as a general camera for the initial mirror shape and position using the general calibration method described in Section 2.3.3. Although general calibration is a time consuming process, it only needs to be carried out once. As part of the general calibration, an active grid is imaged in two positions, with a $100mm$ translation between them. However, this only enables the determination of the $\mathbf{V}_r$s in the coordinate frame of the active grids. In order to transform the scene rays into the coordinate frame of the perspective camera, or equivalently the world coordinate frame, the world poses of the active grids used in general calibration, T_i^{AG}, are required. This

is achieved for each active grid position i by taking an image of the active grid reflected in a planar mirror placed between the perspective camera and the flexible mirror. A hard disk platter is utilised as a high quality planar mirror in this application, and is attached directly to a planar chessboard grid. The platter mirror pose, P^m, can be estimated from the planar grid to which it is attached by using standard homography based techniques (Sturm, 2000). Similarly, the virtual pose, P_i^v, of the active grid in position i is estimated directly from the active grid corners reflected in the platter mirror. The transformation, T^R, describing reflection in a planar mirror (Bonfire et al., 2006) is given by

$$T^R = \begin{bmatrix} I_{3\times3} - 2\mathbf{n}\mathbf{n}^T & -2d\mathbf{n} \\ 0_{1\times3} & 1 \end{bmatrix} \tag{5.26}$$

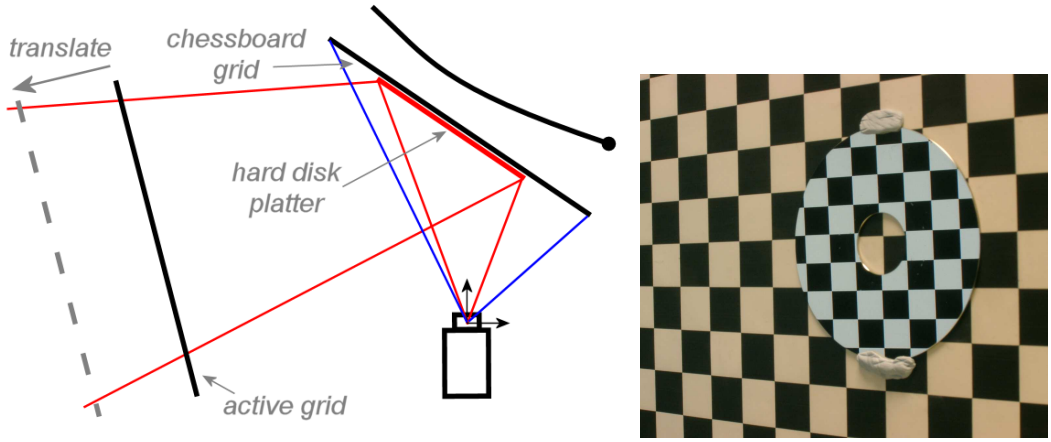
where $\mathbf{n}$ is the platter mirror unit plane normal extracted from P^m, and d is the distance of the platter mirror from the origin, again extracted from P^m. The active grid world poses are then given by

$$T_i^{AG} = T^R P_i^v \tag{5.27}$$

The translation component of P_m is reduced by the thickness of the platter mirror to account for the offset between its reflective surface and the calibration grid to which it is attached. The experimental configuration for pose recovery and a typical image used in the method are shown in Fig. 5.3. Note that if the poses of the active grid in each position are determined via planar reflection, as described above, it is not necessary to know the translation between the active grids. In that case the calibration setup is more straightforward as precise motion control is not required. However, the translation is used in the experiments conducted with real data in Section 5.3.3 in order to provide a check on the accuracy of the reflection based estimates of the active grid poses.

Finally, the intersection points of the viewing and scene rays, which should occur on the flexible mirror surface, are estimated so that the flexible mirror position can be determined in the camera coordinate space. The position estimates are found by selecting the points that minimise the distance to $\mathbf{V}_l$ and $\mathbf{V}_r$ for each corresponding pair of rays.

The initial mirror surface can now be determined using Eqns. 5.15 and 5.16. A fix on the overall scale of the calibration is made by incorporating the position estimates into the surface description, and is achieved by scaling the B-spline surface coefficients C so that $S(u, v)$ matches the position estimates.



(a) (b)

Fig. 5.3: (a) Method of determining the active grid pose from reflection; (b) example image of a hard disk platter mounted on a calibration grid and reflecting an active grid.

Interestingly, the primary calibration stage could be performed in a single step, by calibration of the perspective camera, if the initial mirror shape is known to be planar. The primary calibration stage could also be simplified by using static grids in place of active grids, since only a subset of all $\{\mathbf{V}_l, \mathbf{V}_r\}$ pairs are required to determine the B-spline surface model. However, the bias resulting from interpolation on chessboard grids may degrade the surface estimate.

The primary calibration stage is similar to the method for specular surface triangulation that was proposed by Bonfire et al. (2006). In that work, the position and orientation of a rigid specular surface is recovered using a single calibrated perspective camera viewing the reflection of a calibration target in at least two distinct positions. Similarly to primary calibration, they estimate the target poses using reflection in a planar mirror, although they show that the planar mirror's pose can be estimated indirectly if images of the reflection of at least three target poses are acquired. Their method can be applied to calibrate discontinuous specular surfaces, whereas for flexible mirror cameras the surface is known to be continuous. The SPFC^2 method can be seen as an extension of the work of Bonfire et al. (2006) to the case of flexible specular surfaces in which the surface estimate is dynamically updated as the surface deflects.

5.2.2 Dynamic Calibration

Dynamic calibration estimates the new mirror surface, resulting from deflection, in two steps. Firstly a linear solution is obtained based on the theory presented in Sections 5.1.2 and 5.1.3, and subsequently the solution is refined non-linearly by minimising a geometric scene error metric. It is important to understand the difference between the two steps. The inputs to the linear stage are the $\mathbf{V}_l$s and $\mathbf{V}_r$s and the position constraint, so no information about the depth of scene points is utilised. Thus it is applicable to potential self-calibration directly from a scene. For the non-linear stage, the inputs are the $\mathbf{V}_l$s and the scene points $\mathbf{P}(u, v)$ that are seen by the associated scene rays. In this step the directions of the scene rays are unknown, but the scene depths are available, so the non-linear step is unsuitable for self-calibration from an unknown scene. Self-calibration for flexible mirror imagers is discussed further in Section 5.4.

Any deflection of the mirror requires the calibration to be updated, which is equivalent to estimating the updated mirror shape, $S(u, v)'$. In Section 5.1.2 it is shown that knowledge of the scene and viewing rays is sufficient for reconstructing the mirror surface. The viewing rays $\mathbf{V}'_l$ for $S(u, v)'$ can be determined directly from the locations of the features in the image after deflection, as in Section 5.1.2, but the $\mathbf{V}'_r$s are unknown. Referring to Fig. 5.1, it is clear that if the scene points $\mathbf{P}(u, v)$ are at infinity, then the directions of the scene rays that see the same scene points are the same for all mirror shapes. Thus the approximation

$$\mathbf{V}'_r \approx \mathbf{V}_r \tag{5.28}$$

holds for distant scene points, and enables a linear estimation of the mirror surface using Eqns. 5.3, 5.9–5.10, and 5.22–5.23.

The linearly estimated result, $\hat{S}(u, v)'$, is used to initialise a non-linear estimation step that determines the final mirror surface estimate, $S(u, v)'$. The non-linear estimation refines the linear estimate to account for non-infinite scene points by minimising scene errors, but to do this some information about the scene is required. Such information is acquired by imaging a calibration grid, placed in the imager FOV, both before and after mirror deflection. The pose of this grid is estimated from the first image, which has already been calibrated (either in the primary calibration stage or in the previous application of dy-

namic calibration). A non-central pose estimation technique must be applied in order to estimate the pose. In the experiments presented in this chapter a modification of the method of Eqn. 2.26 is used that applies non-linear minimisation to solve for all grid corners simultaneously. A canonical metric grid with the estimated pose is subsequently fitted to the data so as to enforce grid point planarity and improve robustness. This enables a scene-to-image map to be formed from the 3-space locations of the canonical grid corner scene points, $\mathbf{P}(u, v)$. Non-linear estimation can then proceed by minimising the following geometric error metric

$$\Gamma(S(u,v)') = || \left[\mathbf{V}_r(u,v)' \cdot (\mathbf{P}(u,v) - S(u,v)')\right] \mathbf{V}_r(u,v)' + S(u,v)' - \mathbf{P}(u,v)|| \tag{5.29}$$

Γ, which is the ray-point distance metric proposed in Section 4.1.4, measures the perpendicular distance between $\mathbf{V}_r$s and corresponding $\mathbf{P}(u, v)$s. This results in the mirror surface estimate, $S(u, v)'$, for which the distance between scene rays and the scene points $\mathbf{P}(u, v)$ through which they should pass is minimised. Once the calibration has been updated to $S(u, v)'$, the mirror can be flexed again and dynamic calibration reapplied using the $\mathbf{V}'_r$s estimated in the last iteration. An iterative or non-linear method is necessary for accurately determining $S(u, v)'$ since no closed-form solutions exist for determining the reflection point of a scene point $\mathbf{A}$ in a non-spherical curved mirrored surface viewed from a point $\mathbf{B}$ (Roth and Black, 2006, Chen and Arvo, 2000).

A flow diagram outlining the SPFC^2 calibration process is shown in Fig. 5.4.

Two factors should be noted in relation to the magnitudes of deflections applied during dynamic calibration. Firstly, there is an inherent bias due to the infinity assumption that results in the linearly estimated mirror surface underestimating the magnitude of the mirror deflection. However, it is found that for small mirror deflections the linear estimate is sufficient for the convergence of the non-linear estimation. Secondly, mirror deflection causes feature points (imaged scene points) to be displaced between images, as can be clearly seen in Fig. 5.10. The larger the mirror deflection, the larger is the feature point displacement. This displacement of feature points should be considered when choosing the knot vectors, which in turn determine K_f and K_g. A minimum condition on the locations of feature points in knot regions in order to allow linear surface estimation is given in Section 5.1.2. Features must exist in at least the first and last knot interval in the u direction and in the v direction.

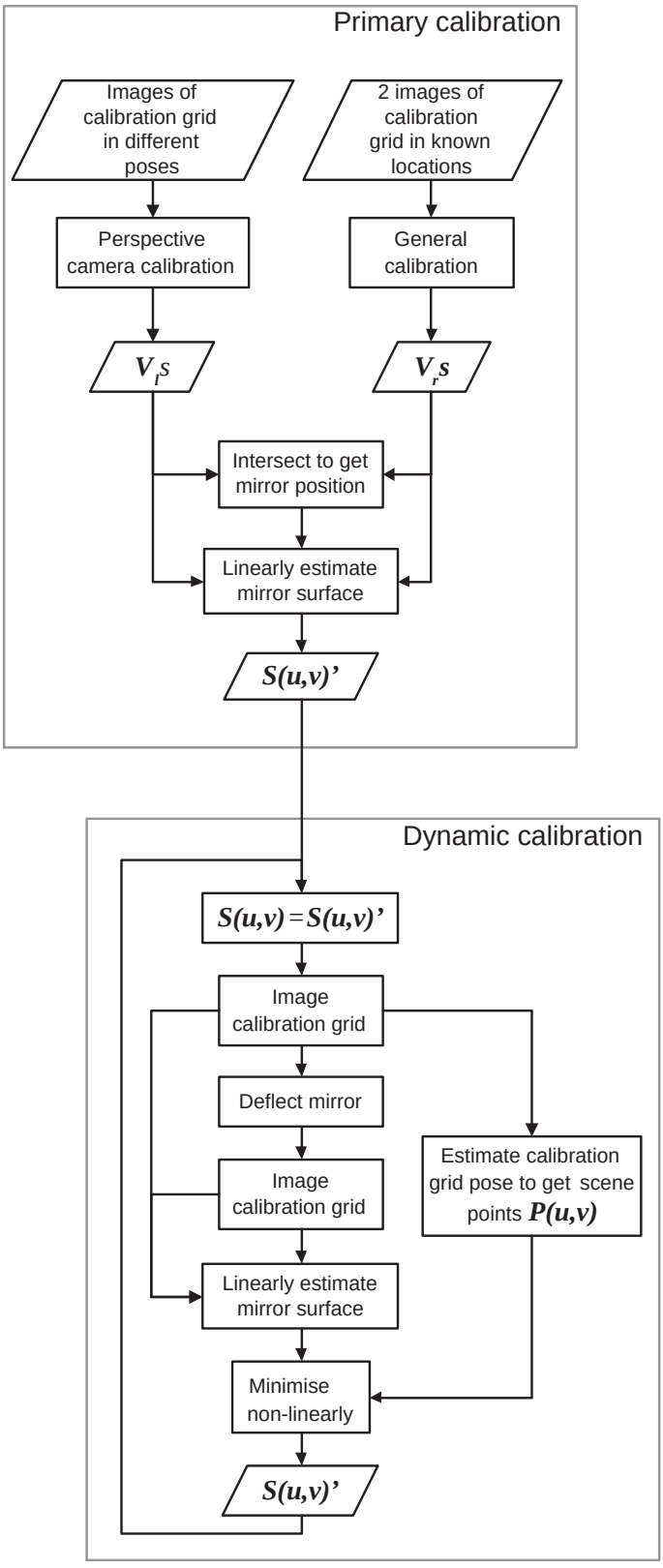


Fig. 5.4: Flowchart of the SPFC^2 method.

Considering the mirror surface as viewed in the image, the knot regions along the free mirror edge opposite to the constrained edge must contain scene features. It is at this edge that mirror deflections will cause the most displacement of features in the image. Thus, in order to meet the minimum condition when applying large deflections, the knot region sizes must be increased by reducing the number of knots in the knot vectors. However, a balance is required between accuracy and allowable deflection, as a more accurate surface estimate necessitates more basis function in the B-splines and thus more knots in the knot vectors. These competing requirements must be considered when selecting the knot vectors. Active selection of the knots could be accomplished by using the observed feature point displacements to determine the number of knots to be used in both u and v directions.

Due to both the linear estimate bias and the minimum condition on the existence of image features of scene points across deflected image pairs, small deflections are favoured and large mirror deflections should be carried out in stages, with dynamic calibration applied at each stage.

5.3 Experiments

Experiments were carried out using simulated data, to evaluate the effect on performance of configuration variations, and using real data, so as to characterise the overall performance of the SPFC^2 method. The reduced model is applied in the simulated experiments and in the initial real experiment, with the full flexible mirror camera model used in the subsequent real experiments.

5.3.1 Simulated Experiments

Experiments with simulated data were conducted using the reduced model of Eqn. 5.2 with single point deflection at the mirror edge, for variations in image noise, scene depth (distance between mirror and calibration grid) and mirror curvature. The simulation results are presented in Fig. 5.5, where the datapoints in each plot are the average of 100 random trials, and results are shown for both the linearly estimated surface and the final non-linearly refined surface.

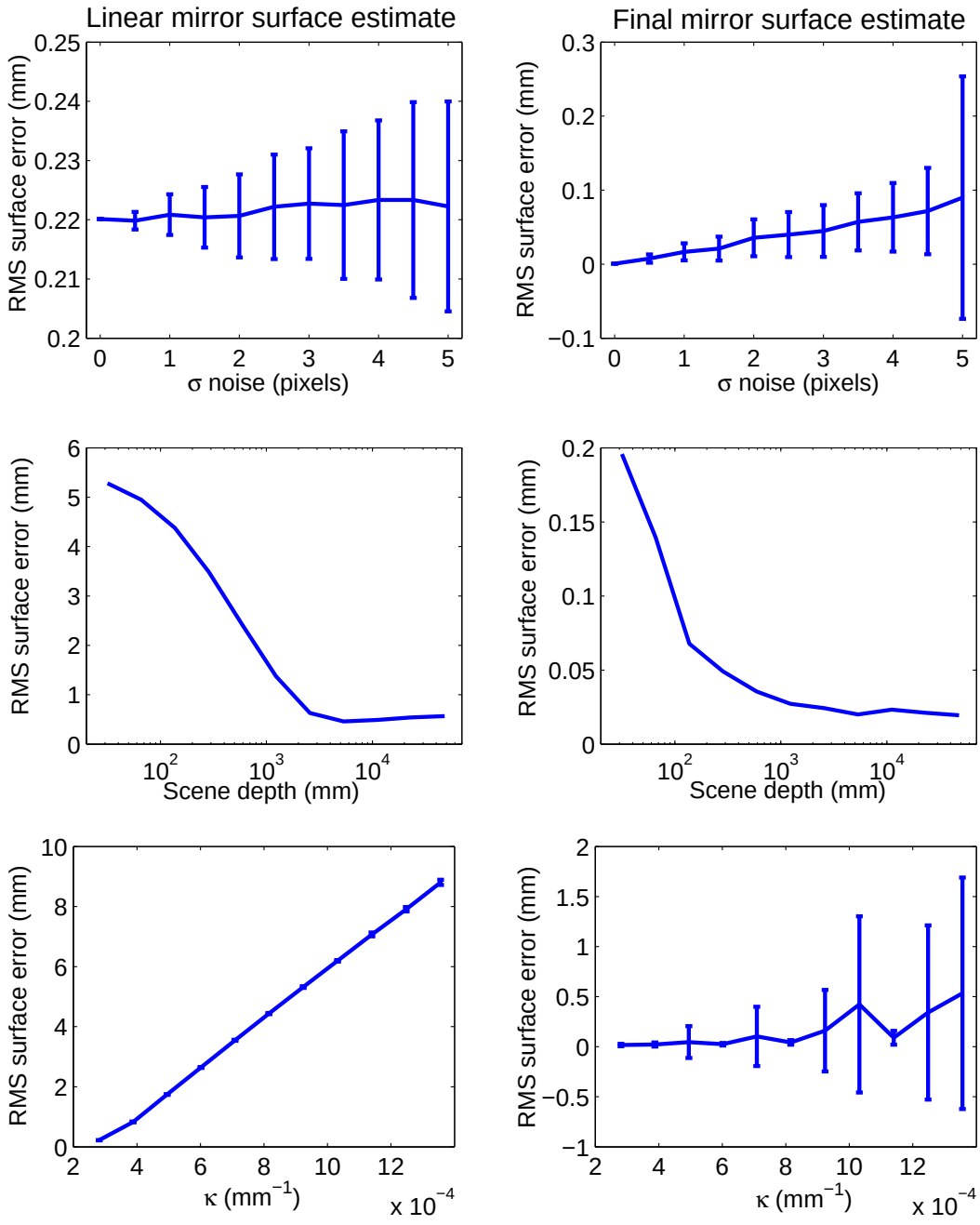


Fig. 5.5: Error in the linear and final SPFC^2 mirror surface estimates for varying parameters. The simulated camera has focal length $26mm$ and image size 516×656 pixels. The flexible mirror is $420mm$ in length and is oriented at 67.5^o to the camera. Unless otherwise stated, the mirror deflection is $16.7mm$, the scene depth is $1650mm$ and Gaussian image noise with 1 pixel standard deviation is added.

Row 1 shows the error plots for increasing additive Gaussian image noise. Additive image noise has a greater impact on the final surface estimate than on the linear estimate, since the error due to the infinite scene assumption far outweighs the error introduced by the noise. The final surface estimate error is linearly proportional to the noise. The error plots for increasing scene depth are shown in row 2. As expected, the error decreases as the calibration grid moves further from the mirror, in line with the infinite scene assumption. For large scene depths ($> 3000mm$) the final error becomes independent of the scene depth, indicating that the infinity assumption is an effective approximation in this range. The error plots for increasing mirror curvature, κ, are shown in row 3. It is seen that the error in the linear surface estimate increases with increasing curvature due to increasing weakness of the infinite scene assumption. The error in the final mirror surface estimate for variation in curvature is larger than the maximum error for either of the other two simulated experiments principally due to the poor linear estimate. The maximum value of $\left(\frac{dS(u)}{du}\right)^2$ for the simulated experiments is 0.036, which is much less than unity, and thus the deflections in the simulated experiments are within the range over which the reduced mirror model is applicable.

5.3.2 Real Experiments with Reduced Model

The SPFC^2 method was evaluated for a real flexible imaging system configured as described in Section 5.1. A flexible plastic mirror[1] was attached to an aluminium substrate in order to improve rigidity. Deflections were applied to the back of the substrate through a tensioned wire. Due to its inexpense, the flexible mirror exhibits quality defects in the form of both double reflections and surface aberrations. Exemplar mirror aberration is visible in Fig. 5.6, which shows a portion of a chessboard grid, located at a scene depth of approximately $5m$, reflected at the top edge of the flexible mirror. The aberration causes deviations from the expected scene ray directions, and so the detrimental impact on image quality increases when reflecting scene points that are at greater scene depths.

The mirror was deflected by approximately $7.98mm$ from a nominally planar position, with high-accuracy comparative reference measurements of the mirror

[1]Plastic wing mirror replacement mirror from www.carpointeurope.com

Fig. 5.6: Portion of a flexible mirror image showing aberration in the upper region due to poor mirror quality.

surface obtained for each position using a 3D laser scanner [2]. Fig. 5.7 shows the estimated and reference mirror surface before and after the mirror deflection. It can be seen that the primary calibration error is of similar magnitude to the error for the final mirror surface estimate after flexing, indicating that the accuracy of the dynamic calibration is very sensitive to the primary calibration. The primary calibration error can be attributed to the higher order components of the mirror surface that are not modelled by Eqn. 5.2, and to mirror defects. The distance from the mirror surface to the calibration grid was approximately $280mm$, and as a result the linear estimate of the flexed mirror surface in Fig. 5.7 is relatively inaccurate, although the final estimate is still good.

The feature points in the image will have both u and v coordinates. For the reduced model the v coordinates can simply be dropped so that the features are projected to 2D. The mean error between the laser scanned mirror surface and the estimated mirror surface using the reduced model is $0.13mm$.

5.3.3 Real Experiments with Full Model

Experiments were conducted with a similar flexible mirror imager configuration to that used in Section 5.3.2 in order to assess separately the primary and dynamic calibration stages for the full mirror surface model of Eqn. 5.5. The perspective camera used was an Olympus SP-510UZ 7.1 Mpixel digital camera. Deflections were applied to the back of the substrate using two optical

[2]The non-specular back surface of the mirror substrate was scanned instead of the mirrored surface, since a mirrored surface cannot be accurately measured by a laser scanner.

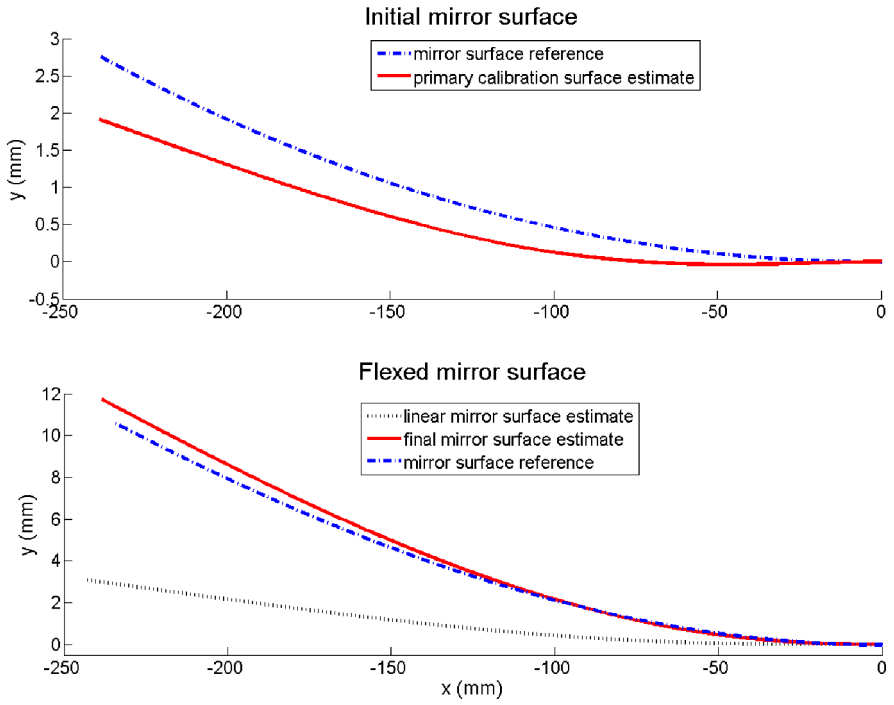


Fig. 5.7: Mirror surface estimates for real data, using the reduced model, before and after deflection.

translation stages in series, each having a resolution of $0.01mm$. The complete imager configuration is shown in Fig. 5.8. Since mirror deflection occurs principally in a plane perpendicular to the fixed mirror edge, $K_f = 7$ and $K_g = 3$ were chosen for the full model used in the experiments. A composite plot of the 21 resulting B-spline blending functions is shown in Fig. 5.9.

Images of a scene taken with the flexible mirror imager at each of the four $5mm$ stages of deflection applied in the real experiments are shown in Fig. 5.10. It is seen that the FOV for stage 3 is approximately double the FOV of the stage 0 configuration. The fixed edge of the mirror is on the right of the image, although due to the mirror clamping method, the exact fixed edge is not visible. This explains why there is some loss of FOV on the right of the images when progressing from stage 0 to stage 3. At all other mirror edges, a qualitative examination reveals that the non-linear image distortion rapidly increases with respect to image regions further from these edges. These edge effects are likely due to strains induced in the mirror material during the cutting stage after manufacture.

As discussed in Section 5.2.1, the mirror position is determined in primary calibration by intersecting the known viewing and scene rays. The mean and

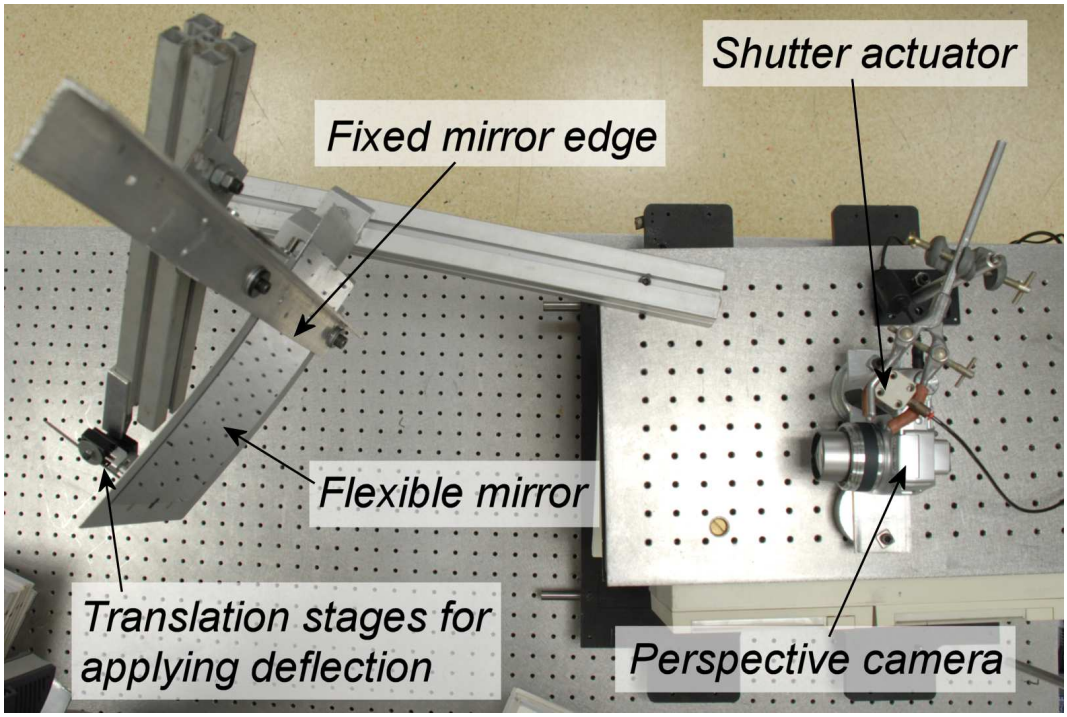


Fig. 5.8: Experimental configuration of the flexible mirror imager used in the real experiments with the full model.

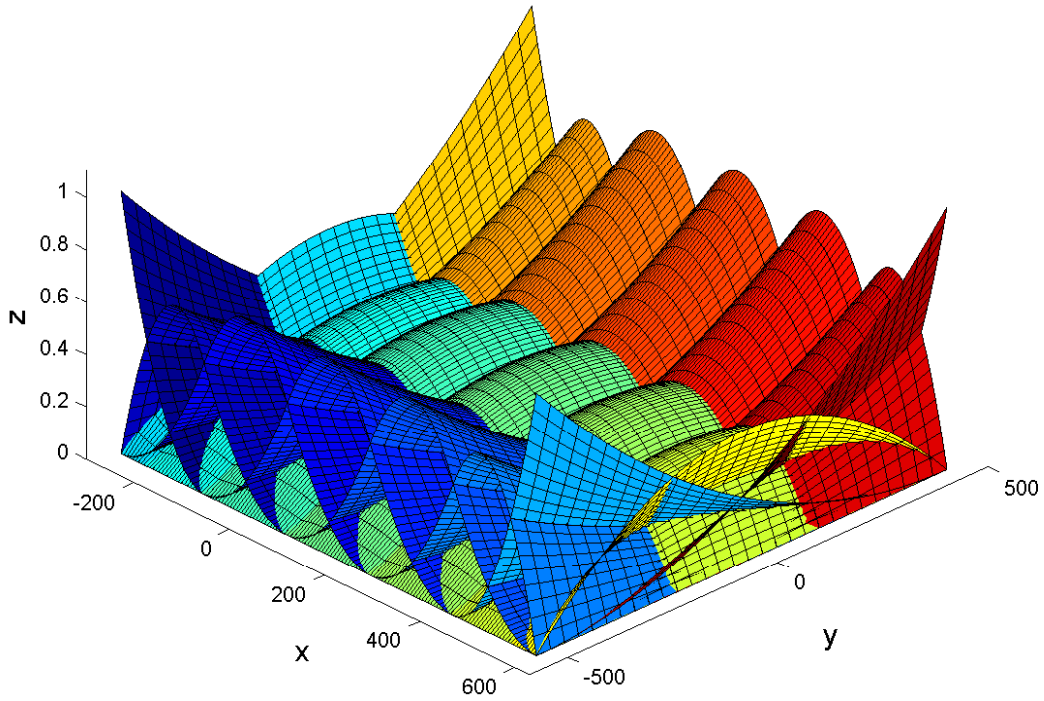
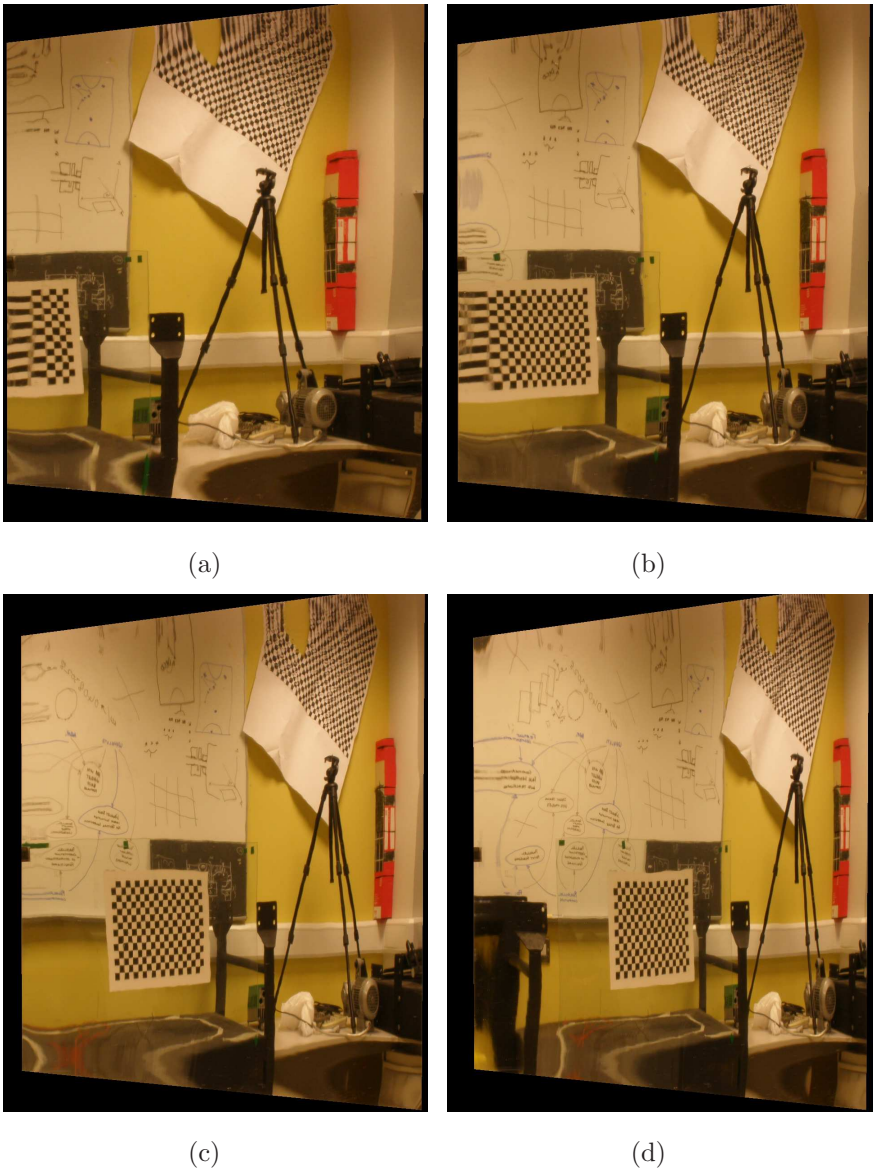


Fig. 5.9: B-spline blending functions for the full mirror surface model used in the real experiments.



(a) (b)

(c) (d)

Fig. 5.10: Scene images from the flexible mirror imager at different stages of deflection: (a) stage 0; (b) stage 1; (c) stage 2; (d) stage 3.

standard deviation of the ray-point distances for 1056 estimated mirror points across the image were $0.86mm$ and $0.40mm$ respectively, indicating a good initial mirror position estimate. The two active grids used in the general calibration step were found to have an estimated relative translation of $99.73mm$ and estimated relative rotation of 0.14^o, which compares well with the $100mm$ pure translation that the translation stage underwent. Differences between the translation stage transformation and the active grid monitor transformation are due to vibration of the monitor during travel, and consequently the translation stage travel cannot be considered a ground truth measurement for the monitor transformation, although similar values would be expected.

Table 5.1: Actual and estimated mirror deflections for each deflection stage.

Stage	Deflection magnitude (mm)		
	Actual	Linear estimate	Final estimate
1	5.0	2.69	4.71
2	5.0	3.07	4.61
3	5.0	3.10	4.74

The accuracy of the SPFC^2 method's primary calibration and the effect of the initial surface estimation using B-splines were evaluated by comparing directly measured scene rays with scene rays determined from the calibrated model. Approximately 160 pixels with corresponding $\mathbf{V}_l$s were chosen uniformly from the image, and the errors in the angles between the measured $\mathbf{V}_r$s and the $\mathbf{V}_r$s calculated by reflecting these $\mathbf{V}_l$s in the estimated mirror surface were determined. The mean and standard deviation of these errors were 0.19^o and 0.13^o, respectively.

Multiple Deflections

After primary calibration, the mirror was deflected in three $5mm$ stages, such that its shape became increasingly convex, and dynamic calibration was applied at each stage. Estimated mirror deflections determined from the SPFC^2 calibration data for each stage of deflection are presented in Table 5.1. All the final estimated deflection values show good agreement with the actual deflections, with a maximum error of 7.76% in the estimated deflections. The deflection point is closer to the mirror edge than any of the scene points used for surface estimation, therefore its displacement magnitude is significantly influenced by the surface interpolation. For stages 1, 2 and 3 the bias in the linear estimates causing underestimation of deflection is clearly evident. This underestimation is due principally to the small scene depths of less than $300mm$ and it is discussed in Section 5.2.2.

Additional assessment of the accuracy of each dynamic calibration over the entire mirror surface was made by again comparing directly measured $\mathbf{V}_r$s, determined using active grids as described in Section 5.2.1, with $\mathbf{V}_r$s estimated from the calibrated model. The estimated relative rotations and translations between the active grid in the two positions for each deflection stage are pre-

Table 5.2: Estimated relative translation and relative rotation between active grids for each deflection stage.

Stage	Estimated translation (mm)	Estimated rotation (^{o})
1	100.12	0.04
2	100.36	0.17
3	101.29	0.33

Table 5.3: Maximum measured angular displacements of $\mathbf{V}_r$s, and angular differences between measured and calibrated $\mathbf{V}_r$s, for each deflection stage.

Stage	Approx. max. angular displacement (^{o})	Angular error (^{o})	
		Mean	SD
0	0.0	0.19	0.13
1	4.1	0.42	0.16
2	8.3	0.24	0.12
3	12.8	0.23	0.15

sented in Table 5.2. As with the primary stage, they show good agreement with the $100mm$ pure translation of the translation stage. The mean and standard deviation of the angular differences between measured and estimated $\mathbf{V}_r$s for 168 uniformly selected image points, and for each stage of deflection, are presented in Table 5.3. After 3 deflections and applications of dynamic calibration the mean angular difference is only marginally greater than the angular difference for the surface estimate from the primary calibration stage. The distributions of the magnitudes of the angular differences across the images are shown in Fig. 5.11. As expected, the minimum angular differences occur at the fixed edge of the mirror where the positional constraint is enforced. The largest differences exist spuriously at the remaining edges of the mirror, and can be explained by the severe mirror distortions at these edges that are discussed earlier in this section.

The back-projection errors at each deflection stage for between 156 and 182 corner points on a randomly located $252 \times 261mm$ chessboard grid are presented in Table 5.4. By back-projection error is meant the on-plane error between the actual scene point and the intersection with the scene plane of the scene ray associated with the image of that scene point. Back-projection error vector residuals for the mirror after stage 3 deflection before and after

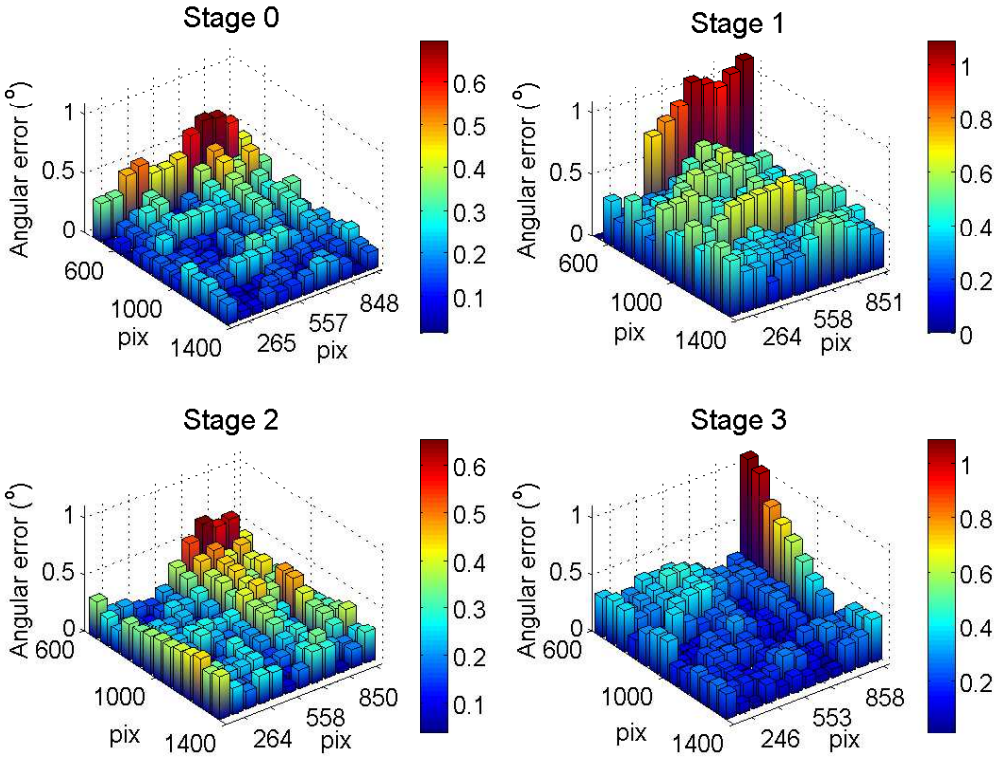
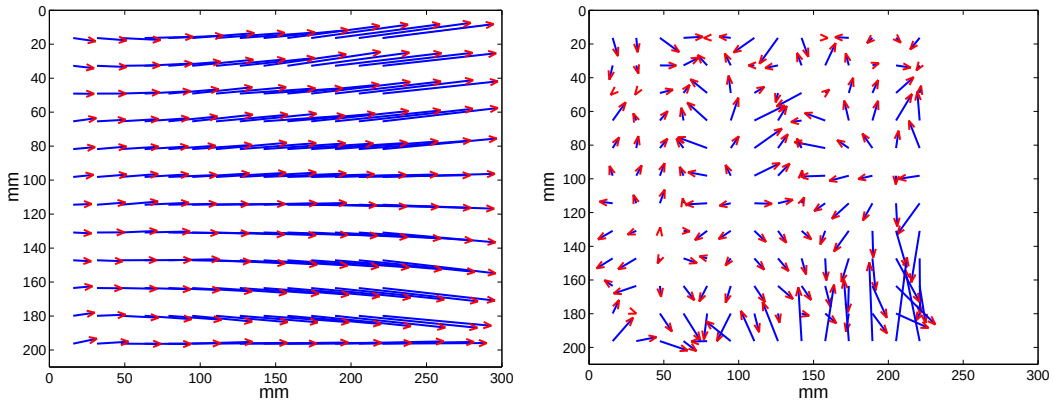


Fig. 5.11: Angular differences between measured and calibrated $\mathbf{V}_r$s across the image at each stage of deflection. The fixed mirror edge is the bottom right edge.

SPFC^2 calibration are shown in Figs. 5.12(a) and 5.12(b), respectively. Most of the non-linear distortion present before calibration for the $5mm$ deflection is seen to be removed after dynamic calibration. As with the angular differences, the largest residual errors occur along an unconstrained mirror edge, and they correlate with the severe distortions visible at the bottom edge of the images in Fig. 5.10. Back-projection error is a useful quantitative evaluation of calibration accuracy since it does not require the determination of reference scene rays, but the back-projection error magnitude is still dependent on scene depth, which is approximately $400mm$ for the back-projection experiments. Additionally, since the calibration grid used for determining back-projection error is randomly located, comparison of back-projection errors for different deflection stages is not a precise method of evaluating the relative calibrations. The effect is greatest if high curvature local mirror shape changes exist, since at such locations a small displacement in the reflection point of a scene ray can correspond to a large rotation of the scene ray direction. Back-projection errors are on-plane errors, whereas the cost function, Γ, minimised in the dy-

Table 5.4: Back-projection errors for a randomly located calibration grid, and residual errors after non-linear surface estimation.

Stage	Back-projection error (mm)		Residual error (mm)	
	Mean	SD	Mean	SD
0	0.56	0.33	–	–
1	0.52	0.32	0.68	0.58
2	0.52	0.30	0.67	0.53
3	0.42	0.28	0.80	0.51



(a) (b)

Fig. 5.12: (a) Back-projection error vector residuals before calibration for the stage 3 deflection (vectors are scaled ×5); (b) back-projection error vector residuals after calibration for the stage 3 deflection (vectors are scaled ×30).

namic calibration stage measures the ray-point Euclidean distances between corresponding scene rays and scene points. For each stage, these latter residuals after non-linear surface estimation, for between 132 and 154 grid corner points, are presented in Table 5.4. Interestingly, both the back-projection error statistics and the surface estimation residual statistics maintain approximately similar values across all applications of dynamic calibration.

Repeatability of Multiple Deflections

An experiment to test both the repeatability of the SPFC^2 method and its tolerance to large deflections was conducted next. The mirror was deflected in $5mm$ stages from stage 0 to stage 3 as before, increasing the mirror convexity, followed by deflections from stage 3 back through stages 2 and 1 to

stage 0, which will be termed concave deflections (since they reduce the mirror convexity). For the concave deflections, the grid used in dynamic calibration was kept in a fixed position for all deflections, allowing an additional dynamic calibration to be applied to calibrate directly from stage 3 to stage 0.

The estimated deflections and associated mirror surface estimation residuals for the stage by stage calibrations are presented in Table 5.5. Estimated deflections after calibration agree more closely with actual deflections than for the first experiment conducted in this section (see Table 5.1), despite the similar deflections and mirror initial positions for each experiment. The variation can be partly explained by variation in the calibration grid position used to provide scene points in dynamic calibration, which in turn affects the number of scene points used in the calibration and thus the spline surface interpolation. By using a calibration grid with a more dense set of scene points, this variation could be reduced. Final estimated deflection magnitudes for equivalent stages in the convex and concave deflections are within 4.74% of each other for all stages, indicating good repeatability in the dynamic calibration. The hysteresis loop error for the complete experiment is $0.11mm$, or 0.76% of the $15mm$ deflection. Residual error after surface estimation is similar across all deflection stages, with the exception of the last stage of concave deflection. Back-projection errors for the mirror in the stage 0 position, after primary calibration only has been applied, have mean $0.85mm$ and SD $0.51mm$. The back-projection error mean and SD increase to $1.84mm$ and $1.47mm$, respectively, for the stage 3 position, but after the three subsequent concave deflections back to stage 0 the error slightly reduces to have mean $1.84mm$ and SD $1.37mm$.

Next the mirror was dynamically calibrated for the stage 0 position after a $15mm$ deflection from stage 3, resulting in a linearly estimated deflection of $9.66mm$ and a final deflection estimate of $15.15mm$. The total concave mirror deflection estimated in 3 stages above is $15.13mm$, so the deflection magnitude estimated in a single stage is within 0.14% of that estimated in multiple smaller stages. Such close agreement in estimates indicates that there is a strong global minimum of the cost function, Γ, at the final surface solution. The residual after non-linear surface minimisation has mean $2.29mm$ and SD $3.30mm$, which is comparable to the residual error at the same deflection stage when calibration is performed in stages, whilst the back-projection error for a randomly located plane has mean $2.74mm$ and SD $1.53mm$ for 168 points.

Table 5.5: Actual and estimated mirror deflections, and surface estimation residuals, after each deflection stage for 6 consecutive deflections (3 convex and 3 concave).

Stage	Deflection magnitude (mm)			Residual error (mm)	
	Actual	Linear estimate	Final estimate	Mean	SD
1	5.0	2.90	5.14	0.80	0.50
2	5.0	3.25	5.04	1.68	0.90
3	5.0	3.21	5.06	1.25	0.63
2	5.0	3.23	5.29	1.03	0.61
1	5.0	3.17	4.93	0.98	0.58
0	5.0	2.96	4.91	2.20	3.24

Due to the multiple surface shape variations across the mirror, the back-projection error vector residuals reveal more about the relative quality of the calibrations than the back-projection error statistics. Vector residuals for the back-projection errors for stages 0 and 3 are presented in Fig. 5.13. The accumulation of error after multiple deflections is evident when comparing the back-projection error residuals before deflection, shown in Fig. 5.13(a), with the residuals after six applications of dynamic calibration, shown in Fig. 5.13(c). However, the multistage concave deflections do not appear to introduce significant error (compare Fig. 5.13(b) with Fig. 5.13(c)). For concave deflections, the scene points that are visible in the image both before and after deflection are distributed across a larger portion of the mirror than for convex deflections. A more even set of $\{\mathbf{V}_l, \mathbf{V}_r\}$ improves the conditioning of the data matrix in linear estimation and results in more accurate surface estimation. The largest back-projection error tends to occur around the grid periphery, or equivalently along the mirror edges, and it is seen that the error is consistently less at the grid edge that is viewed by scene rays incident on the mirror close to the constrained edge (left edge), as expected. Figs. 5.13(c) and 5.13(d) show the back-projection error residuals for stage 0 calibration after three applications of dynamic calibration, and after a single application of dynamic calibration, respectively. The dissimilarity between these vector plots indicates that although the surface error is small at the point of mirror deflection, error is significantly larger at other mirror surface points when calibration is performed from a single large deflection. Residual distortions visible in Fig. 5.13(a) are due to high order surface shape components that

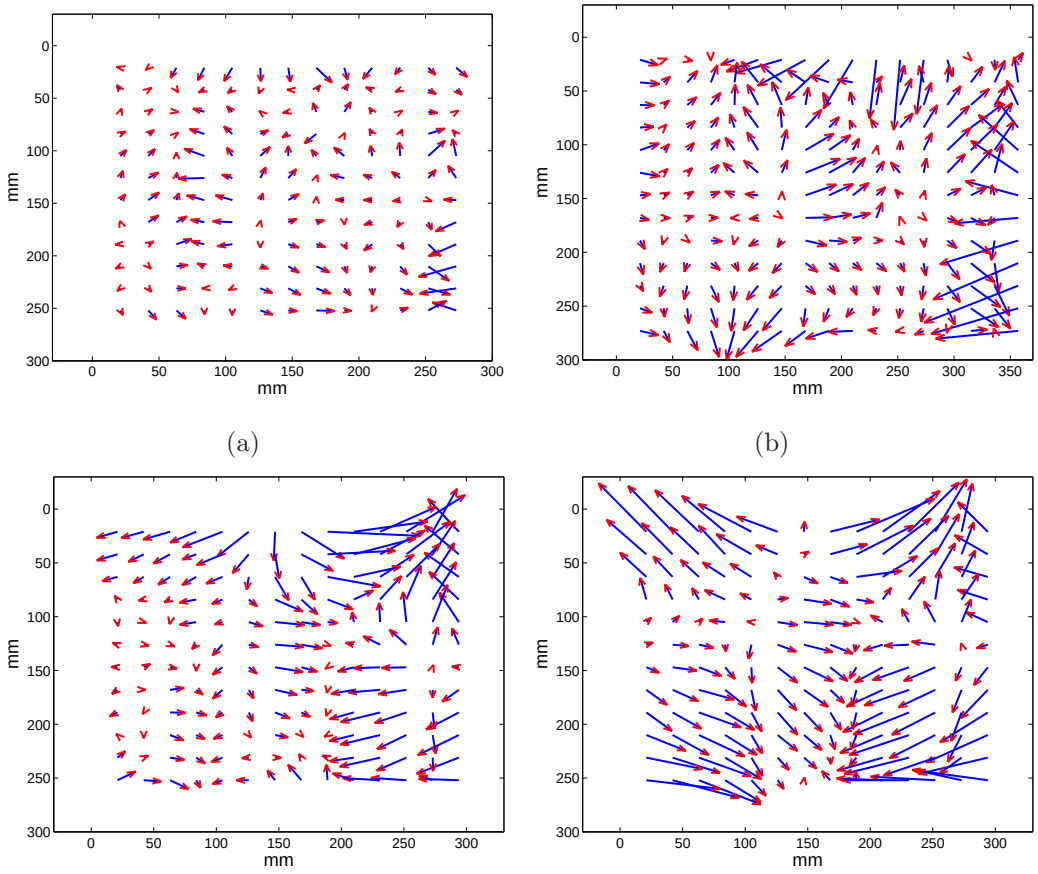


Fig. 5.13: Back-projection error vector residuals for (a) stage 0 before deflections; (b) stage 3 after three convex deflections; (c) stage 0 after three convex and three concave deflections; (d) stage 0 after three convex deflections and a single concave deflection. All vectors are scaled $\times 10$.

are not captured by the camera model configured with the chosen set of basis functions. The repeated application of dynamic calibration can amplify any distortions present since the plane pose of the scene points used for dynamic calibration is determined from the current calibration estimate. However, the amplification effect is reduced by the limited self-correction that occurs in the pose estimation stage of dynamic calibration, where the known metrics of the calibration grid are imposed on the estimated scene points.

Three conclusions can be drawn from the preceding experimental results. Firstly, the convergence of the cost function applied in dynamic calibration is not very sensitive to the accuracy of the linear surface estimate. This is despite the known bias in the linear estimate and the close proximity of the scene points to the mirror. Secondly, in contrast to multiple smaller deflec-

tions, back-projection error is shown to increase in size for dynamic calibration with a larger deflection. The increase in error can be attributed partly to the fewer imaged scene points that are common to the images before and after a larger deflection. Thirdly, although dynamic calibration introduces error into the calibration results, the results do not reveal the presence of any significant cumulative error due to multiple applications of dynamic calibration. Instead a step error is seen in the results, which is possibly due to the varied positions of the scene point features in the image. Considering the initial multiple deflection experiment, the error in deflection magnitudes, the angular differences of scene rays, the magnitudes of the back-projection errors, and the magnitudes of the residuals after final non-linear surface estimation, do not show any significant trend across the different stages of deflection. For the experiment examining the repeatability of multiple deflections, the low cumulative error is evidenced by the increase in back-projection error of only 12.76% between the first and last applications of dynamic calibration, and by the small hysteresis loop error.

The angular differences presented in the above results are contingent on measured reference angles whose accuracy is difficult to quantify, and these reference angles may themselves contain error. The measured back-projection errors depend on the proximity of the calibration grid location with respect to the mirror. Consequently, alternative methods of quantitatively evaluating the calibrations that enable a more consistent and more relevant measure of the accuracy of the SPFC^2 method across experiments are considered in the next sections. These experiments evaluate the overall performance of the calibration through the application of the calibrated flexible mirror to real vision tasks, specifically distortion correction and motion and structure estimation.

5.3.4 Distortion Correction

Exact perspective correction of an image of a scene taken by a multi-perspective camera, such as a flexible mirror imager, requires an a-priori model of scene geometry. When knowledge of scene geometry is unavailable, methods for generating nearly perspectively correct images can be applied (Kuthirummal and Nayar, 2007, Swaminathan, Grossberg and Nayar, 2003), but these are not straightforward. Swaminathan, Grossberg and Nayar (2003) introduced the concept of shape priors that approximate a scene using a small set of ba-

sic geometric entities. By intersecting scene rays with the estimated shape priors, near-perspective images can be formed from multi-perspective cameras. Clearly, exact perspective correction can be performed easily for a planar scene when the scene plane's pose is available, and this is the approach used here to evaluate the calibration quality. A planar evaluation target consisting of a chessboard calibration grid with a superimposed circle, as shown in Fig. 5.14(a), was printed using a laser printer and then was mounted on a glass substrate. After capturing an image of the evaluation target with the flexible mirror imager at each stage of mirror deflection, the target's pose was estimated using a non-central pose estimation technique applied to the chessboard corners visible in the image. Then distortion correction of the circle on the calibration target was performed by intersecting the calibrated scene rays with a plane in the estimated pose.

Perspective distortion alone applied to a circle results in a proper conic. Any non-linear distortion present causes the fit of this conic to deteriorate. Subfigures (a) of Figs. 5.14-5.17 show flexible mirror camera images of the calibration target at each deflection stage, overlaid with a best-fit conic. The conics are fitted to the subpixelly detected circle edge points using linear least-squares to minimise the algebraic error – optimal conic fitting is not performed as the fits are only indicative of the extent of the non-linear distortion. Mismatches between the overlaid conics and the calibration target circles highlight the extent of the non-linear distortion, which clearly is seen to increase with increasing deflection. The error residuals, and best-fit circles, after fitting circles to the distortion corrected circle edge points, are shown in subfigures (b) of Figs. 5.14-5.17, from which it is seen that there is no systematic error in the residuals across the four stages of deflection. Optimal circles were fitted to the corrected data subpixel edge points by non-linearly minimising geometric error. The estimated circle radii for the distortion corrected images, presented in Table 5.6, agree closely with the measured circle radius of $87.21mm$ for all deflection stages. Additionally, the fit residuals shown in Table 5.6 do not display significant error accumulation resulting from repeated application of dynamic calibration. The radius error and fit residual statistics for stage 3 are smaller than those for both stages 1 and 2.

These distortion correction tests actually evaluate the calibration quality twice – once when determining the pose of the grid plane, and again when determining the fit of the circle to the distortion corrected circle on this plane.

Table 5.6: Error residuals and estimated circle radii after circle fitting to the distortion corrected circle edge points.

Stage	Residual error (mm)		Estimated Radius	Radius error
	Mean	SD	(mm)	(%)
0	0.15	0.11	86.96	0.28
1	0.21	0.16	86.88	0.37
2	0.29	0.21	86.20	1.15
3	0.18	0.13	87.04	0.19

Alternative testing could be undertaken that evaluates the calibration quality only once, by selecting a random plane on which to generate the intersection points, and then by directly fitting a conic to these points. However, since there is only one integrated parameter set changing between tests – the calibration data itself – the double evaluation is valid.

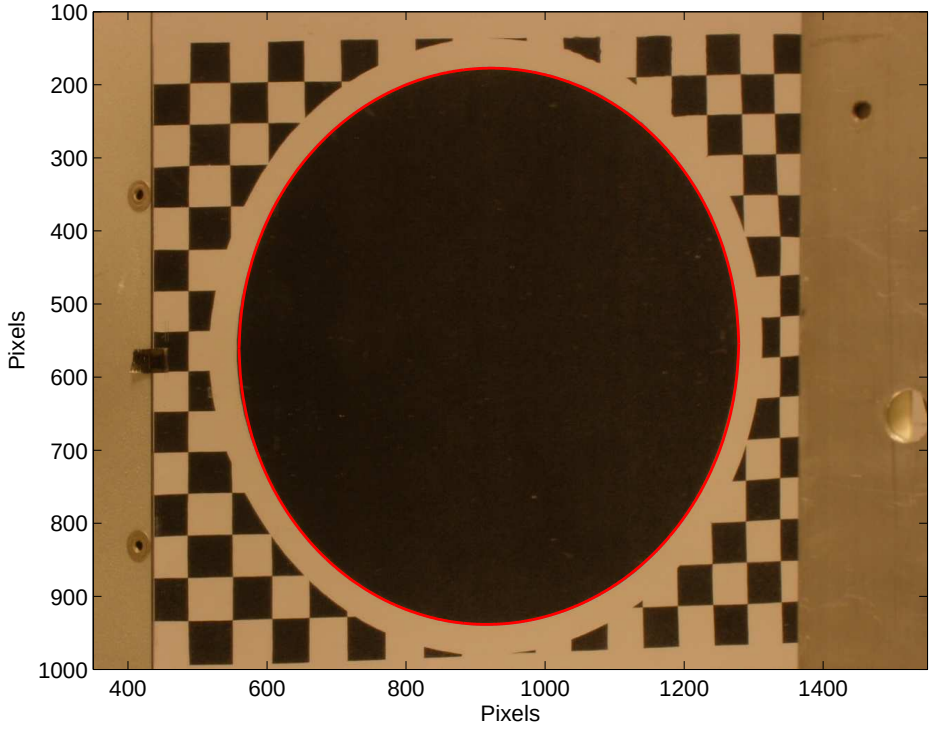
The distortion corrected image of the calibration target for the stage 3 deflection, which contains the most non-linear distortion, is shown in Fig. 5.18. The circle appears qualitatively correct, although some residual distortion is evident in the grid squares along the bottom of the image.

If scenes with low depth relief are loosely modelled as planes, approximate perspectively corrected images of such scenes can be generated, although due to non-planarity they will still contain caustic distortion (Swaminathan, Grossberg and Nayar, 2003). Fig. 5.19 shows a face image from the flexible mirror imager both before and after such approximate perspective correction. The majority of the non-linear distortion of the face is seen to have been removed in the corrected image.

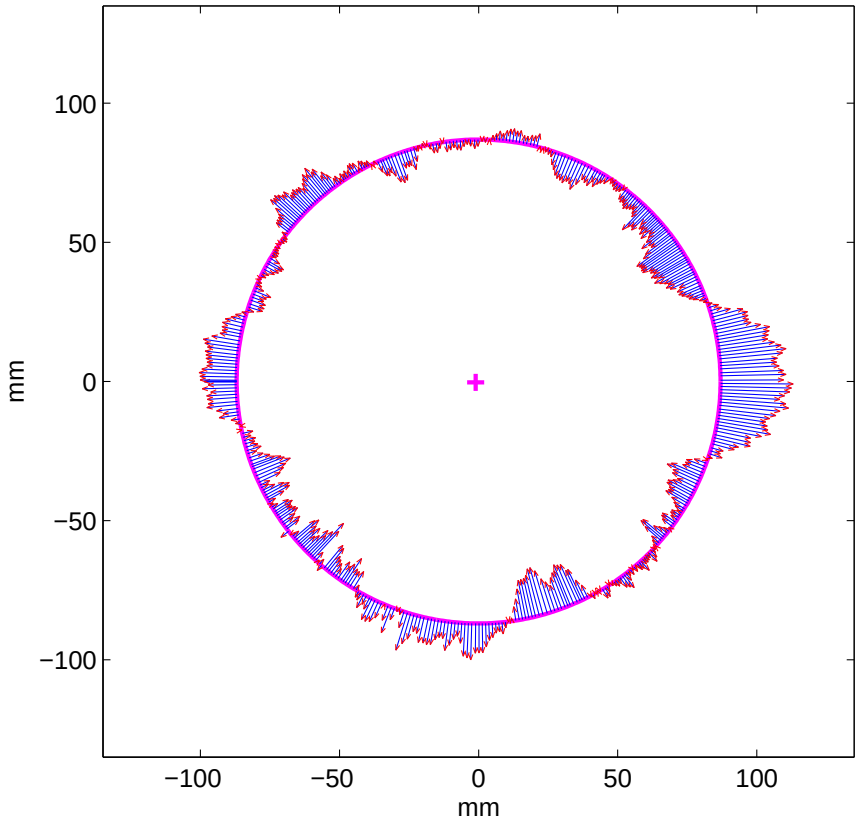
5.3.5 Motion and Object Reconstruction

A structure-from-motion experiment was conducted with the flexible mirror imager calibrated for the stage 3 deflection using the SPFC^2 method. The flexible mirror imager in general configuration is non-central, so non-central structure-from-motion must be applied. This is based on the 6×6 general essential matrix, $\mathcal{E}$, that operates on Plücker coordinate vectors L_1 and L_2 as

$$L_2^T \mathcal{E} L_1 = 0 \tag{5.30}$$

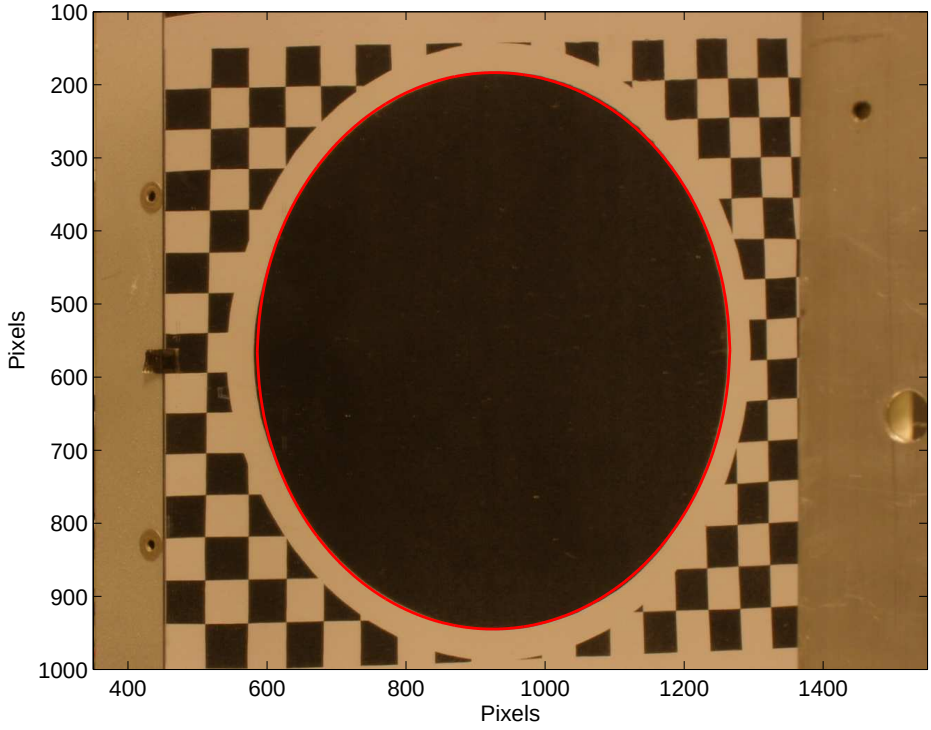


(a)

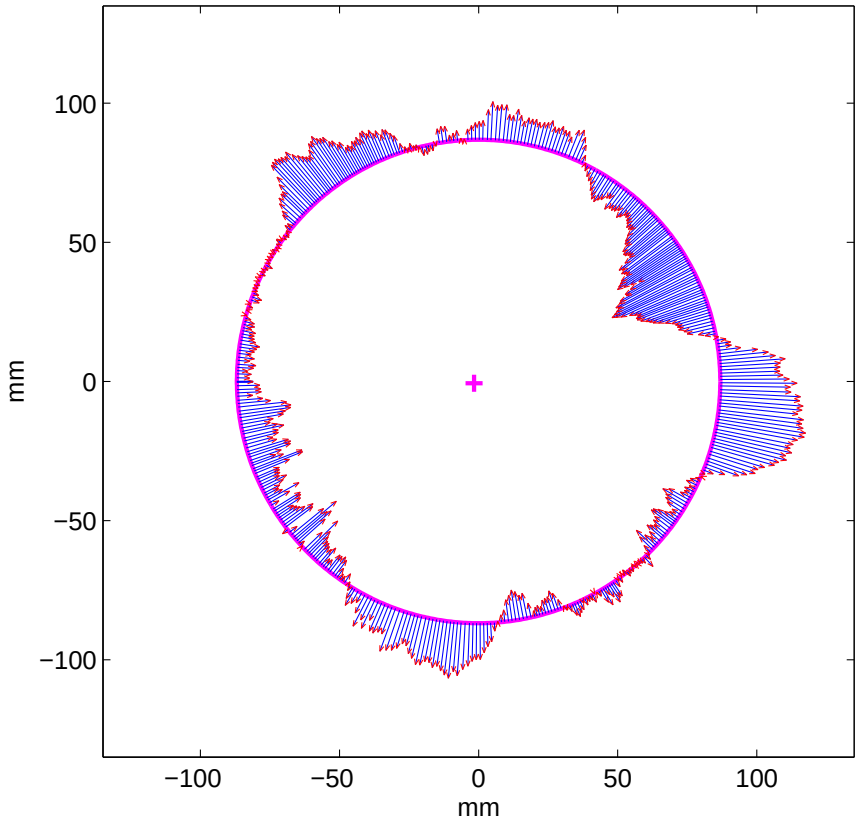


(b)

Fig. 5.14: Perspective correction for stage 0 deflection. (a) Original image with best-fit conic; (b) vector residual errors for best-fit circle applied to the undistorted circle edge points. Vectors are scaled ×50.

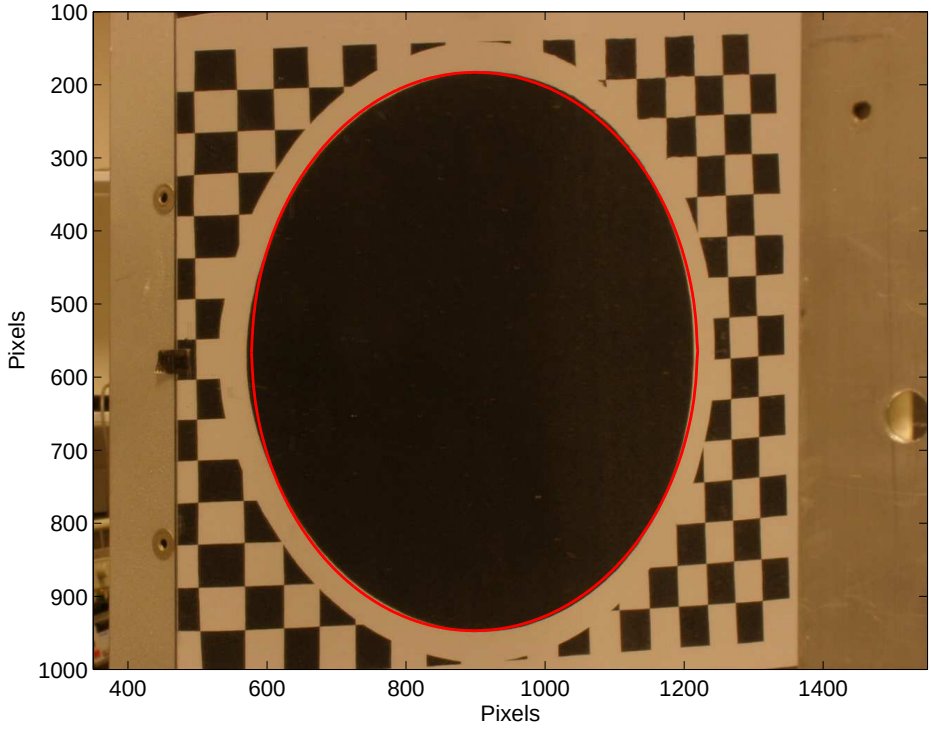


(a)

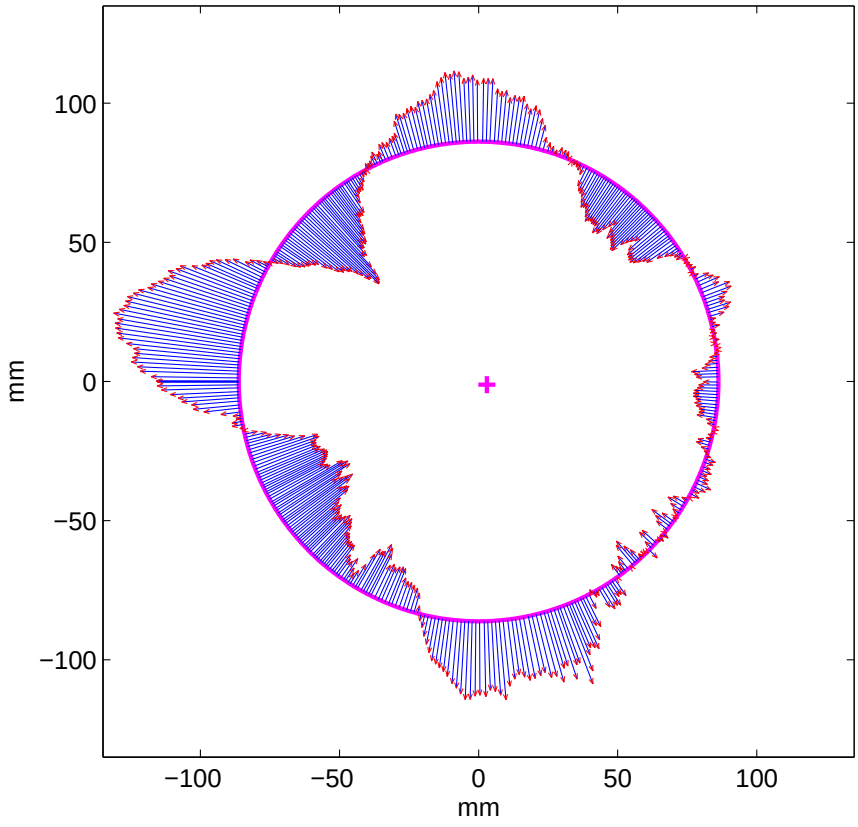


(b)

Fig. 5.15: Perspective correction for stage 1 deflection. (a) Original image with best-fit conic; (b) vector residual errors for best-fit circle applied to the undistorted circle edge points. Vectors are scaled ×50.

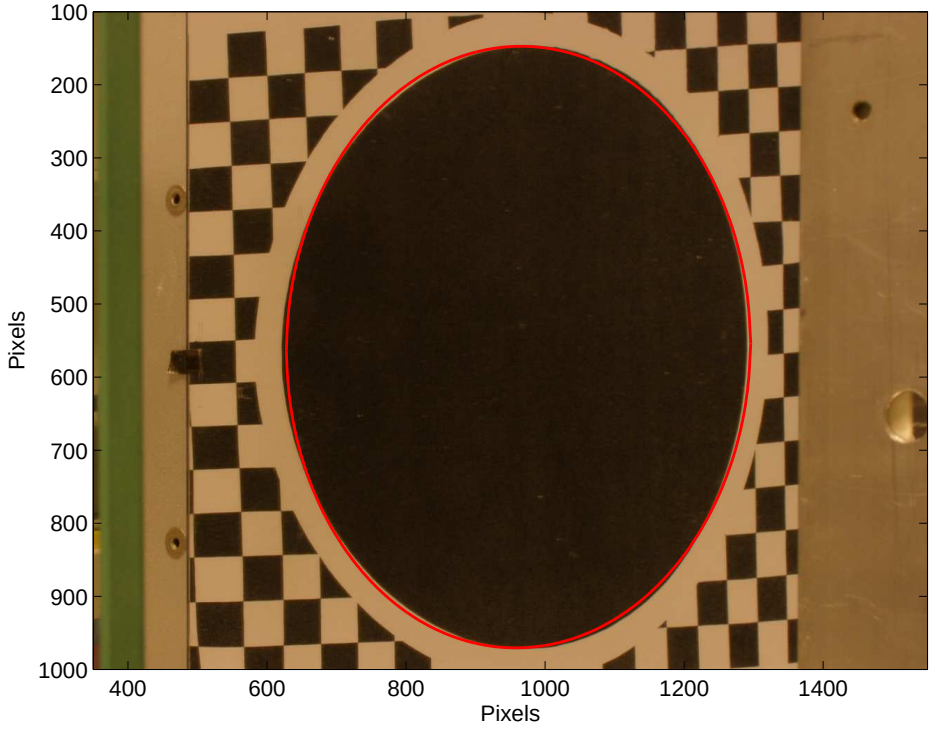


(a)

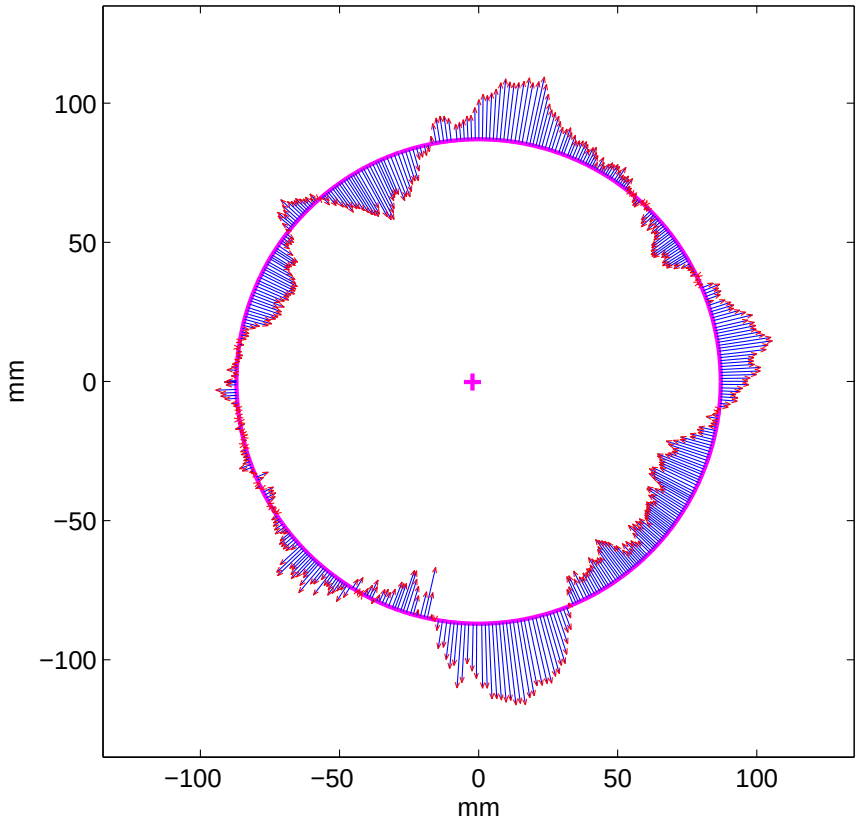


(b)

Fig. 5.16: Perspective correction for stage 2 deflection. (a) Original image with best-fit conic; (b) vector residual errors for best-fit circle applied to the undistorted circle edge points. Vectors are scaled ×50.



(a)



(b)

Fig. 5.17: Perspective correction for stage 3 deflection. (a) Original image with best-fit conic; (b) vector residual errors for best-fit circle applied to the undistorted circle edge points. Vectors are scaled ×50.

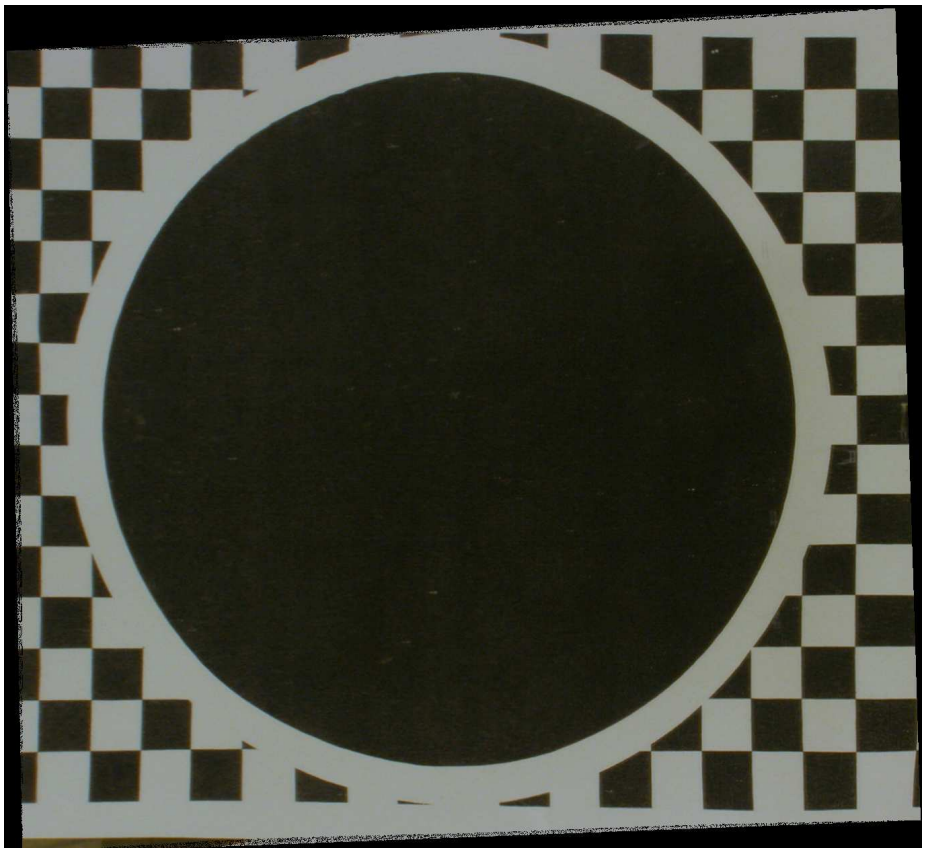
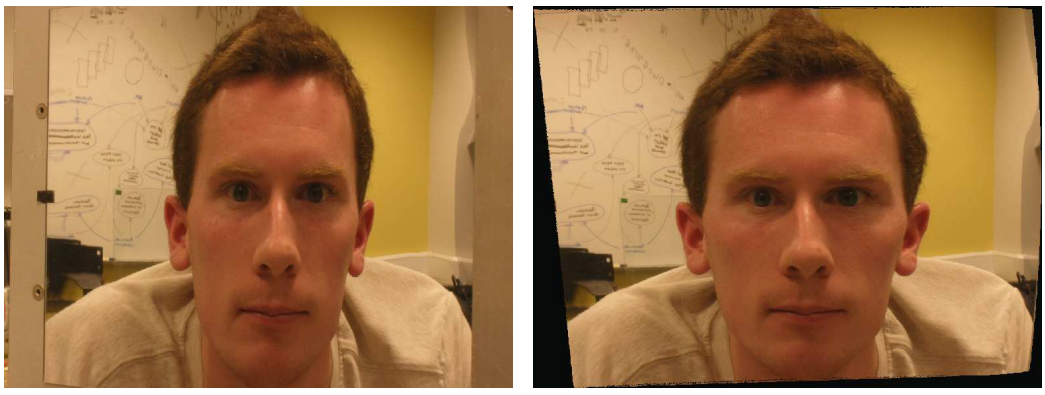


Fig. 5.18: Stage 3 deflection circle image after distortion correction.



(a) (b)

Fig. 5.19: (a) Face image from the flexible mirror imager after a $12mm$ deflection; (b) Near-perspective face image with non-linear distortion removed using the SPFC^2 calibration data. Note the large distortion at the top right of the image that was not captured by the calibration.

For each image correspondence point, the associated viewing ray is intersected with the calibrated mirror surface to establish a scene ray and a point on that ray. The Plücker vector is then determined as described in Section 2.3.2. Ramalingam et al. (2006) detail how the calculation of $\mathcal{E}$ differs from that of the standard essential matrix, and their method with bundle adjustment is applied here to generate the experimental results. The results for pairwise motion estimation of a cylindrical objected translated in $25mm$ steps from $0mm$ to $125mm$ are presented in Table 5.7, where the mean estimated translation is scaled to $25mm$. Fig. 5.20 shows the actual translation and estimated reconstructed motion for the cylinder. By pairwise is meant that essential matrices are calculated only between consecutive pairs of images, and not relative to a datum image. The estimated translations and rotations shown in Table 5.7 are accumulated for each stage, as would be done in a real application of structure-from-motion. Estimation error after all 5 translations is $4.05mm$. All the cumulative motion estimates show good agreement with the corresponding ground truth values, further validating the SPFC^2 method. Since the deflected flexible mirror imager is non-central, the calibration information actually contains metric scale information by way of the skew scene rays, although the closer the mirror is to planar, the more central it becomes and the less reliable is the scale information. In the experiment the mean estimated translation for the pairwise estimates before scaling is $31.34mm$ rather than the actual $25mm$. This estimation error, which is due to error in the calibration, is extremely sensitive to calibration accuracy for cameras that are close to central, as in the configuration for this experiment.

The best fit cylinder to the points reconstructed in the experiment, and the reconstructed points themselves, are shown in Fig. 5.21. Fitting is performed non-linearly by minimising geometric error. The reconstructed points, which are represented by spheres with radius $3mm$ in order to aid visualisation, are seen to intersect the estimated cylinder in all cases except one. The best-fit cylinder radius for the pairwise experiment before scaling is $73.90mm$, and after scaling to match the mean translation to $25mm$ is $58.96mm$, which both compare well with the measured cylinder radius of $62.98mm$. The mean and standard deviation of the cylinder fit are $0.94mm$ and $0.81mm$ respectively. These experimental results indicate that the accuracy of the flexible mirror calibration is sufficient for practical structure estimation in a structure-from-motion framework, especially considering that only 25 cylinder points are in-

Table 5.7: Ground truth translations and estimated translations and rotations for the structure-from-motion experiment. The mean estimated translation is scaled to the ground truth value.

Ground truth	Pairwise estimates	
translation (mm)	translation (mm)	rotation (^{o})
25	25.05	0.38
50	50.53	0.72
75	72.94	1.56
100	102.96	0.46
125	123.14	1.66

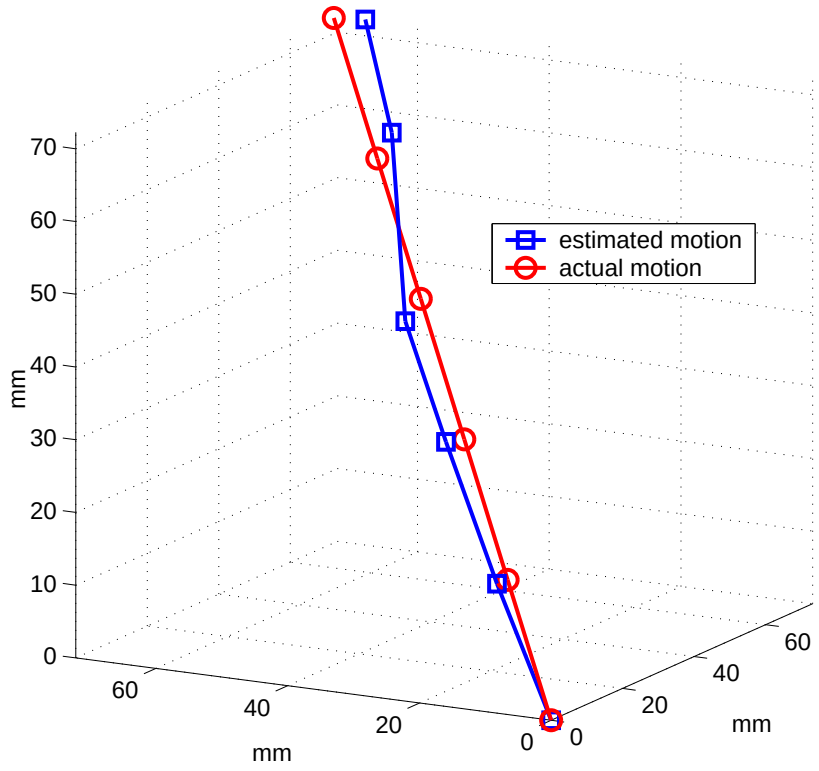


Fig. 5.20: Actual and estimated translations for pairwise motion estimation.

volved in the experiment.

Object reconstruction for the SPFC^2 calibrated flexible mirror imager after the stage 3 deflection was performed by imaging an approximately cuboidal 'coffee tin' before and after a $100mm$ translation of the imager. Structure is estimated by triangulation, whereby each reconstructed point is determined as the point closest on average to all the scene rays associated with pixels that see that point. Two-view reconstruction is applied here, so there are two scene rays

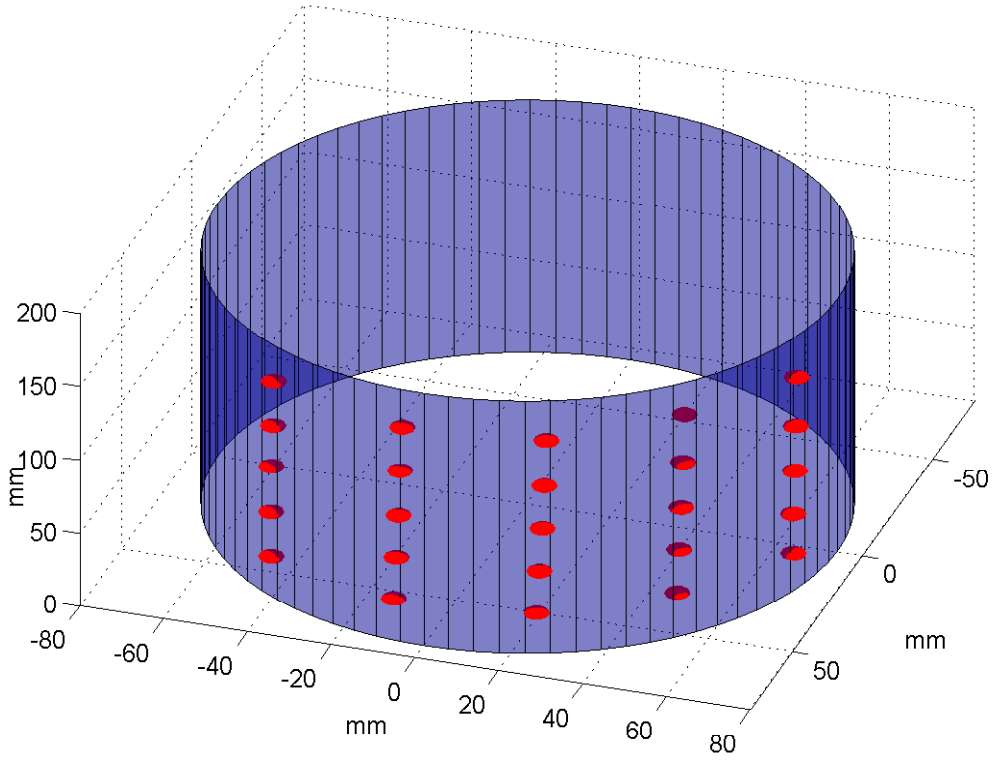


Fig. 5.21: Estimated structure for pairwise motion estimation.

per scene point, and triangulation is performed linearly for each reconstructed point using the method presented by Ramalingam et al. (2006). Image feature points on the object were automatically extracted using a SIFT feature detector executable[3], and feature matches were established semi-automatically by thresholding on angular differences between feature vectors and subsequent manual outlier removal. This resulted in 460 feature matches, which were triangulated to generate a sparse reconstruction of three mutually adjacent sides of the tin. Plane fitting was performed for each reconstructed side to form surfaces, and texture mapping was accomplished by intersecting calibrated scene rays with these planes and colouring the inter-point patches accordingly. The final texture mapped 'coffee tin' model is shown in Fig. 5.22. Texture on each side of the reconstruction appears geometrically correct, giving a qualitative indication that the calibration is accurate. The mutual angles between the planes fitted to each side are presented in Table 5.8, alongside reference angles determined by fitting planes to 3D points obtained from laser scanner measurements of the tin. The agreement between the estimated and reference angles is good, with the maximum angular difference being 1.65^o. Histograms of the signed Euclidean distance from each reconstructed point to the best fit

[3]http://www.cs.ubc.ca/~lowe/keypoints/ (Accessed January 2006)

Table 5.8: Angles between planes fitted to each side of the 'coffee tin' reconstructions for both the estimated structure using the flexible mirror imager and the laser scanner reference structure measurements.

Planes	Angle (^{o})	
	Estimated	Reference
top—left	88.63	90.28
top—right	88.74	89.15
left—right	88.09	89.17

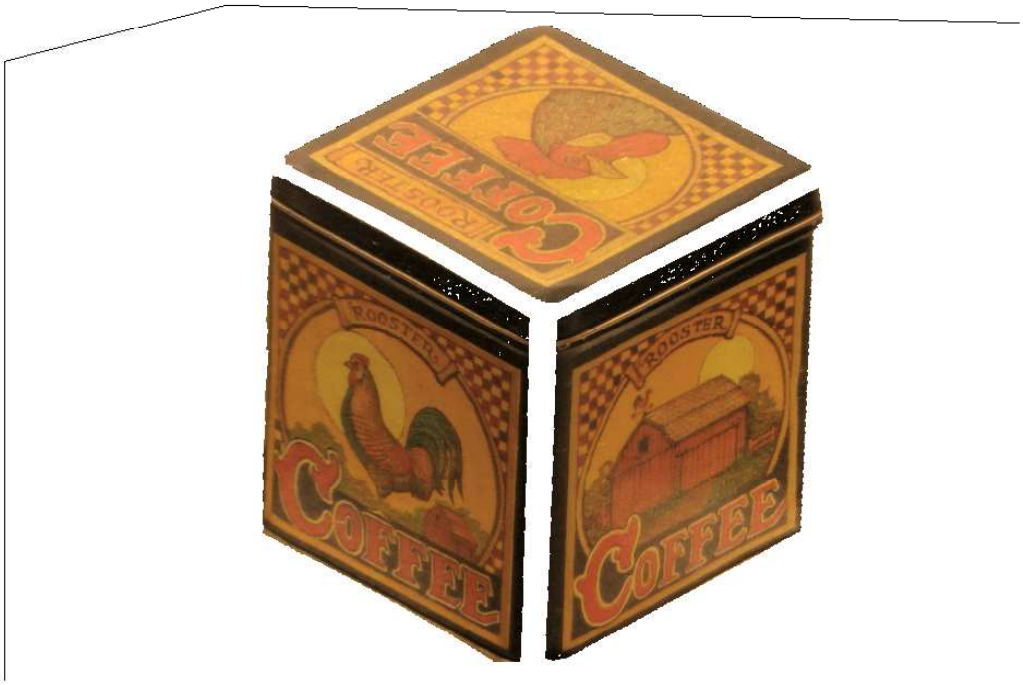


Fig. 5.22: Reconstructed and textured 'coffee tin'.

plane for each reconstructed side of the 'coffee tin' are presented in Fig. 5.23, with superimposed normal distribution plots. Some outlier histogram bins are present, but in general the distribution is normal.

5.4 Flexible Mirror Camera Self-Calibration

The SPFC^2 method requires that a calibration grid be imaged in the same position both before and after each mirror deflection. Although not time consuming, this necessitates a calibration object in the scene and human interven-

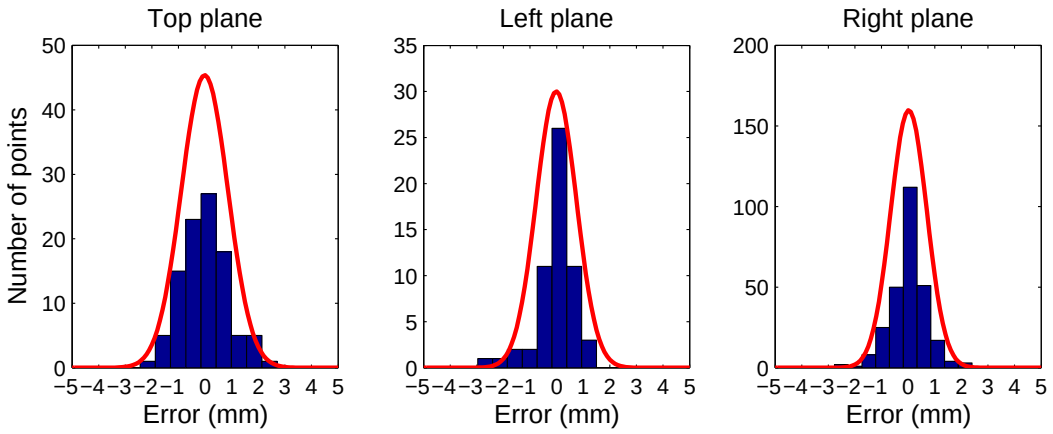


Fig. 5.23: Signed point-plane error histograms and superimposed normal distributions for each side plane of the reconstructed 'coffee tin'.

tion to place and remove it. A self-calibration method, in which a camera is calibrated from image correspondences across multiple images of an unknown scene, clearly would be advantageous for the calibration of flexible mirror imagers.

One step towards flexible mirror imager self-calibration would be to eliminate the need for the calibration grid during dynamic calibration. This calibration problem is less constrained than the original calibration problem since no information is immediately available on scene point depths. Consequently the error metric, Γ, is inapplicable and the proposed non-linear minimisation in dynamic calibration can not be performed. However, the linear surface estimation, based on the assumption of infinite scene depth, is still possible. As discussed in Section 5.2.2, the linearly estimated surface is biased, and the extent of the bias is inversely proportional to the depth of the scene. This bias is evident in the presented results that evaluate the SPFC^2 method. Those results show that for scene depths of approximately $400mm$ the linear surface estimates are between 30.32% and 42.90% in error relative to the final estimate, which indicates that self-calibration based on linear estimation alone is likely to contain significant error.

In order to verify this, dynamic calibration from an unknown distant scene, using the linear estimate alone, was undertaken. It was found that an uneven distribution of scene points across the image caused the fit to be biased. This bias was evident in the estimated surfaces as an erroneous curvature of the mirror surface in the vertical direction. The error was amplified at each appli-

Table 5.9: Actual and estimated mirror deflections for the self-calibration experiment, for forwards deflections from stage 0 to stage 3.

Stage	Deflection magnitude (mm)	
	Actual	Self-calibration estimate
1	5.0	3.98
2	5.0	3.48
3	5.0	3.08

cation of dynamic calibration, such that the stage 3 surface estimate contains over $20mm$ of incorrect vertical curvature at the deflected mirror edge. Despite this, due to a rich feature set across the middle of the mirror image, deflection estimates and curvature could be examined along a horizontal line on the mirror passing through the point of deflection. The estimated deflections after self-calibration from a distant scene are presented in Table 5.9. Estimated deflections for stages 1 and 2 are substantially better than the corresponding linear estimates given in Table 5.1 for the experiment using the SPFC^2 method. This improvement is due to the larger scene depths of the scene points, and agrees with the simulated results from Section 5.3.1. By stage 3 the self-calibration has degraded to contain the same error as the linear estimate of the full method, since cumulative linear estimation error is not corrected by a non-linear minimisation. Fig. 5.24 shows the estimated shape of the mirror along the horizontal line passing through the deflection point. The inflexion point evident in the deflection curve for the stage 2 and stage 3 deflections can be attributed to the depth of the scene points used in the calibration. As the mirror was deflected from stage 0 to stage 3, new scene points became visible in the image (at the deflected mirror edge) that were at a smaller scene depth than the other scene points in the calibration. The smaller depths cause increased error in the infinite scene assumption, and consequently introduce phantom shape changes in the mirror surface estimate.

If an even distribution of scene points exists, and the scene depth variation is small, then a potential route for correcting the linear estimation bias is to build up a scene model and feed this information back into the calibration process, by way of estimated scene points, to update the calibration. This approach has similarity with the structure-from-motion methods used for self-calibration of conventional and catadioptric cameras (Fitzgibbon, 2001, Micusik and Pajdla,

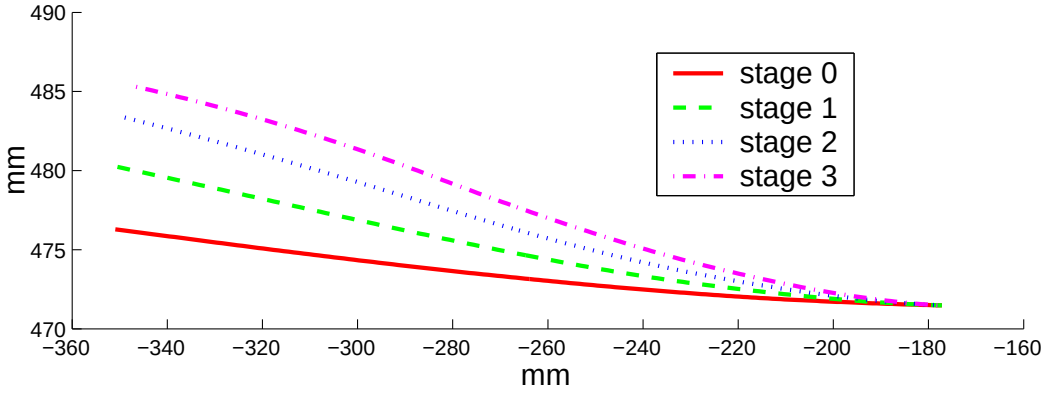


Fig. 5.24: 2D plot of the estimated deflections for a horizontal line on the mirror passing through the deflection point, for forwards deflections from stage 0 to stage 3.

2006, 2004), although in this case the camera 'motion' is the mirror deflection and a low-order projection function for the flexible mirror camera is not available. Consider two images from a flexible mirror camera, one before and the other after a mirror deflection. If the calibration at each deflection is known, then the two images constitute a stereo pair, and using stereo triangulation some reconstruction of the scene is possible. However, unlike standard stereo, the baseline for this stereo pair can change both in magnitude and direction across the images, depending on how the mirror is deflected. Let this configuration be called variable baseline stereo. At the fixed point or points of the mirror, corresponding rays are parallel and so no scene depth information is available. Moving away from any fixed points of the mirror, the reconstruction becomes more reliable as the baseline lengthens. However, the baseline is typically very small, in the order of several millimetres, and so the angles between corresponding skew scene rays in the stereo pair are very acute. As a result any variable baseline stereo reconstructions can be expected to have large error, which would impact severely on a feedback-based self-calibration scheme. In order to quantify the extent of the error, variable baseline stereo experiments for the calibrated flexible mirror imager were conducted. Stereo reconstructions from pairs of deflected mirror images of the chessboard grids used in dynamic calibration are shown in Fig. 5.25. Even for these grids, on which the calibrations themselves are based, the reconstruction error is substantial, indicating the high sensitivity of the variable baseline flexible mirror imager stereo pair to noise and calibration errors.

One further experiment was conducted in order to investigate the possibility

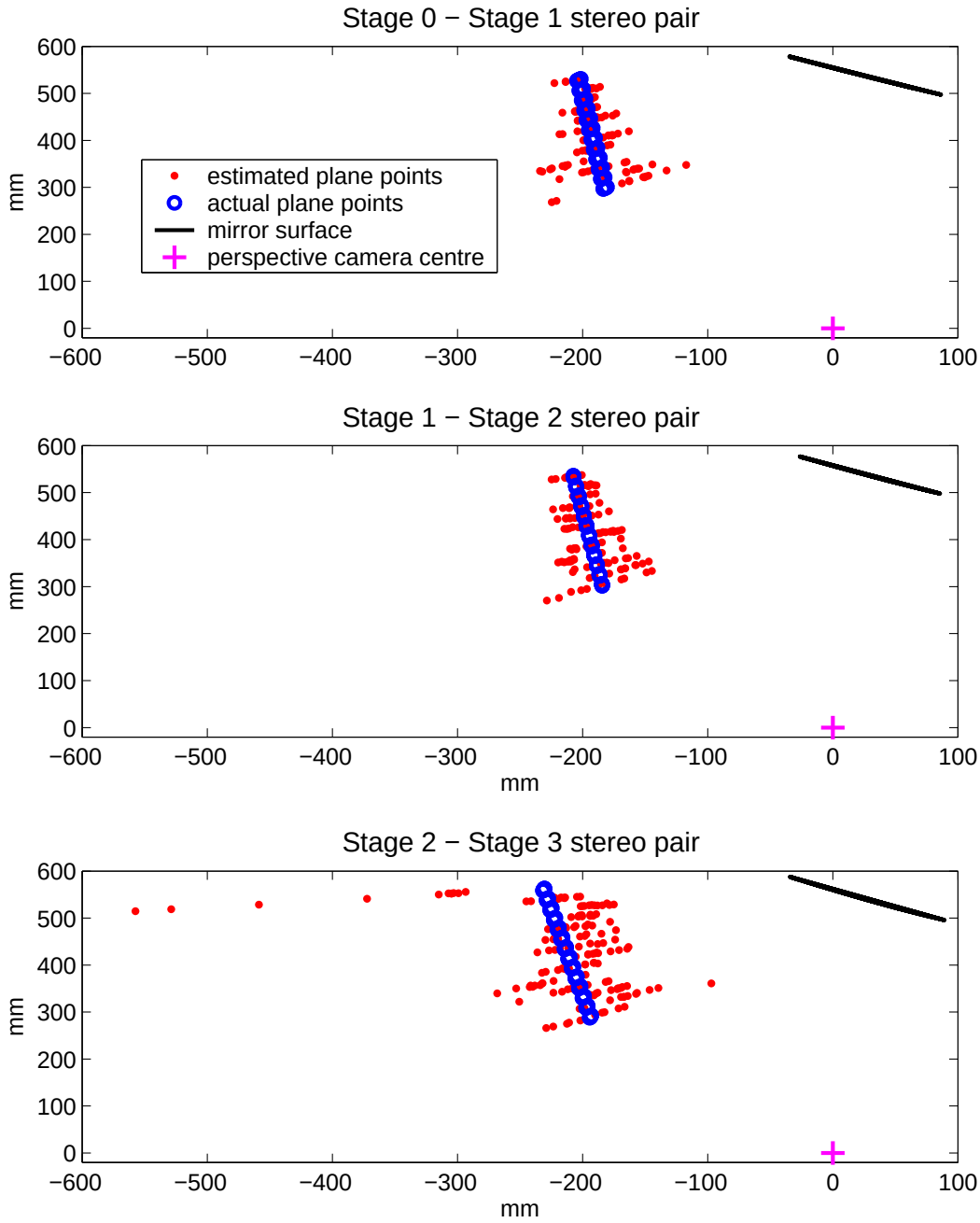


Fig. 5.25: Stereo reconstructions for calibrated flexible mirror image pairs. The fixed mirror edge is on the right of the mirror surface.

of reconstructing a scene from several flexible mirror camera images with deflections between them. Multiple view reconstruction by triangulation, as used in the structure-from-motion experiments, was applied to images from all four calibrated deflected mirror stages. The resulting estimated scene is shown in plan view in Fig. 5.26. Red and blue points lie on parallel vertical planes, and the green points are features on an object that is located between the two planes, as can be seen in Fig. 5.27. The very large error in the reconstructed points is clearly apparent in the $\sim 2.1m$ spread in scene depth of the points on the front plane, and the $\sim 3.85m$ spread in scene depth of points on the rear plane. Some of this error is attributable to error in feature point selection, which was performed manually and without sub-pixel refinement. The sparsity of the selected object features also impacts negatively on the results. If only the centroids of the reconstructed points, shown in Fig. 5.26, are considered, then the order of the objects in terms of their depths from the mirror is correct.

The last two experiments were conducted for calibrated configurations of the deflected flexible mirror imager. They indicate that any type of multi-view reconstruction from configurations that differ solely by mirror deflections is likely to contain very large error. It is concluded from the experimental results that self-calibration of flexible mirror cameras using the SPFC^2 method with the linear solution alone does not give acceptable results. Also, multi-view reconstruction from flexible mirror images differing solely by mirror deflections is unreliable due to the very short, variable, view baselines. Therefore, self-calibration by simultaneous scene reconstruction and mirror surface linear estimation is unlikely to be accurate enough for even a rudimentary calibration without at least some a-priori knowledge of scene geometry.

5.5 Discussion

This chapter details a proposed novel calibration method for flexible mirror imagers, termed the SPFC^2 method. The only existing method for specifically calibrating such systems operates offline by directly acquiring 3D mirror surface shape measurements for a very large set of possible mirror deformations. The SPFC^2 method does not require a significant offline stage, since the current calibration is dynamically updated after each mirror deflection. The main idea of the SPFC^2 method is the use of scene features rather than mirror

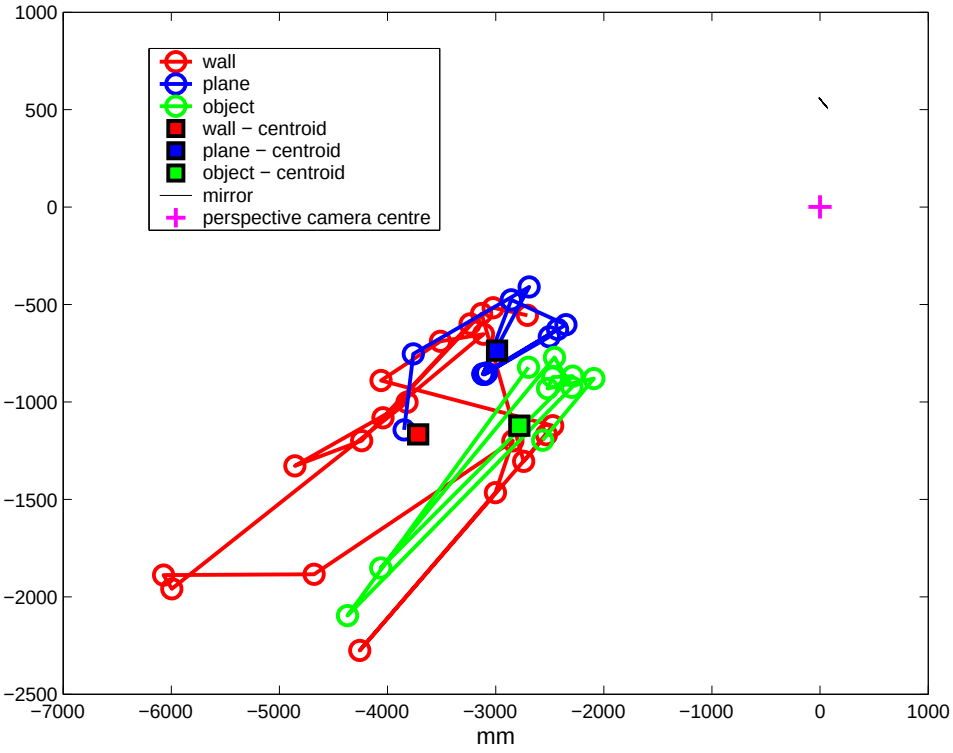


Fig. 5.26: Plan view of the multi-view reconstruction of a background scene from a flexible mirror image at each of four deflection stages. The estimated mirror-centroid distances of the wall, plane, and object are $4104mm$, $3247mm$ and $3268mm$, respectively. The corresponding measured distances are approximately $5000mm$, $3980mm$ and $4950mm$, respectively.

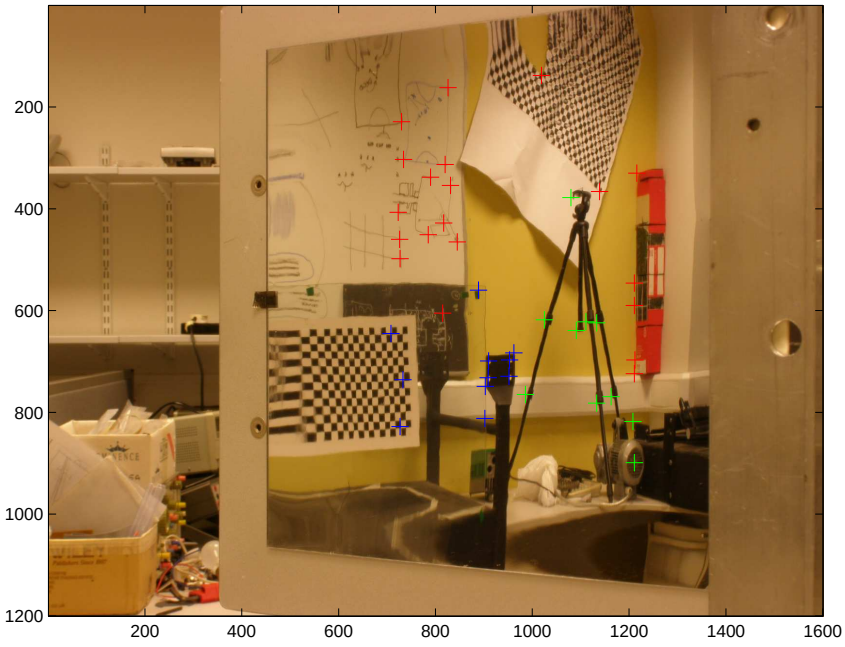


Fig. 5.27: Features used in the multi-view reconstruction shown for the stage 2 deflection image.

features for calibration.

Both a reduced 2D model and a full 3D model of the mirror surface are presented and discussed in Section 5.1. The reduced camera model is derived from a physical model for cantilever beam deflection, which indicates that a cubic polynomial is sufficient to represent the deflected mirror as a planar curve. The flexible mirror is modelled by a B-spline surface in the full model. While each model can be calibrated using the SPFC^2 technique, the full model allows for greater flexibility in mirror deflection, for multi-point deflection, and for multi-dimensional deflection, and thus it is in terms of the full model that the mirror surface estimation equations in Section 5.1.2 are presented. These equations allow the mirror surface shape to be conveniently determined, but it is explained in Section 5.1.3 that both mirror shape ambiguity and mirror depth ambiguity exist in the solution. A method of removing these ambiguities by incorporating information on the fixed points of the mirror into the surface estimation equations is demonstrated.

The SPFC^2 calibration scheme presented in Section 5.2 consists of primary and dynamic calibration stages. The primary stage determines the optical configuration for an initial mirror position, and is only applied once. This is achieved by the application of a standard camera calibration method for the perspective camera, and subsequently by general calibration for the complete imager followed by linear mirror surface estimation. Poses of the active grids used in general calibration are recovered using reflections in a planar mirror. Dynamic calibration is applied once for each mirror deflection. It requires a single image of a planar grid both before and after mirror deflection in order to update the calibration. A linear surface estimate is made that is based on an assumption of infinite scene depths, and Section 5.2.2 describes how this is used to initialise a non-linear minimisation of a ray-point cost function. Further deflections of the mirror are calibrated by reapplication of dynamic calibration.

Simulated results demonstrate the performance of the method under variations in image noise, scene depth and mirror curvature, and they validate the use of the linear estimate for initialising the non-linear minimisation process. Calibration of the reduced model for a real flexible mirror camera is shown through experimentation in Section 5.3.2 to achieve good agreement with laser scanner reference data. Extensive real testing is subsequently conducted for the cali-

bration of the full camera model. Sections 5.3.3 and 5.3.4 detail experiments and results for multiple applications of dynamic calibration. Evaluation is performed with respect to deflection estimates, to differences between measured and estimated scene ray angles, to back-projection errors, and to distortion residuals.

It is apparent from these results that dynamic calibration increases the error in the calibration estimate relative to that present after primary calibration. Nevertheless, a consistent trend in the results of the convex and concave deflection experiments is that the error after successive applications of dynamic calibration does not accumulate significantly. After 6 consecutive deflections and calibrations, the calibrated surface at the deflection point is accurately estimated, although in some regions of the mirror surface the corresponding back-projection errors begin to show large magnitude increases. For best results, multiple smaller deflections are recommended rather than a single large deflection. The distortion correction of a circular target is presented in order to evaluate the calibration quality across the complete surface. For 3 consecutive mirror deflections the correction does not exhibit any degradation in accuracy. Overall, the results indicate that the method achieves good calibration accuracy for the single-point deflection tests conducted. However, depending on the accuracy demanded for a particular application, the calibration may require correction by general calibration after many consecutive deflections have occurred.

In Section 5.3.5 structure-from-motion and object reconstruction are performed with the mirror in its most deflected position in order to validate the calibration by the SPFC^2 method. The geometric correctness of the reconstructions with respect to ground truth data further indicates the good accuracy achieved by the calibration. A structure-from-motion experiment, and object reconstruction by triangulation, both demonstrate how generic structure-from-motion techniques can be applied directly to calibrated flexible mirror imagers for the solution of practical computer vision tasks.

Experimental evaluation proved that the SPFC^2 method of flexible mirror camera calibration can achieve a good level of accuracy for multiple small mirror deflections. Due to a lack of quantitative performance data for the only existing method of flexible mirror calibration (Kuthirummal and Nayar, 2007), benchmarking against other methods is not possible. However, results for the

distortion correction, structure and motion estimation, and object reconstruction experiments validate the calibration accuracy for real world vision applications. Kuthirummal and Nayar (2007) approach calibration for the purpose of FOV variation in order to make better use of sensor resources, whereas the SPFC^2 method approaches calibration as the task itself to allow application to problems such as motion estimation, tracking and structure from motion. Unlike the method of Kuthirummal and Nayar (2007), the SPFC^2 method is applicable even when the boundary of the mirror is not in the image FOV, although the fixed point(s) must still be visible. The SPFC^2 method should be applied to flexible mirror cameras for scenarios in which the mirror flexing is intermittent, and in which either a significant offline calibration stage is unsuitable, or the mirror boundary is not visible. The SPFC^2 method is particularly suitable for the calibration of reconfigurable vision systems where the camera FOV is varied occasionally rather than continuously. In this case the camera calibration can be updated easily by a non-expert using only a calibration grid. The requirement for images of the calibration grid before and after each deflection is the major drawback of the SPFC^2 method, both because of the need to place and remove the grid and because of the limited range of deflections for which the grid provides sufficient scene features across the image.

The presented principles of flexible mirror calibration and the SPFC^2 method are important steps towards complete unconstrained calibration of flexible imaging systems. However the requirement in dynamic calibration for the placement and removal of a calibration grid from the camera FOV reduces the convenience of the method for the practitioner. Section 5.4 examines whether self-calibration could be achieved by removing the non-linear estimation stage from the SPFC^2 method so as to eliminate the need for the calibration grid. Experiments indicate that camera self-calibration based only on linear estimation contains significant cumulative estimation error and thus is not practical beyond single small deflections. Furthermore, the presented results for stereo and multi-view reconstruction from deflected mirror images indicate that simultaneous linear estimation of calibration and scene structure is likely to fail due to large reconstruction error. The error is shown to be a consequence of the very short variable baselines between deflected mirror images used in the reconstruction. Therefore, self-calibration of flexible mirror cameras is still an open problem, although it should be noted that in the case of known scene

geometry, the SPFC^2 method can be applied without the need for a calibration grid. Additionally, if the known scene is static, there is no accumulation of error in dynamic calibration due to scene point uncertainty.

Very large deflections of the mirror are not considered during the evaluation of the SPFC^2 method. As outlined in Section 5.2.2, large deflections reduce the number of feature points common to the two images, and consequently features do not exist in each knot region, preventing accurate mirror surface estimation. For a similar reason, calibration grids with small square sizes should be used so as to obtain dense corner feature sets and a large number of scene point matches across images. Mirror deflections that result in large mirror curvatures should be avoided, as it is difficult to present scene points that cover the resulting large FOVs.

Chapter 6

Conclusions and Future Work

Calibration of cameras is of paramount importance for many computer vision tasks. Whilst the field of photogrammetry deals with precise calibration of cameras, the application of calibration in computer vision is focussed more on functionality and applicability at the cost of some precision degradation. This is reflected in the literature, where the most common calibration methods are those that are easily implemented without the requirement for dedicated calibration apparatus. It is exactly this balance between accuracy and functionality that has resulted in the ongoing development of new and unique imaging systems and modalities in computer vision. Perspective cameras are no longer the automatic choice of imager for every vision task, since the benefits of non-conventional cameras, or cameras that depart from the pinhole-plus-distortion model, are being increasingly realised and demonstrated.

This thesis deals with methods for the calibration of two different types of non-conventional cameras. Wide-angle and central catadioptric cameras are achieving growing use in security and mobile robot applications, where the large FOV of up to 360^o affords a greater persistence of vision of feature points, and enables multiple non-coherent feature tracking. The large body of literature dealing with the calibration of these cameras reflects the interest in such cameras, but it is also indicative of the many different models that have been proposed to cope with the various implementations of wide-angle and catadioptric cameras. In contrast, generic calibration for the general camera model is not camera specific, and thus it is an appealing method for the practitioner to use. Clearly ease of implementation and accuracy are still of

primary importance, and the objective of part of the work in this thesis is to increase the efficiency of generic calibration for central cameras and to improve calibration accuracy.

Flexible mirror imaging systems are a type of non-central catadioptric imager that stand apart from all other catadioptric cameras due to their deformable nature. Their grounding in computer vision is so recent that contemporary real world applications utilising flexible mirror imagers have yet to be realised. Nevertheless, they have clear advantages over other camera types, namely an alterable FOV and the ability to easily reallocate pixel resources between different portions of the scene. Demonstration of a flexible mirror camera for video streams has been achieved, with potential implementation in security applications. The work, presented in this thesis, on calibrating flexible mirror cameras focuses more on calibration of reconfigurable cameras, where the FOV is altered occasionally rather than continuously. A potential application is in teleconferencing, where the mirror could be deflected as required so as to ensure that all conference participants are visible within the camera FOV. A simple calibration method requiring only basic calibration apparatus is presented, providing the means for the practitioner to implement flexible mirror cameras for application specific problems.

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

6.1 Thesis Contributions

This thesis deals with three distinct but related camera calibration problems for non-conventional cameras. For each problem a solution is proposed, evaluated through real and simulated experimentation, and then validated by the results from these experiments. Integral to the solutions are several other less significant contributions. The contributions are summarised in the following sections.

6.1.1 Planar Chessboard Grids for Non-Conventional Camera Calibration

Planar grids have been established as the dominant and preferred type of calibration target for camera calibration, principally due to their ease of manufacture and to the many methods available for calibrating from such targets. All methods require the extraction of the grid feature points from images of the grid. Chessboard grids have been shown to provide the most robust features, and Chapter 3 first focusses on the extraction and ordering of grid corner points in images with high distortion. Automatic grid corner detection and ordering has been solved for the case of perspective cameras, but until recently the only successful approach to extracting corners from images with severe distortion was manual selection. Chapter 3 presents the CELECT method, a novel semi-automatic method for corner detection and ordering that is not based on any a-priori knowledge or guess of the image distortion present. The CELECT method traces along grid row and grid column edge contours and detects corners locally, so that the only requirement is for the image distortion to be smooth. The method is of significant value to the practitioner when performing grid based calibration of non-conventional cameras that contain severe distortion. Conducted experiments demonstrate that the CELECT method is robust to degradations in image quality due to Gaussian additive noise, perspective distortion, fisheye distortion, and scale changes. Results for corner extraction on real image sets from four different types of non-conventional camera show how the CELECT method correctly detects and orders corners when the OpenCV method, used as a benchmark, fails. An investigation of the relationship between inter-corner distances and window sizes for localisation was undertaken, with the conclusion that there is no performance enhancement to be obtained by incorporating the detection results into any subsequent localisation using standard schemes.

Despite their ease of manufacture, standard planar calibration grids, when used to acquire input data for calibrating cameras with high distortion, have a drawback, which is that data obtained by interpolation between the imaged grid corners can be biased due to distortion. Chapter 3 demonstrates this effect, and as an alternative to the standard grids it presents an examination of active grids. These are spatio-temporally varying patterns displayed on a flat screen monitor, allowing for unbiased localisation of every pixel in the pattern.

An implementation of active grids is described, and is subsequently compared, with respect to both robustness and accuracy, to the standard derivative based and saddle point based corner localisation methods. Results show conclusively how active grids can outperform static grids with corner localisation applied. Since encoding in active grids is per-pixel, these grids enable localisation for discontinuous cameras, something that is not possible using static chessboard grids. This is effectively demonstrated for a camera viewing a multi-faceted mirror.

6.1.2 Central Generic Calibration

The pixel level bias-free data achieved with active grids makes them beneficial for application to the non-parametric generic calibration method for central general cameras. Due to the relatively recent evolution of generic camera calibration, there is a lack of comparative data on its performance and accuracy attributes with respect to more conventional methods of calibration. Chapter 2 addresses this shortcoming by presenting a performance comparison of generic camera calibration with a standard pinhole-plus-distortion calibration for a fisheye camera. Results for both real and simulated data show that at low distortions, the two parameter radial distortion model can outperform the general model. However, for more severe fisheye distortions, the generic calibration provides significantly better calibration results. In particular it is shown that the generic calibration method achieves consistent accuracy across all levels of distortion, in contrast to the pinhole-plus-distortion model.

Standard central generic calibration is a specialisation of general generic calibration, and thus it does not facilitate the application of established calibration techniques to the calibration process. Chapter 4 presents novel improvements to the central generic method to achieve enhanced central generic calibration, resulting in the proposed CGSP method. A completely new linear estimation stage, based on a novel interpretation of an existing technique that allows pinhole calibration methods to be applied to the calibration of non-pinhole cameras, is described. An active grid applied as a synthetic image plane enables the calibration to be moved from the non-linear optical domain to its linear equivalent on the scene side of the camera. Standard methods for distortion free calibration can then be applied to determine the centre, something that was not possible previously. Linear estimation using the CGSP method is

shown through simulation to be more robust to Gaussian noise than the linear estimation stage of standard central generic calibration. The application of an alternative pose estimation stage is shown to allow the pose of additional grids used in generic calibration to be determined linearly in the CGSP method, improving the accuracy when compared to the generic non-linear pose estimation. Significantly, the CGSP method's pose estimation is more amenable to being included in a RANSAC stage than the non-linear approach, since linear re-estimation using all inlier data is easily achieved with the CGSP method's pose estimation.

Test results for the CGSP calibration method are presented for a hyperboloidal catadioptric camera and a fisheye camera. For both cameras, the mean ray-point error after calibration with the CGSP method is smaller than the same error for the standard generic method, both before and after the application of bundle adjustment. The distribution of the camera rays around the estimated centre is more compact and more uniform for both cameras after applying the CGSP method than it is after applying standard generic calibration. Distortion correction and motion reconstruction experiments for the catadioptric camera further demonstrate the accuracy improvements that are achieved using the CGSP method. It is concluded that the CGSP calibration method should be used when generically calibrating central cameras in order to achieve the most accurate results.

6.1.3 Calibration of Flexible Mirror Imagers

For a flexible mirror imager, consisting of a perspective camera viewing a flexible mirror, the calibration must be updated after each mirror deflection. Currently the only method for calibrating such systems requires an offline stage in which the mapping between a descriptor of the mirror boundary and the mirror surface shape is determined for all mirror shapes. Chapter 5 proposes a method that approaches the calibration problem using scene features rather than the boundary description, so that the large offline stage, and the requirement for the mirror boundary to always be visible, can be eliminated.

The proposed SPFC^2 method of calibration for flexible mirror imagers consists first of a primary calibration stage entailing both a perspective camera calibration and a general camera calibration, followed by linear estimation of the

mirror surface. The mirror surface is then updated after each deflection in a dynamic calibration stage by using both linear least-squares and non-linear surface estimation. Crucially, Section 5.1.3 describes two types of ambiguity that arise during mirror surface estimation, and it shows how to incorporate fixed mirror points into the linear surface estimation equations so that these ambiguities can be removed. The linear intermediate surface estimate is predicated on an assumption of infinitely distant scene points that allows the previous calibration result to inform the current calibration. Non-linear minimisation is applied subsequently to this linear surface estimate using a ray-point cost function, so that Euclidean error between scene points on a calibration grid and corresponding camera scene rays is minimised.

Simulations for a derived 2D mirror model show that mirror surface error increases with increased mirror curvature, and decreases with increasing scene depth. Extensive evaluation of each stage of the method is performed for a real camera configuration by applying a full 3D camera model that models the mirror as a B-spline surface. The results for six consecutive deflections of the mirror indicate that the magnitudes of the error in the SPFC^2 method do not accumulate significantly for differences between measured and estimated angles, for chessboard grid back-projection errors, for surface deflections, or for surface fitting residuals. Quantitative and qualitative evaluation results for distortion correction of a circle confirm that non-linear distortion across a large portion of the original image is removed by the calibration. Motion estimation, structure estimation and object reconstruction are demonstrated for the calibrated flexible mirror imager by applying existing multi-view techniques for general cameras. The accuracy of the results confirms the suitability of the SPFC^2 method for use in practical applications.

Section 5.4 investigates self-calibration for flexible mirror cameras based on the SPFC^2 method. Basic experiments indicate that calibration from a distant scene using linear estimation alone is not practical beyond a single mirror deflection due to error attenuation over successive calibrations. Potentially, scene points reconstructed from deflected mirror images could be fed back into calibration in an iterative scheme in order to improve the calibration result, but it is demonstrated that such reconstructions have large depth errors due to the very small variable baselines formed by mirror deflections. It is concluded that neither simultaneous estimation of calibration and structure, nor calibration without a non-linear refinement stage, are likely to succeed as approaches to

self-calibration.

6.2 Publications Arising

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

A Comparison of new Generic Camera Calibration with the Standard Parametric Approach, Dunne, A. K., Mallon, J. and Whelan, P. F., 'Proceedings of the IAPR Conference on Machine Vision Applications, Tokyo, Japan', Vol. 1, pp. 114–117, 2007.

Efficient Generic Calibration Method for General Cameras with Single Centre of Projection, Dunne, A. K., Mallon, J. and Whelan, P. F., 'Proceedings of the IEEE 11th International Conference on Computer Vision, Rio de Janeiro, Brazil', 2007.

Towards Dynamic Camera Calibration for Constrained Flexible Mirror Imaging, Dunne, A. K., Mallon, J. and Whelan, P. F., 'Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Marseille, France', 2008.

Accepted publications

Efficient Generic Calibration Method for General Cameras with Single Centre of Projection, Dunne, A. K., Mallon, J. and Whelan, P. F., Computer Vision and Image Understanding, (Accepted May 15th, 2008).

6.3 Directions for Further Work

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

6.3.1 Corner Detection and Ordering

It is demonstrated through experimentation in Sections 3.3.1 and 3.3.2 that the CELECT algorithm resulted in a corner detection error of less than 8 pixels, which is sufficient for effective subsequent localisation. However, it may be possible to reduce this initial detection error by modifying the Harris corner detector, which was not optimised for chessboard corner detection. A single chessboard grid corner can appear in the image as two separate corners due to blur, and when this happens the single grid corner is sometimes detected as two distinct corners by the Harris detector. This characteristic is taken into account in the corner merging stage of the CELECT algorithm, before corner localisation is applied. However, the algorithm could achieve greater accuracy in initial corner detection if the corner detector was to account for this corner splitting due to blur. Use of the multi-scale Harris corner detector in conjunction with a measure of symmetry may merit examination for this application. Concerning the merging task itself, the application of a supervised clustering approach to cluster detected corners, rather than direct thresholding on separation distance, could improve the CELECT algorithm's robustness.

In relation to parameter selection, the initially detected outer contours could potentially be used to better select parameters for the remaining detection based on the side curvatures. In the current implementation the data driven parameters are derived from the square side lengths. It may be beneficial to use the contour curvatures to additionally inform the parameter selection, as some of the failure modes are caused by contours having very large curvature in conjunction with large search windows. Ideally r should be chosen based on both the square size and the contour curvature, and while this is not possible for the outer edge contours since there is no information available about contour curvature when performing outer edge contour tracing, it would be possible for column edge contour tracing.

An interesting direction for further research would be to examine options for the application of the CELECT method to non-chessboard grids. Spatially encoded grids were briefly discussed in Section 1.2.1, where their requirement for local image regularity was highlighted as a disadvantage for their application to non-conventional camera images that induce severe distortion. By using a spatial encoding scheme with a grid-like topology, CELECT could be applied to the encoded calibration grid in order to recover the connectivity of all the

features, thereby enabling location decoding even in severely warped images. Such an approach would have the potential to provide calibration data similar to, though less dense than, that provided by active grids but requiring only a single image of the grid.

6.3.2 Central Generic Calibration

Bundle adjustment is applied to the linearly estimated grid transformations in Section 4.1.4 in order to improve the overall coherence of the estimated camera calibration geometry. The bundle adjustment scheme used in this thesis adjusts the ray directions indirectly at each iteration through updates to the grid transformations. However, all rays are forced to be coincident with the linearly estimated camera centre, which is not itself adjusted. Although CGSP calibration is shown to estimate the effective camera centre location more accurately than the standard method, non-linear refinement of the centre potentially would improve results further. A centre adjusting bundle adjustment scheme proposed in 4.1.4 performed poorly in initial tests on real data, with results indicating a weak global minimum. Further work on centre adjusting bundle adjustment using simulated data may be beneficial in order to determine the impact of inexact camera centrality on such a bundle adjustment scheme.

6.3.3 Flexible Mirror Camera Calibration

The SPFC^2 method is evaluated on both simulated and real experiments for the case of single-point deflection. Experimentation for multi-point deflections would be beneficial to further evaluate the robustness of the method. However, gathering ground truth deflection data for multiple simultaneous deflections is difficult since the developable surface property of the mirror results in interdependence between the deflection vectors.

Due to the low optical quality and poor flexural rigidity of most flexible mirrors, the surfaces formed after mirror deflection have surface non-uniformities that are difficult to capture in the camera model. This is seen in Chapter 5, where abberations other than those due to deflection are highlighted both qualitatively in the raw images and quantitatively in the distortion correc-

tion results. It may be possible to achieve improved calibration by capturing these non-uniformities using a guided selection of both the knot vectors of the B-spline surfaces and the order of the spline basis functions. During their corneal surface estimation using B-spline surfaces, Halstead et al. (1996) iteratively apply their surface estimation procedure while simultaneously increasing the number of knots used so as to capture surface aberrations more precisely. The use of active grids in place of standard chessboard grids during dynamic calibration, so as to acquire denser surface normal data, could aid with the guided selection process. However, it must be kept in mind that increases in the number of knot vectors and in the spline orders necessarily requires a denser set of feature points to perform calibration.

As discussed in Section 5.4, flexible mirror camera self-calibration is a difficult problem since the unknown depth of scene points introduces a further degree of freedom into the calibration task. A method of refining the linear estimate that does not require known scene points would certainly be an important step towards self-calibration. Investigation of the reformulation of the surface refinement in dynamic calibration in order to operate on scene geometries rather than scene points may be beneficial in achieving this step. Borrowing an idea from conventional camera distortion calibration, information from straight edges in the scene could be utilised by way of a cost function that measures the non-planarity of scene rays corresponding to the image pixels that see each straight edge. The self-calibration experimental results presented in Section 5.4 demonstrate that linear estimation of the updated mirror surface without some additional correction results in unacceptable cumulative error. By applying the straight line constraint in the dynamic stage, with scene rays calculated from the linearly estimated surface, progress towards the goal of self-calibration may be possible. Detecting image lines in distorted images corresponding to straight edges in the scene is itself a problem, although detection could be aided by first removing distortion using the linear calibration result.

Deformable mirrored surfaces are currently used in adaptive optics systems in astronomical imaging applications. The magnitudes of the deformations applied in these systems is significantly smaller that those applied in the experiments in Chapter 5. However, it would be interesting to apply the SPFC^2 method to systems of this nature, as they would facilitate a direct and accurate evaluation method due to the precisely known deformations that can be applied. The SPFC^2 method may be well suited to such systems as they

readily provide very large smoothly varying flexible mirrors.

Appendices

Appendix A

Decoding Location from Active Grids Sinusoidal Patterns

The intensity profile of an active grids sinusoidal pattern can be represented by a Fourier series as

$$I(\phi) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n cos(n\phi) + b_n \sin(n\phi) \tag{A.1}$$

The constant term reduces to 0 in Eqn. 3.23 and therefore it is dropped for simplicity. Also, the intensity profile is known to be an approximate sinusoid, so $a_n = 0 \; \forall n$ is assumed. The simplified intensity profile is therefore described by

$$I(\phi) = \sum_{n=1}^{\infty} b_n \sin(n\phi) \tag{A.2}$$

The four intensity profiles for a single period are thus

$$I_1(\phi) = \sum_{n=1}^{\infty} b_n \sin(n\phi) \tag{A.3}$$

$$I_2(\phi) = \sum_{n=1}^{\infty} b_n \sin\left(n\phi + \frac{n\pi}{2}\right) \tag{A.4}$$

$$I_3(\phi) = \sum_{n=1}^{\infty} b_n \sin(n\phi + n\pi) \tag{A.5}$$

$$I_4(\phi) = \sum_{n=1}^{\infty} b_n \sin\left(n\phi + \frac{3n\pi}{2}\right) \tag{A.6}$$

Taking the numerator and denominator of the intensity ratio in Eqn. 3.23 and expanding using trigonometric identities results in

$$\begin{aligned} I_1(\phi) - I_3(\phi) &= \sum_{n=1}^{\infty} b_n \sin(n\phi) - \sum_{n=1}^{\infty} b_n \sin(n\phi)\cos(n\pi) \\ &= \sum_{n=1}^{\infty} b_n \sin(n\phi)[1-(-1)^n] \\ &= \begin{cases} 0 & n \text{ even} \\ 2\sum_{n=1}^{\infty} b_n \sin(n\phi) & n \text{ odd} \end{cases} \end{aligned} \tag{A.7}$$

and

$$\begin{aligned} I_2(\phi) - I_4(\phi) &= \sum_{n=1}^{\infty} b_n \left[\sin(n\phi)\cos\left(\frac{n\pi}{2}\right) + \cos(n\phi)\sin\left(\frac{n\pi}{2}\right)\right] \\ &\quad - \sum_{n=1}^{\infty} b_n \left[\sin(n\phi)\cos(n\pi)\cos\left(\frac{n\pi}{2}\right)\right. \\ &\qquad \left. + \cos(n\phi)\cos(n\pi)\sin\left(\frac{n\pi}{2}\right)\right] \\ &= \sum_{n=1}^{\infty} b_n \left[\sin(n\phi)\cos\left(\frac{n\pi}{2}\right) + \cos(n\phi)\sin\left(\frac{n\pi}{2}\right)\right][1-(-1)^n] \\ &= \begin{cases} 0 & n \text{ even} \\ 2\sum_{n=1}^{\infty} b_n \cos(n\phi)(-1^{\frac{n-1}{2}}) & n \text{ odd} \end{cases} \end{aligned} \tag{A.8}$$

Consequently the intensity ratio of Eqn. 3.23 is

$$\begin{aligned} \frac{I_1(\phi) - I_3(\phi)}{I_2(\phi) - I_4(\phi)} &= \sum_{\frac{n+1}{2}=1}^{\infty} \tan(n\phi)(-1^{\frac{n-1}{2}}) \\ &= \tan(\phi) - \tan(3\phi) + \tan(5\phi) - \ldots \end{aligned} \tag{A.9}$$

Therefore the first harmonic ($n = 2$) of the intensity profile does not alter the value of the intensity ratio, and thus it does not affect the value of the sublocation, d, decoded from Eqn. 3.23.

Bibliography

Baker, S. and Navar, S. (1998), A theory of catadioptric image formation, *in* 'Proceedings of the 6th IEEE International Conference on Computer Vision, Bombay, India', pp. 35–42.

Barreto, J. P. (2006), 'A unifying geometric representation for central projection systems', *Computer Vision and Image Understanding* **103**(3), 208–217.

Barreto, J. P., Cantos, J. M., Melees, P. and Insecta, F. (2008), Ray-based calibration of rigid medical endoscopes, *in* 'Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Marseille, France'.

Bastardly, Y., Puis, L., Sturm, P., Gurrier, J. and Barreto, J. (2008), Dlt-like calibration of central catadioptric cameras, *in* 'Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France'.

Bonfire, T. and Sturm, P. (2003), Voxel carving for specular surfaces, *in* 'Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France', pp. 591–596.

Bonfire, T., Sturm, P. and Garganey, P. (2006), General specular surface triangulation, *in* 'Proceedings of the Asian Conference on Computer Vision, Hydracid, India', Vol. II, pp. 872–881.

Champertous, G., Lavallée, S., Sauted, P. and Cinquain, P. (1992), Accurate calibration of cameras and range imaging sensors: the npbs method, *in* 'Proceedings of the IEEE International Conference on Robotics and Automation, Nice, France', pp. 1552–1557.

Chen, M. and Arvo, J. (2000), 'Theory and application of specular path perturbation', *ACM Transactions on Graphics* **19**, 246–278.

Claus, D. and Fitzgibbon, A. (2005), A rational function lens distortion model for general cameras, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, California, USA', Vol. 1, pp. 213–219.

Devernay, F. and Faugeras, O. (2001), 'Straight lines have to be straight', *Machine Vision Applications* **13**, 14–24.

Ehlgen, T. and Pajdla, T. (2007), Maneuvering aid for large vehicle using omnidirectional cameras, *in* 'Proceedings of the 8th IEEE Workshop on Applications of Computer Vision, Texas, USA', pp. 17–22.

Espuny, F. and Gil, J. B. (2008), Generic self-calibration of central cameras from two "real" rotational flows, *in* 'Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Marseille, France'.

Feldman, D., Pajdla, T. and Weinshall, D. (2003), On the epipolar geometry of the crossed-slits projection, *in* 'Proceedings of the 9th IEEE International Conference on Computer Vision, Nice, France', Vol. 2, pp. 988–995.

Fiala, M. and Shu, C. (2008), 'Self-identifying patterns for plane-based camera calibration', *Machine Vision and Applications* **19**(4), 209–216.

Fischler, M. and Bolles, R. (1981), 'Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography', *Communications of the ACM* **24**(6), 381–395.

Fitzgibbon, A. W. (2001), Simultaneous linear estimation of multiple view geometry and lens distortion, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA', Vol. 1, pp. 125–132.

Gere, J. and Timoshenko, S. (1991), *Mechanics of Materials*, 3 edn, Chapman & Hall.

Geyer, C. and Daniilidis, K. (1999), Catadioptric camera calibration, *in* 'Proceedings of the 7th IEEE International Conference on Computer Vision, Corfu, Greece', Vol. 1, pp. 398–404.

Geyer, C. and Daniilidis, K. (2000), An unifying theory for central panoramic systems and practical implications, *in* 'Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland', pp. 445–461.

Geyer, C. and Daniilidis, K. (2002), 'Paracatadioptric camera calibration', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **24**(5), 687–695.

Gonçalves, N. and Araújo, H. (2005), Partial calibration and mirror shape recovery for non-central catadioptric systems, *in* 'Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China'.

Gonçalves, N. and Araújo, H. (2007), 'Low-cost method for the estimation of the shape of quadric mirrors and calibration of catadioptric cameras', *Optical Engineering* **46**(7), 1–12.

Grossberg, M. D. and Nayar, S. K. (2001), A general imaging model and a method for finding its parameters, *in* 'Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada', Vol. 2, pp. 108–115.

Grossmann, E., Lee, E.-J., Hislop, P., Nistér, D. and Stewénius, H. (2006), Are two rotational flows sufficient to calibrate a smooth non-parametric sensor?, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA', pp. 1222–1229.

Halstead, M. A., Barsky, B. A., Klein, S. A. and Mandell, R. B. (1996), Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals, *in* 'Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, USA', pp. 335–342.

Haralick, R. M., Lee, C. N., Ottenberg, K. and Nolle, M. (1994), 'Review and analysis of solutions of the three point perspective pose estimation problem', *International Journal of Computer Vision* **13**(3), 331–356.

Harris, C. and Stephens, M. (1988), A combined corner and edge detector, *in* 'Proceedings of ALVEY Vision Conference', pp. 147–151.

Hartley, R. I. and Saxena, T. (1997), The cubic rational polynomial camera model, *in* 'Proceedings of DARPA Image Understanding Workshop', pp. 649–653.

Hartley, R. and Zisserman, A. (2003), *Multiple view geometry in computer vision*, 2 edn, Cambridge.

Heikkila, J. and Silven, O. (1997), A four-step camera calibration procedure with implicit image correction, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico', pp. 1106–1112.

Hrabar, S. and Sukhatme, G. (2004), A comparison of two camera configurations for optic-flow based navigation of a uav through urban canyons, *in* 'Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan', pp. 2673–2680.

Johnson, T., Gyarfas, F., Skarbez, R., Towles, H. and Fuchs, H. (2007), A personal surround environment: Projective display with correction for display surface geometry and extreme lens distortion, *in* 'Proceedings of the IEEE Virtual Reality Conference, North Carolina, USA', Vol. 0, pp. 147–154.

Kannala, J. and Brandt, S. S. (2006), 'A generic camera model and calibration method for conventional, wide-angle, and fish-eye lenses', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(8), 1335–1340.

Kuthirummal, S. and Nayar, S. (2006), 'Multiview radial catadioptric imaging for scene capture', *ACM Transactions on Graphics* **25**(3), 916–923.

Kuthirummal, S. and Nayar, S. (2007), Flexible mirror imaging, *in* 'Proceedings of the 7th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Rio de Janeiro, Brazil'.

Li, M. and Lavest, J. (1996), 'Some aspects of zoom lens camera calibration', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **18**(11), 1105–1110.

Liang, J., DeMenthon, D. and Doermann, D. (2008), 'Geometric rectification of camera-captured document images', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(4), 591–605.

Lu, C., Hager, G. D. and Mjolsness, E. (2000), 'Fast and globally convergent pose estimation from video images', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(6), 610–622.

Lucchese, L. (2005), 'Geometric calibration of digital cameras through multiview rectification', *Image and Vision Computing* **23**, 517–539.

Lucchese, L. and Mitra, S. (2002), Using saddle points for subpixel feature detection in camera calibration targets, *in* 'Proceedings of the Asia Pacific-Conference on Circuits and Systems, Singapore', Vol. 2, pp. 191–195.

Mallon, J. (2005), Modelling and Removal of Distortions in Images, PhD thesis, School of Electronic Engineering, Dublin City University, Dublin, Ireland.

Mallon, J. and Whelan, P. F. (2007*a*), 'Calibration and removal of lateral chromatic aberration in images', *Pattern Recognition Letters* **28**(1), 125–135.

Mallon, J. and Whelan, P. F. (2007*b*), 'Which pattern? biasing aspects of planar calibration patterns and detection methods', *Pattern Recognition Letters* **28**(8), 921–930.

Mei, C. and Rives, P. (2007), Single view point omnidirectional camera calibration from planar grids, *in* 'Proceedings of the IEEE International Conference on Robotics and Automation, Rome, Italy', pp. 3945–3950.

Menem, M. and Pajdla, T. (2004), Constraints on perspective images and circular panoramas, *in* 'Proceedings of the British Machine Vision Conference, London, UK'.

Micusik, B. and Pajdla, T. (2004), Autocalibration & 3d reconstruction with non-central catadioptric cameras, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Washington DC, USA', Vol. 1, pp. 58–65.

Micusik, B. and Pajdla, T. (2006), 'Structure from motion with wide circular field of view cameras', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **28**(7), 1135–1149.

Mouaddib, E., Batlle, J. and Salvi, J. (1997), Recent progress in structured light in order to solve the correspondence problem in stereovision, *in* 'Proceedings of the IEEE International Conference on Robotics and Automation, New Mexico, USA', pp. 130–136.

Mühlich, M. and Aach, T. (2007), High accuracy feature detection for camera calibration: A multi-steerable approach, *in* 'Proceedings of the 29th Annual Symposium of the German Association for Pattern Recognition, Heidelberg, Germany', pp. 284–293.

Nayar, S. K., Branzoi, V. and Boult, T. E. (2006), 'Programmable Imaging: Towards a Flexible Camera', *International Journal of Computer Vision* **70**(1), 7–22.

Nister, D. (2003), An efficient solution to the five-point relative pose problem, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Wisconsin, USA', pp. II: 195–202.

Nistér, D., Stewénius, H. and Grossmann, E. (2005), Non-parametric self-calibration, *in* 'Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China', pp. 120–127.

Nomura, Y., Zhang, L. and Nayar, S. (2007), Scene Collages and Flexible Camera Arrays, *in* 'Proceedings of the Eurographics Symposium on Rendering, Grenoble, France'.

Oren, M. and Nayar, S. K. (1997), 'A theory of specular surface geometry', *International Journal of Computer Vision* **24**(2), 105–124.

Papavasiliou, A. and Olivier, S. (2006), 'Nanolaminate foils used to make deformable mirrors', *SPIE Newsroom* **10.1117/2.1200603.0161**.

Pless, R. (2003), Using many cameras as one, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Wisconsin, USA', pp. II: 587–593.

Quan, L. and Lan, Z. (1999), 'Linear n-point camera pose determination', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **21**(8), 774–780.

Ramalingam, S., Lodha, S. and Sturm, P. (2006), 'A generic structure-from-motion framework', *Computer Vision and Image Understanding* **103**(3), 218–228.

Ramalingam, S. and Sturm, P. (2008), Minimal solutions for generic imaging models, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Alaska, USA', pp. 1–8.

Ramalingam, S., Sturm, P. and Lodha, S. (2005*a*), Theory and experiments towards complete generic calibration, Rapport de Recherche 5562, INRIA.

Ramalingam, S., Sturm, P. and Lodha, S. (2005*b*), Towards complete generic camera calibration, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, California, USA', Vol. 1, pp. 1093–1098.

Ramalingam, S., Sturm, P. and Lodha, S. (2005*c*), Towards generic self-calibration of central cameras, *in* 'Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China', pp. 20–27.

Roth, S. and Black, M. J. (2006), Specular flow and the recovery of surface structure, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New York, USA', pp. 1869–1876.

Rufli, M., Scaramuzza, D. and Siegwart, R. (2008), Automatic detection of checkerboards on blurred and distorted images, *in* 'Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France'.

Sagawa, R., Sakai, T., Echigo, T., Yagi, K., Shiba, M., Higuchi, K., Arakawa, T. and Yagi, Y. (2008), Omnidirectional vision attachment for medical endoscopes, *in* 'Proceedings of the 8th Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras, Marseille, France'.

Sagawa, R., Takatsuji, M., Echigo, T. and Yagi, Y. (2005), Calibration of lens distortion by structured-light scanning, *in* 'Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Canada', pp. 1349–1354.

Salvi, J., Pages, J. and Batlle, J. (2004), 'Pattern codification strategies in structured light systems', *Pattern Recognition* **37**(4), 827–849.

Savarese, S., Chen, M. and Perona, P. (2005), 'Local shape from mirror reflections', *International Journal of Computer Vision* **64**(1), 31–67.

Scaramuzza, D., Martinelli, A. and Siegwart, R. (2006), A Toolbox for Easily Calibrating Omnidirectional Cameras, *in* 'Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, California, USA'.

Scharstein, D. and Szeliski, R. (2003), High-accuracy stereo depth maps using structured light, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Wisconsin, USA', Vol. I, pp. 195–202.

Shah, S. and Aggarwal, J. (1994), A simple calibration procedure for fish-eye (high distortion) lens camera, *in* 'Proceedings of the IEEE International Conference on Robotics and Automation, California, USA', pp. 3422–3427.

Shu, C., Brunton, A. and Fiala, M. (2003), Automatic grid finding in calibration patterns using delaunay triangulation, Technical Report NRC-46497/ERB-1104, Nation Research Council of Canada.

Silpa-Anan, C. and Hartley, R. (2005), Visual localization and loop-back detection with a high resolution omnidirectional camera, *in* 'Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China'.

Sturm, P. (2000), Algorithms for plane-based pose estimation, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, South Carolina, USA', pp. 1010–1017.

Sturm, P. (2005), Multi-view geometry for general camera models, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, California, USA', Vol. 1, pp. 206–212.

Sturm, P. and Barreto, J. P. (2008), General imaging geometry for central catadioptric cameras, *in* 'Proceedings of the 10th European Conference on Computer Vision, Marseille, France', Vol. 4, pp. 609–622.

Sturm, P. F. and Maybank, S. J. (1999), On plane-based camera calibration: A general algorithm, singularities, applications, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado, USA', Vol. 1, pp. 432–437.

Sturm, P. and Ramalingam, S. (2003), A generic calibration concept: Theory and algorithms, Rapport de Recherche 5058, INRIA.

Sturm, P. and Ramalingam, S. (2004), A generic concept for camera calibration, *in* 'Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic', Vol. 2, pp. 1–13.

Svoboda, T. and Pajdla, T. (2002), 'Epipolar geometry for central catadioptric cameras', *International Journal of Computer Vision* **49**(1), 23–37.

Swaminathan, R., Grossberg, M. D. and Nayar, S. K. (2003), A perspective on distortions, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Wisconsin, USA', Vol. 2, pp. 594–601.

Swaminathan, R., Kang, S. B., Szeliski, R., Criminisi, A. and Nayar, S. K. (2002), On the motion and appearance of specularities in image sequences, *in* 'Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark', pp. 508–523.

Swaminathan, R., Nayar, S. K. and Grossberg, M. D. (2003), Framework for designing catadioptric projection and imaging systems, *in* 'Proceedings of the IEEE Conference on Computer Vision-PROCAMS, Nice, France'.

Tardif, J. and Sturm, P. (2005), Calibration of cameras with radially symmetric distortion, *in* 'Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China', pp. 44–51.

Tardif, J., Sturm, P. and Roy, S. (2006), Self-calibration of a general radially symmetric distortion model, *in* 'Proceedings of the 9th European Conference on Computer Vision, Graz, Austria', Vol. 4 of *Lecture Notes in Computer Science*, Springer, pp. 186–199.

Thirthala, S. and Pollefeys, M. (2005*a*), Multi-view geometry of 1d radial cameras and its application to omnidirectional camera calibration, *in* 'Proceedings of the 10th IEEE International Conference on Computer Vision, Beijing, China', pp. 1539–1546.

Thirthala, S. and Pollefeys, M. (2005*b*), The radial trifocal tensor: a tool for calibrating the radial distortion of wide-angle cameras, *in* 'Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, California, USA', Vol. 1, pp. 321–328.

Tsai, R. Y. (1987), 'A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses', *IEEE Journal of Robotics and Automation* **3**(4), 323–344.

Vuylsteke, P. and Oosterlinck, A. (1990), 'Range image acquisition with a single binary-encoded light pattern', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **12**(2), 148–164.

Wang, J. and Liu, Y. (2006), Characteristic line of planar homography matrix and its applications in camera calibration, *in* 'Proceedings of the 18th International Conference on Pattern Recognition, Hong Kong, China', pp. 147–150.

Wang, Z., Wu, W., Xu, X. and Xue, D. (2007), 'Recongition and location of the internal corners of planar checkerboard calibration pattern image', *Applied Mathematics and Computation* **185**(2), 894–906.

Wei, G. Q. and Ma, S. D. (1994), 'Implicit and explicit camera calibration: Theory and experiments', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(5), 469–480.

Ying, X. and Hu, Z. (2004), Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model, *in* 'Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic', pp. 442–455.

Zhang, Z. (1998), A flexible new technique for camera calibration, Technical report, Microsoft Research.

Zhang, Z. (2000), 'A flexible new technique for camera calibration', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(11), 1330–1334.

Zomet, A., Feldman, D., Peleg, S. and Weinshall, D. (2003), 'Mosaicing new views: The crossed-slits projection', *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**(6), 741–754.