

THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Camera Modelling and Calibration with Machine Vision Applications

ANDERS RYBERG



Department of Signals and Systems
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2010



Department of Engineering Science
UNIVERSITY WEST
Trollhättan, Sweden, 2010

Camera Modelling and Calibration with Machine Vision Applications

ANDERS RYBERG

ISBN 978-91-7385-365-1

Doktorsavhandlingar vid Chalmers tekniska högskola

Ny serie Nr 3046

ISSN 0346-718X

Department of Signals and Systems

Automation Research Group

Chalmers University of Technology

SE-412 96 Göteborg, Sweden

Telephone +46 (0) 31 772 10 00

© 2010 Anders Ryberg

Printed by Chalmers Reproservice

Göteborg, Sweden 2010

to my wife, Zhi

Abstract

Camera Modelling and Calibration with Machine Vision Applications

Anders Ryberg

Department of Signals and Systems

Chalmers University of Technology

Camera modelling and calibration are important parts of machine vision. They can be used for calculating geometric information from images. A camera model is a mathematical projection between a 3D object space and a 2D image. The camera calibration is a mathematical procedure calculating parameters of the camera model, usually based on several images of reference points. These fundamental parts of machine vision are improved in this thesis. One large part is the development of a generic camera model, GCM, that is accurate, computationally efficient and can be used for both conventional, fisheye and even catadioptric cameras. Different models were used in the past for conventional and omnidirectional cameras and this is a well-known problem, the solution of which is described in this thesis.

The accuracy of camera models is improved by introducing new ways of compensating for different distortions, such as radial distortion, varying entrance pupil point and decentring distortion. Calibration is improved by introducing new means of calculating start estimates of camera parameters, from analysing shapes, sizes and positions of the reference points in the images. These start estimates are needed in order to make the calibration converge. Methods for calculating better reference centre points than the centres of gravity are developed in order to increase the accuracy further. Non-trivial null spaces that occur during calibration are identified. Awareness of these improve the calibration.

Calibrations with different camera models are implemented and tested for real cameras in order to compare their accuracy. Certain models are better for certain situations, but the overall performance and properties are favourable for the GCM. A stereo vision welding robot system is developed, using the new model. It determines the geometry of a 3D weld joint, so that a robot can follow it. The same system is implemented in a virtual environment using a simulation software. Such simulation is important since it makes it possible to develop robot vision systems off-line.

KEYWORDS: Camera Model, Camera Calibration, Fisheye Camera, Catadioptric Camera, Stereo Vision.

Acknowledgments

First I want to thank the Knowledge Foundation, the EU research project Affix and University West for financing this work, which has been very interesting to me.

A number of people have been helpful, whom I want to thank. My main supervisor professor Bengt Lennartson has been guiding me and he has stayed up late nights to review my articles and give me valuable comments. I want to thank my local supervisor doctor Anna-Karin Christiansson, for being enthusiastic and for always taking the time to help me when needed, although a very busy person. Professor Kenneth Eriksson deserves a special thanks since he has always been interested and involved in my projects. Professor Lars Asplund deserves many thanks for letting me borrow high quality camera equipment for my camera model and calibration experiments, and for being the co-author of one of the papers. Mr Jan Fäger, for introducing me to problems needing to be solved in the field. The solutions to those problems constitute a large part of this thesis, such as leaning detector surface compensation and estimating camera parameters based on the shape of references in the images. I thank doctor Mikael Ericsson for working as an additional supervisor; you have been a valuable extra resource when needed and also the first author of one of the papers. I also want to thank Martin Byröd for useful comments and suggestions. Master thesis students Jim Nilsson and Mathias Larsson have done a lot of work on implementing some of my robot vision algorithms in the robot cell; I thank you for that. I thank Leona Johanson-Bunting for proofreading. I also want to thank my wife YingZhi Zhu Ryberg, to whom this thesis is dedicated. Thank you for standing by my side and always supporting me.

*Anders Ryberg
Trollhättan, February 2010*

Publications

The thesis is mainly based on the following five publications. The first is an early conference paper. The second is a published book chapter. The third and fifth are submitted journal publications. The fourth is an accepted conference paper.

Anders Ryberg, Anna-Karin Christiansson, Kenneth Eriksson. Accuracy Investigation of a Vision Based System for Pose Measurements. In *The 2006 9th IEEE International Conference on Automation, Robotics, Control and Vision*, Singapore, Dec 2006.

Anders Ryberg, Anna-Karin Christiansson, Bengt Lennartson and Kenneth Eriksson. Camera Modelling and Calibration - with Applications. In book *Computer Vision*, Publisher: I-Tech Education and Publishing, Vienna, Austria, 303-332, 2008.

Partially published as A new Camera Model for Higher Accuracy Pose Calculations. In *Proceedings of the 2006 IEEE International Symposium on Industrial Electronics*, Montreal, Canada, July 2006, and A new Camera Model and Algorithms for higher Accuracy and better Convergence in Vision based Pose Calculations. In *Proceedings of the 2006 IEEE International Conference on Mechatronics and Automation*, Luoyang, China, June 2006.

Anders Ryberg, Bengt Lennartson, Anna-Karin Christiansson, Lars Asplund and Mikael Ericsson. Analysis and Evaluation of a Generic Camera Model. Submitted to Elsevier *Journal of Computer Vision and Image Understanding* 2010

Anders Ryberg, Mikael Ericsson, Anna-Karin Christiansson, Kenneth Eriksson, Jim Nilsson and Mathias Larsson. Stereo Vision for Path Correction in Off-Line Programmed Robot Welding. In *The 2010 IEEE International Conference on Industrial Technology*, Viña del Mar - Valparaiso, Chile, March 2010.

Mikael Ericsson, Anders Ryberg, Jim Nilsson, Anna-Karin Christiansson and Bengt Lennartson. Off-line Simulation of Advanced Stereo Vision Welding Application. Submitted to Springer *Journal of Machine Vision and Applications* 2010

Abbreviations and Notations

Abbreviation / Notation	Description
CCM	Conventional Camera Model
CMM	Coordinate Measuring Machine
FCM	Fisheye Camera Model
GCM	Generic Camera Model
PCM	Pinhole Camera Model
α	Angle between optical axis and object line
γ_d, γ_r	Image angle to x_d and x_r
λ	Parameter
μ	Aspect ratio
M, B	Coordinate transformation matrices
e_x, e_y, e_z	Unit vectors spanning camera coordinate system
f	Focal distance
$f_f(r_p), f_r(r)$	Functions compensating for radial distortion
$f_i(r)$	Function compensating for distortion in the GCM
$f_l(r)$	Defines the exit pupil point position
$f_o(r)$	Defines the entrance pupil point position
k_i	Camera parameters
R	Rotation matrix, for rotation between coordinate systems
r, r_d, r_p	Distance from x_r, x_d and x_p to the principal point
s	Compensation for non perpendicular image coordinate system
t	Translation vector for translation between coordinate systems
x_{ca}	Principal point position
x_0^{dc}	Image detector coordinate of principal point
x_d	Decentring distorted image point
x_{dr}, x_{rr}	Coordinates used for leaning detector compensation
x_{fi}	Auxiliary point in the GCM
x_{fo}	Entrance pupil point
x_o	Object point or reference point coordinates
x_p	Undistorted image point (pinhole)
x_r	Radially distorted image point

Index lower/upper	Description
p	PCM coordinate (lower index)
r	Radially distorted point
d	Decentering distorted point
c	Detector chip coordinate
$1,2,3$	x -, y - and z - components of a vector, or numbering of camera parameters
w	3D world coordinate system (upper index)
c	3D camera coordinate system
i	2D image coordinate system
dc	2D detector chip coordinate system

Contents

Abstract	v
Acknowledgments	vii
Publications	ix
Abbreviations and Notations	xi
Contents	xiii

Introductory Chapters

Chapter 1: Introduction	3
1.1 Background	3
1.2 Objectives	4
1.3 Research Questions	4
1.4 Main Contributions	5
1.5 Limitations	5
1.6 Outline of Thesis	6
Chapter 2: Camera Modelling and Calibration	7
2.1 Camera Models	7
2.1.1 Introduction	7
2.1.2 Pinhole Camera Model	8
2.1.3 Radial Distortion	10
2.1.4 Decentering Distortion	12
2.1.5 Varying Entrance Pupil Point	13
2.1.6 Catadioptric Cameras	13
2.2 New Generic Camera Model	14
2.2.1 Radial Distortion and Varying Entrance Pupil Point	14
2.2.2 Decentering Distortion	17
2.2.3 Detector Chip Coordinates	19
2.2.4 Alternative Camera Model Representation	20
2.2.5 Variable Focus and Zoom	21
2.3 Camera Calibration	21
2.3.1 Pre-Processing Algorithms	23
2.3.2 Non-Trivial Null Spaces	23
2.3.3 Total Calibration Process	24
2.4 Experimental Verification of Camera Models	24
2.4.1 Error Norms	24
2.4.2 Experimental Setup for Comparison	26

2.4.3	Model Comparison	26
2.4.4	Discussion	29
Chapter 3:	Geometric Calculations from Camera Models	33
3.1	Pose Calculation	33
3.2	Stereo Vision	34
3.3	Experimental Verification of Pose Measurement and Stereo Vision	36
Chapter 4:	Related Computer Vision	39
4.1	Image Processing	39
4.2	Image Part Recognition	40
4.2.1	Feature Selection	40
4.2.2	Image Segmentation	40
4.2.3	Feature Extraction	41
4.2.4	Classification	41
4.2.5	Recognition of Objects with Known Geometries	41
4.2.6	Point Feature Detectors and Descriptors	42
4.3	Tracking	42
4.4	Visual Servoing	42
4.5	Virtual Vision	43
4.6	Other Computer Vision Applications	43
Chapter 5:	Conclusions and Future work	45
Chapter 6:	Summary of Appended Papers	47

Bibliography

Bibliography	53
--------------	----

Included Papers

Paper I: Accuracy Investigation of a Vision Based System for Pose Measurements	61
Paper II: Camera Modelling and Calibration - with Applications	75
Paper III: Analysis and Evaluation of a Generic Camera Model	105
Paper IV: Stereo Vision for Path Correction in Off-Line Programmed Robot Welding	131
Paper V: Virtual Vision for off-line Programing of a Robot Vision System	145

Introductory Chapters

Introduction

1.1 Background

This thesis considers applications where certain tasks are performed with the aid of cameras. Such systems are often denoted “machine vision” systems. This work was initiated for machine vision in robot applications, but has led to more general solutions and developments of camera systems. A vision system can guide traditional industrial robots, see Figure 1.1, as well as different types of autonomous robots. The main focus of this thesis is to calculate geometric information based on images to e.g. be used by robots. Camera modelling and calibration are important parts of vision systems for determining such geometric information.

When starting this work a system called PosEye, developed by MEEQ, Sweden, was available. The system calculated camera poses, i.e. positions and orientations, based on images. This system was meant to measure robot poses in the industry. By mounting the camera on the robot hand and using a transformation between the camera pose and the robot hand pose, the pose of the robot could be measured. However, when analysing the accuracy of the PosEye system it turned out to be too low for robot welding applications. It was identified that a new camera model could improve the accuracy. That is the reason why a new camera model has been developed. It would be advantageous if it could be improved in other aspects as well, for e.g. generality, efficiency and simplicity, and be useful for a wider range of camera types. A well-known problem for the use of camera models is that certain models are normally used for different camera types [54], like fisheye and conventional cameras. This is no longer needed with the introduction of the model presented in this thesis.

A camera model is a mathematical projection between a 3D “object space” and a 2D image. The camera calibration calculates parameters of the camera model based on several images. The initial PosEye system needed reliable start values of the parameters for the camera calibration in order to robustly converge. Therefore methods were developed to estimate these, based on the shapes of the image references.

One specific problem to be solved in the robotic welding industry is the automatic determination of a 3D path to be followed by the robot. Due to heat induced deformations from the welding process and part tolerances, this path can vary from time to time. The problem was solved using a stereo-vision robotic system. In order to develop this kind of robot vision systems off-line, robot simulation software is useful and a virtual camera is used to simulate the machine vision system.

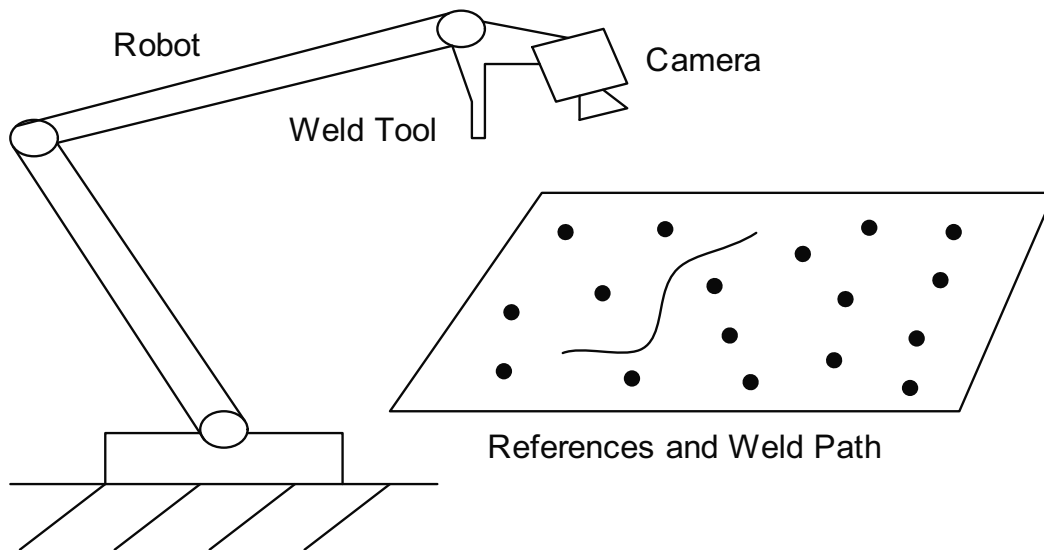


Figure 1.1: A vision system can measure the pose of a robot if the camera can see references. It can also determine the geometry of a curve to be welded by the robot. In that case it needs to first see the path from at least two directions and use stereo vision.

1.2 Objectives

The objectives of this work is to develop new and better methods for the calculation of geometric information from images to be used by robots. This includes identifying possible improvements and suggesting new solutions where cameras are used in industrial applications. Improvements that are needed mainly involve accuracy and efficiency and methods for making the systems more general, e.g. for a wider range of camera types. The whole chain of sub-systems from image processing to camera modelling and camera calibration needs to be as accurate and efficient as possible.

1.3 Research Questions

The research questions for this work are

- What problems need to be solved when cameras are used for robot positioning in welding?
- What camera types are frequently used for machine vision? How are they modelled and calibrated with required accuracy?
- How can cameras, including wide angle optics, be calibrated to achieve enough accuracy for positioning?
- How can a robot vision welding system compensate for varying geometry of weld joints?
- How can a robot vision welding system be developed in a virtual environment?

1.4 Main Contributions

To answer the research questions the focus has been on accuracy of camera models and calibration, and an emphasis has been to develop a multipurpose camera model and experimentally investigate its accuracy. The novelties are mainly:

- Introduction of a generic camera model (GCM) and its different distortion compensations. It includes conversions both from object to image space and vice versa. A major benefit of the GCM is that it includes wide angle (fisheye) cameras as well as ordinary cameras and even catadioptric cameras within the same unified model structure. Radial distortion, decentring distortion and varying entrance pupil point are modelled in novel ways. The radial distortion and varying entrance pupil compensations are more general, computationally efficient and accurate for a wide range of camera types, compared to previous state of the art. A geometric interpretation of the GCM is presented, giving a better understanding of how the GCM works. An alternative formulation of the GCM is presented, which includes both radial distortion, decentring distortion and varying entrance pupil point. Methods for including variable focus and zoom in modelling and calibration are presented.
- A thorough analysis of conventional camera models that has led to improvements making them more general, accurate and computationally efficient.
- Analysis of nontrivial null spaces, or ambiguities, that occur in the calibration, and means to overcome them. This analysis improves the calculations and guides how to robustly capture the calibration images.
- Algorithms for initial estimates of intrinsic and extrinsic camera parameters as well as reference positions for the camera calibration, based on the shapes, sizes and locations of the references in the images. This includes also methods for calculating accurate image centre points of references.
- A simple expression for stereo vision calculations has been suggested, especially suitable for the new camera model.
- Experimental investigations in which the accuracy of different camera model configurations are analysed.
- A robot stereo vision welding system developed using the new camera model. The system is implemented both in reality and in a virtual environment.

1.5 Limitations

Machine vision is a large field and only parts of it are considered in this thesis. The work only briefly discusses image processing. The camera models presented can be used for both conventional cameras, fisheye cameras and catadioptric cameras, but are not yet tested for catadioptric. The algorithms are not optimised regarding calculation time. MATLAB is used while other programmes, like C, could speed up the calculations. The programmes are not yet made robust enough for industrial implementations.

1.6 Outline of Thesis

In Chapter 2 camera modelling and calibration are described. Different types of previously existing camera models are discussed, as well as the new generic camera model, GCM. Error norms after calibration are calculated in order to compare the accuracy of different models. Chapter 3 presents the use of calibrated cameras to calculate geometric information. Related computer vision methods are then discussed in Chapter 4. Finally, Chapter 5 concludes and discusses future work followed by a summary of the appended papers in Chapter 6.

Camera Modelling and Calibration

An important part of machine vision is camera models together with camera calibration. A calibrated camera model is useful for calculating geometric information from images which in turn can guide robots. A camera model describes a mathematical projection between a 3D object space and a 2D image space. The camera calibration calculates the parameters of the camera model.

Autonomous robots in general use wide angle fisheye or omnidirectional cameras with high distortion, while traditional industrial robots use cameras with lower distortion. Thus they are both important for different applications. Lhuillier [54] claims that *“the automatic reconstruction of 3D models from image sequences is still a very active field of research. All existing methods are designed for a given camera model, and a new (and ambitious) challenge is 3D modelling with a method which is exploitable for any kind of camera”*. The generic camera model, GCM, presented in Section 2.2 and in Papers II and III, has been developed to meet this challenge. It shows that it is possible to have one camera model structure with the advantages of a fisheye camera model for a fisheye camera, and the advantages of a conventional camera model for a conventional low distortion camera. It can even model catadioptric cameras, i.e. cameras with a mirror of some shape in front of it. At the same time, it can model cameras that do not have a single viewpoint, i.e. cameras with a varying entrance pupil point, and it has methods for decentring distortion compensation.

2.1 Camera Models

General aspects of camera models are first introduced, followed by related works in camera modeling and then a presentation of the new generic camera model, GCM.

2.1.1 Introduction

The camera model consists of parameters called intrinsic and extrinsic camera parameters together with a specific algorithm. The extrinsic camera parameters are unique for each image. They are the 6D pose, i.e. the position and orientation in some coordinate system, of the camera. The intrinsic camera parameters, e.g. focal distance, principal point coordinates and distortion parameters, describe how the image is formed inside the camera, based on knowledge of the 3D object space and the extrinsic parameters. If the camera parameters and the position of an object are known, the camera model can compute where the object ends up in the image. Vice versa, if a 2D image point is known, the camera model can

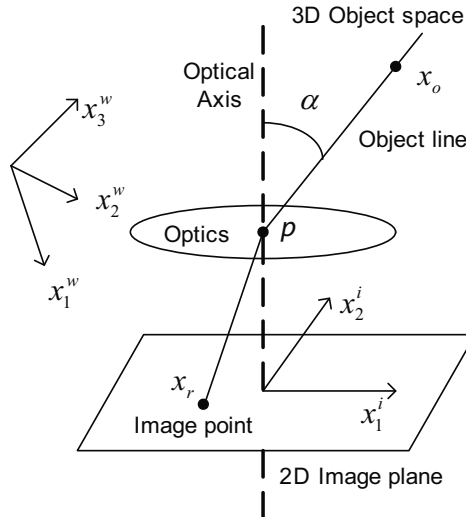


Figure 2.1: Illustration of a camera with image point, x_r , and the corresponding object line.

determine all the possible points in 3D from where the image point may have originated. These points in 3D space are called the object line. See a schematic illustration of a camera in Figure 2.1. The axis of rotational symmetry of the optics is called the optical axis. The angle between the optical axis and the object line is denoted α . The point p , the intersection of the optical axis and the object line inside the lens system is considered the position of the camera.

Upper index w indicates a world coordinate system, upper index i indicates a 2D image coordinate system. Upper index c is used for a 3D camera coordinate system with origin in p and with x, y axes parallel to the image x, y axes and z axis along the optical axis. Indices 1, 2, 3 denote x, y, z coordinates respectively.

Figure 2.2 shows natural conversions between the 3D object space and 2D image planes for different camera models. The different camera models PCM, CCM, FCM and GCM will be presented in more detail in this chapter. The different models project between the 3D surrounding and 2D image planes with or without distortion compensation. In addition to the distortions shown in Figure 2.2 varying entrance pupil point can be included in the GCM.

2.1.2 Pinhole Camera Model

The simplest kind of practical camera model is the pinhole camera model (PCM), see left part of Figure 2.3, as in e.g. Faugeras [29] and Hartley [39]. The orthographic projection is even more simple but not very realistic since it projects objects directly perpendicular to the image plane.

The PCM is such that if a line is drawn from a 3D point observed by the camera to a point p in the centre of the optics, the image point is the intersection with the image plane. This is called perspective projection. The PCM is collinear and has no distortion, i.e. a straight line in object space is mapped to a straight line in the image for the PCM. A lens system usually causes some distortion and the real image thereby differs from a pinhole image. A camera model with distortion can be used to calculate a corresponding non-distorted image, based on a distorted image, see the image of a checkerboard pattern in Figure 2.4. A common way

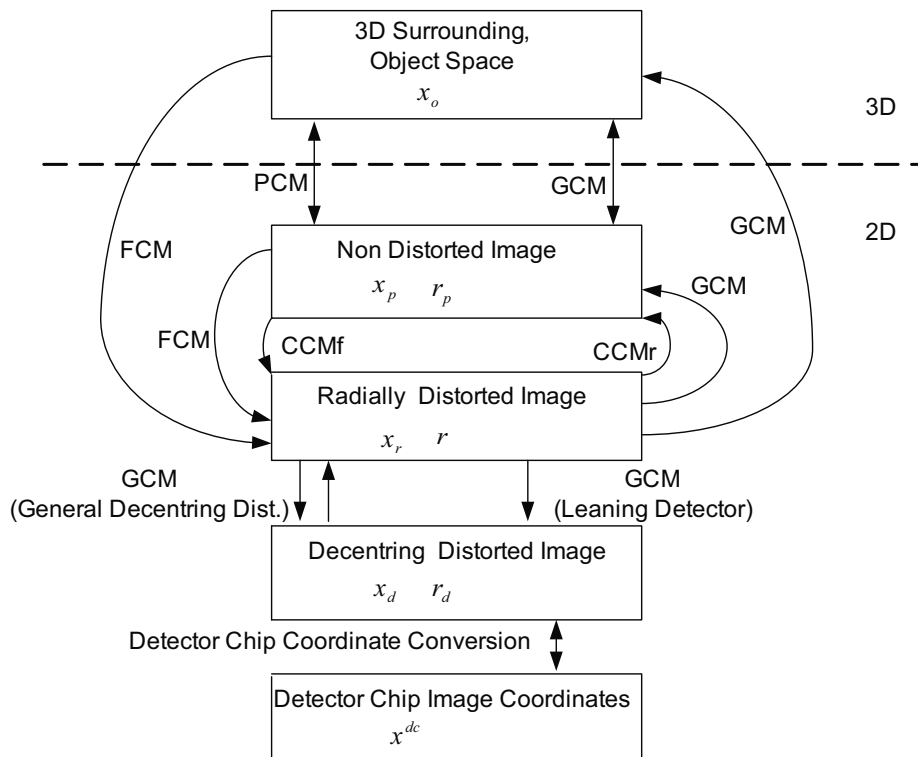


Figure 2.2: Overview of different camera model projections. Arrows show natural projections between 3D object and 2D image spaces for the camera models GCM, PCM, CCMr, CCMf and FCM. Projection in the opposite direction of an arrow requires solving a non linear equation.

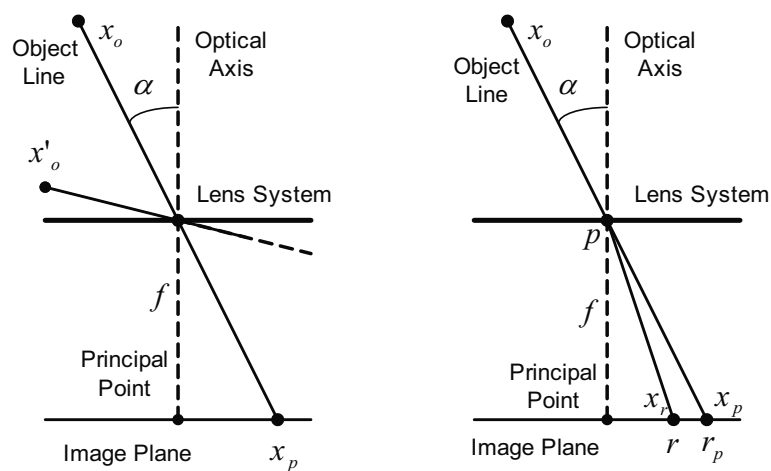


Figure 2.3: Simplified cross section of cameras. Left: PCM, Right: with radial distortion. x_o and x'_o are points in object space. x_p is the PCM image point corresponding to x_o . x_r is the radially distorted point. The image point x_p corresponding to x'_o is far away from the centre of the image.

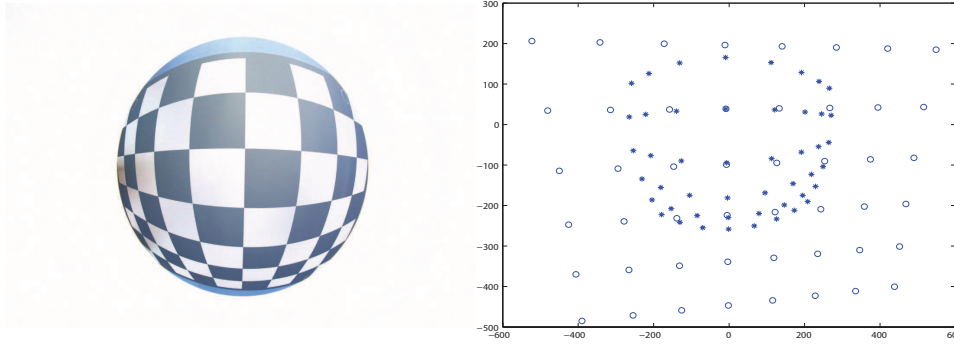


Figure 2.4: Left: Fisheye image of a straight checkerboard pattern. Right: Processed image where coordinates of checkerboard corners are obtained, shown as stars. These coordinates are then made undistorted (using the GCM in Section 2.2), shown as circles.

of modelling a camera with distortion is to use a transformation between a distorted and a non-distorted image plane, and then apply the PCM on the non-distorted image, as can be seen in Figure 2.2.

In order to get an accurate camera model, the distortions should be included in the model. The distortion types considered in this thesis are radial distortion, decentring distortion and varying entrance pupil point. A transformation is also needed between a simplified image coordinate system with the origin in the image centre and the real detector chip pixel coordinates of an image point.

For the PCM, a projection between image and object space using homogeneous coordinates is defined by,

$$\lambda \begin{bmatrix} -x_p^i \\ f \end{bmatrix} = M \begin{bmatrix} x_o^w \\ 1 \end{bmatrix} \quad (2.1)$$

where

$$M = \begin{bmatrix} R_{3 \times 3} & t \end{bmatrix} \quad (2.2)$$

transforms an extended 3D object point $[x_o^{wT} \ 1]^T$ in a world coordinate system, to a camera coordinate system x_o^c . λ is a parameter and f is the focal distance. x_p^i is the 2D image PCM coordinate in a simplified coordinate system with the origin in the image centre. The coordinates x_p^i will be generalised to pixel coordinates in Section 2.2.3. (2.1) is further motivated in Section 2.2.4, see also Heikkilä [41].

The well-developed mathematical theory of projective geometry can be used for different kinds of geometry calculations based on images, see e.g. Faugeras [29] and Hartley [39]. Projective geometry is valid only for non-distorted image calculations. Therefore, it is an advantage if the camera model has the non-distorted PCM camera model as a simple special case, and that the model can be used for transforming an image between a distorted and a non-distorted image, as is the case of the GCM.

2.1.3 Radial Distortion

A camera with only radial distortion is such that its image can be transformed into a non-distorted image by moving image points straight away or towards the image centre, the principal point, see Figure 2.3. The distance between an image point x_r and the principal point

is denoted r , and the distance between the corresponding non-distorted image point x_p to the principal point is r_p . Models using a polynomial function in r or r_p to adjust the image points are here called conventional camera models, CCM. When the function transforms from a non-distorted to a distorted image it is called a forward model, here denoted CCMf, as in (2.3), and when the function transforms from a distorted to a non-distorted image it is called a reverse model, denoted CCMr, as in (2.4).

$$\text{Forward model, CCMf: } r = f_f(r_p) = r_p + k_{p1}r_p^3 + k_{p2}r_p^5 \dots \quad (2.3)$$

$$\text{Reverse model, CCMr: } r_p = f_r(r) = r + k_{r1}r^3 + k_{r2}r^5 \dots \quad (2.4)$$

The polynomial compensation for radial distortion here contains only odd powers of r and r_p . Sometimes both odd and even powers are used as in Hartley [39]. A forward model is more suitable for transforming from a non-distorted to a distorted image, while a reverse model is more suitable for transforming from a distorted to a non-distorted image. When transforming in the opposite direction, using CCMf means that the polynomial equation in r_p (2.3) must be solved, and for CCMr (2.4) must be solved for r . Furthermore, the image coordinates for the distorted point x_r and non-distorted image point x_p are related as $x_p = x_r f_r(r)/r$ and $x_r = x_p f_f(r_p)/r_p$ expressed in a coordinate system in which the origin is the principal point.

An early paper about camera models and camera calibration was written by Brown, [13]. It describes a CCMr for distortion compensation with a polynomial as $f_r(r)$. Tsai [81] describes the same CCMr and shows more details concerning how to project between the 3D object space and the 2D image space. Heikkilä [41] and Nowakowski [65] show both the CCMr and CCMf models using odd power polynomials.

One problem with the methods (2.3) and (2.4) is that $r_p \rightarrow \infty$ for $\alpha \rightarrow 90^\circ$ for the PCM; while r is limited in reality, see Figure 2.3. This cannot be modelled by the polynomials (2.3) and (2.4). Therefore, other methods are used for these wide angle so called fisheye cameras. However the CCMf and CCMr can also handle this problem, as shown in Paper III, by using quotients between polynomials instead of just polynomials in the distortion functions. What is needed for CCMr is a function $r_p(r)$ with the qualitative behaviour of the left plot in Figure 2.5. The figure is a bit exaggerated by considering an extreme fisheye camera with a maximum viewing angle of $\alpha = 180^\circ$. Two constants r_{90} and r_{180} are defined corresponding to the values of r for the angles $\alpha = 90^\circ$ and $\alpha = 180^\circ$ respectively. The behaviour of the left plot can be achieved by using a quotient between polynomials as $r_p(r)$ with zeros in $r = 0$ and $r = r_{180}$ and a pole, i.e. a zero in the denominator, at $r = r_{90}$. This quotient makes

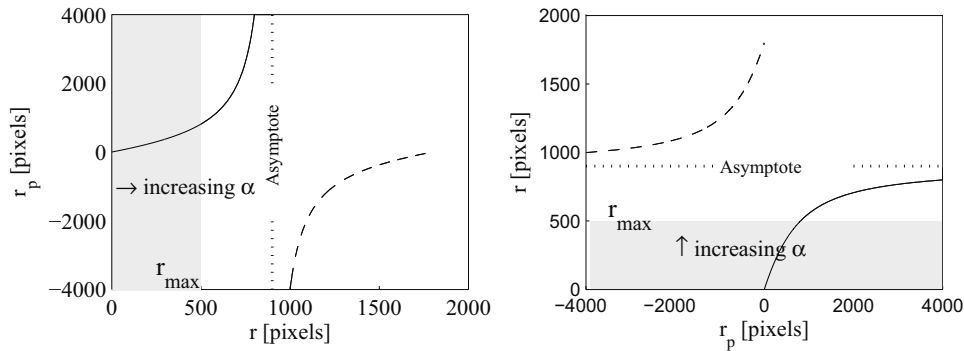


Figure 2.5: Plots of radial distortion functions. Solid for $\alpha \leq 90^\circ$ and dashed for $\alpha > 90^\circ$. Left, CCMr; Right CCMf.

the CCMr model more general, since it also models cameras with high distortion and large angles of view. This is generalised to more distortion parameters in Paper III. At the same time as the model gets more general, it becomes more computationally efficient using the quotients between polynomials. The degree of the polynomial equation from non-distorted to distorted image then gets a lower degree in relation to the number of camera parameters, compared to ordinary CCMr models (2.4). The CCMf should have a behaviour of the right plot in Figure 2.5 which can be achieved using another similar type of approach described in Paper III.

Fisheye Camera Models

A camera model specialised for fisheye cameras is here denoted a fisheye camera model, FCM. Brauer-Burchardt [12] suggests a function between r and r_p suitable for such cameras. It includes both the forward and the reverse model. An FCM, using a logarithmic function with one radial distortion parameter to transform an image from non-distorted to distorted image is described by Basu [7]. Courbon [23] also lists and compares a number of camera models suitable for fisheye cameras. One method is to use a function $r(\alpha)$ instead of a function $r_p(r)$ or $r(r_p)$. Bakstein [4] lists several such functions, which all have one radial distortion parameter. Kannala [45] models radial distortion for fisheye cameras by using a polynomial in α with only odd powers to calculate r :

$$r(\alpha) = f\alpha + k_{k1}\alpha^3 + k_{k2}\alpha^5 + k_{k3}\alpha^7 \dots \quad (2.5)$$

The degree of the polynomial can be adjusted to provide a suitable number of radial distortion parameters. This model naturally has some radial distortion included, even for one parameter, something which is not wanted for low distortion cameras. Such a camera where r is proportional to α is called an ideal fisheye camera.

In order to model a PCM using (2.5) we first observe that for the PCM $r_p = r = f \tan \alpha$. A Taylor series expansion of $\tan \alpha$ around $\alpha = 0$ then results in the relation

$$r(\alpha) = f\left(\alpha + \frac{1}{3}\alpha^3 + \frac{2}{15}\alpha^5 + \frac{17}{315}\alpha^7 \dots\right) \quad (2.6)$$

where α is measured in radians. Thus FCMs based on (2.5) using several camera parameters can be considered generic since they also work for conventional cameras. However, since they do not have the PCM as a simple special case they are unnecessarily complicated and less suitable for low distortion cameras. In a simulated environment it is an advantage if the PCM is a simple special case, as in Paper V. For cameras with a small field of view, (small α), and low distortion the model described by (2.5) is accurate also with no distortion parameter except f , but not for cameras with a wider field of view and low radial distortion.

2.1.4 Decentring Distortion

Decentring distortion is not rotationally symmetric around the optical axis, and can be modelled by another kind of transformation between a distorted and a non-distorted image plane or a transformation between an image plane with decentring distortion and an image plane with only radial distortion. Decentring distortion is caused by e.g. badly aligned lenses in the lens system, a leaning detector surface or a not constant refraction index in the lenses. A common approach for handling decentring distortion is the method in Slama [73] with a

polynomial expression including cross terms in the x - and y - directions. Paper III and Section 2.2.2 describe two methods for decentring distortion, one specialised on leaning detector compensation while the other is more general.

2.1.5 Varying Entrance Pupil Point

Another distortion type is varying entrance pupil point. The entrance pupil point is the intersection between the object line and the optical axis. Varying entrance pupil point occurs if the position of the entrance pupil point on the camera varies with the angle α . A camera with a single effective viewpoint is a camera with a constant entrance pupil point. If the entrance pupil point is not constant, it is not possible to calculate a non-distorted image based on a distorted image, even if the camera parameters are known. A model including varying entrance pupil point is presented in Gennery [31], see also the end of Section 2.2.1.

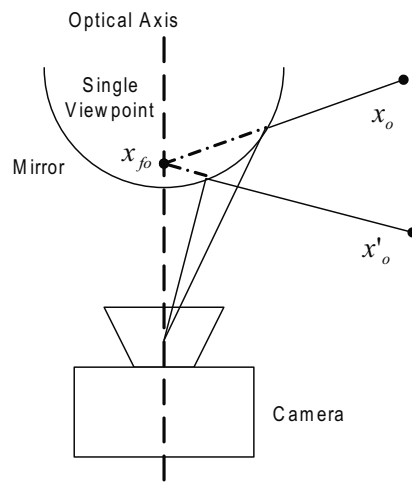


Figure 2.6: Catadioptric omnidirectional camera. A catadioptric camera gives a single viewpoint if all the object lines intersect with the optical axis in the same point, x_{fo} .

2.1.6 Catadioptric Cameras

Catadioptric omnidirectional cameras have a, usually rotationally symmetric, mirror in front of the lens system, see Figure 2.6, to increase the field of view as in e.g. Benosman [9] and Chahl [15]. Catadioptric omnidirectional cameras that provide a single viewpoint are analysed by Baker in [3] and Geyer in [32] and [33]. A single viewpoint means that the entrance pupil point is constant, so the image can be converted to non-distorted images. Another model for single viewpoint catadioptric and other highly distorted cameras is presented by Claus in [22]. That model use quotients between non rotationally symmetric second degree polynomials. Planar, conical, spherical, ellipsoidal and hyperbolic mirror shapes can give a single viewpoint, see Baker [3]. Another work on catadioptric omnidirectional and fisheye cameras with models with a single viewpoint, including how they are used for e.g. autonomous robots, is presented in Scaramuzza [70]. A catadioptric camera model for both single viewpoint and not single viewpoint cameras is presented in [53]. The GCM, presented

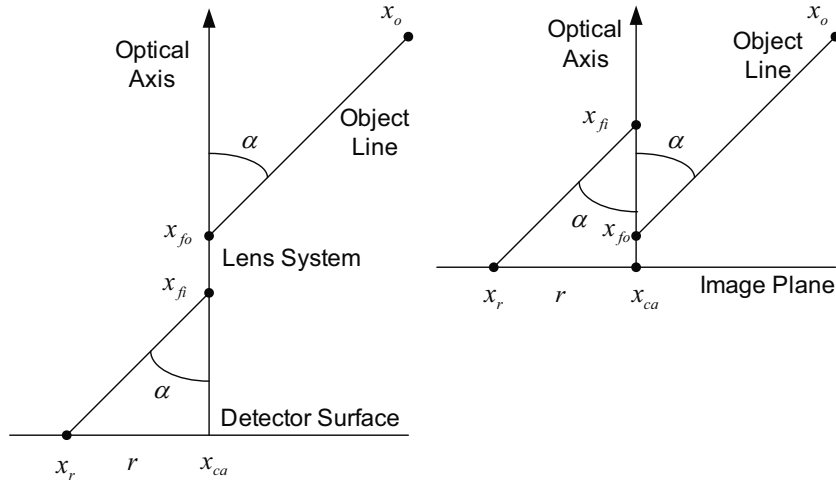


Figure 2.7: 2D illustrations of the GCM. Left, intuitive model. Right, the actual model, in which the mathematical image plane and x_{ca} are moved closer to x_{fo} . x_{ca} is a fixed point in the lens system, while x_{fo} is the entrance pupil point that can move.

in the next section, can handle rotational symmetric catadioptric cameras as described in Paper III, and it can handle cameras that have and do not have a single viewpoint.

2.2 New Generic Camera Model

A new generic camera model, GCM, is presented in Paper II and further developed in Paper III. To explain the GCM, a geometric construction of the model is described, see the left part in Figure 2.7. First an image coordinate system is used in which the origin is the principal point, x_{ca} . This will be transformed into the actual detector pixel coordinates in the end. The GCM first handles both radial distortion and varying entrance pupil point, and the decentering distortion is added afterwards.

2.2.1 Radial Distortion and Varying Entrance Pupil Point

For the GCM two points on the optical axis are defined, called x_{fi} and x_{fo} . These can slide along the optical axis, and their positions depend on r . What is wanted is a relationship between the 2D image point x_r and the corresponding object line. In the model, the object line is such that it goes through the point x_{fo} and is parallel to a line from the image point x_r and the point x_{fi} . The distance between the principal point and the point x_{fi} is $f_i(r)$, and the corresponding distance between the principal point and the point x_{fo} is $f_o(r)$. The function $f_i(r)$ determines the radial distortion, and $f_o(r)$ the entrance pupil point variations. They are similar to the radial distortion compensation functions of the CCM models in (2.3) and (2.4), and their parameters are intrinsic camera parameters.

The principal point in a 3D world coordinate system, called x_{ca}^w , is considered the position of the camera. Unit vectors e_x^w , e_y^w and e_z^w are constructed along the coordinate axes of the 3D camera coordinate system, expressed in world coordinates. e_z^w is along the optical axis and e_x^w and e_y^w are along the image x - and y - axes. If the 2D image coordinate x_r^i is known in an image coordinate system with its origin in the principal point, x_{ca} , the

coordinates in a 3D world coordinate system x_r^w can be calculated as

$$x_r^w = x_{ca}^w + x_{r1}^i e_x^w + x_{r2}^i e_y^w \quad (2.7)$$

where x_{r1}^i and x_{r2}^i are the image coordinates. The points x_{fo}^w and x_{fi}^w can be calculated according to

$$x_{fo}^w = x_{ca}^w + f_o(r) e_z^w \quad (2.8)$$

$$x_{fi}^w = x_{ca}^w + f_i(r) e_z^w \quad (2.9)$$

The following function is proposed for $f_i(r)$ and motivated in Paper III:

$$f_i(r) = \frac{f + k_{q1}r + k_{q2}r^2 \dots}{1 + k_{r1}r + k_{r2}r^2 \dots} \quad (2.10)$$

Negative $f_i(r)$ corresponds to $\alpha > 90^\circ$ shown in Figure 2.8. A first approach using polynomial functions for $f_i(r)$ was discussed in Paper II and is further developed into the polynomial quotient function (2.10) in Paper III. Other functions can also be considered. f is the focal distance as in PCM, CCM and FCM. One advantage of the GCM is that the function $f_i(r)$ does not approach infinity for $\alpha = 90^\circ$, instead $f_i(r) = 0$ for $\alpha = 90^\circ$.

A model in [70] by Scaramuzza is equivalent to a special case of the GCM. There a constraint is used, which is equivalent to

$$\left. \frac{df_i(r)}{dr} \right|_{r=0} = 0 \quad (2.11)$$

This leads to the following constraint between the parameters in (2.10):

$$k_{r1} = \frac{k_{q1}}{f} \quad (2.12)$$

The constraint is only empirically and not theoretically motivated in [70]. However, if this is a valid assumption it can guide the calibration calculation and get the same accuracy for one less camera parameter. This is evaluated in Section 2.4.

There is a relation between the models CCMr and GCM, if $f_o(r)$ is constant, derived from the similarity of triangles, cf. Figure 2.3 and Figure 2.7:

$$\frac{r_p(r)}{f} = \frac{r}{f_i(r)} \quad (2.13)$$

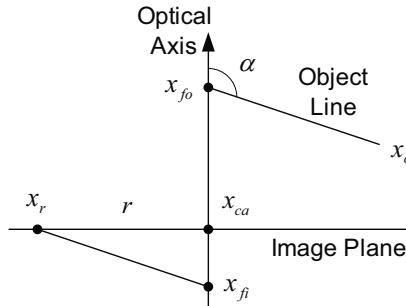


Figure 2.8: Illustration of the GCM with an angle $\alpha > 90^\circ$. $f_i(r)$ is negative so x_{fi} is below the image plane.

Using (2.13), the CCMr and the GCM can have the same behaviour, but the relation is only valid if the entrance pupil point is constant, $f_o(r) = c$. If a camera is calibrated using the CCMr, a corresponding function $f_i(r)$ can be calculated according to (2.13), resulting in the same camera model projection. The same is valid for the other direction, so that if $f_i(r)$ is known the corresponding $r_p(r)$ can be calculated. If a polynomial is used as $f_i(r)$ the equivalent CCMr is achieved using r divided by a polynomial as $r_p(r)$. A corresponding non-distorted image can be calculated with the GCM, using (2.13). This is done when undistorting the fisheye image in Figure 2.4. This is an efficient procedure, there is no need to solve an equation, only to compute the function value.

The following function is proposed, and motivated in Paper III, for $f_o(r)$, modelling the entrance pupil point:

$$f_o(r) = \frac{k_{s1}r + k_{s2}r^2 \dots}{1 + k_{t1}r + k_{t2}r^2 \dots} \quad (2.14)$$

There is an ambiguity between a general definition of $f_o(r)$ and the position of the camera. A constraint solves this problem. A suitable constraint is that $f_o(0) = 0$, which is used in (2.14). This will “move” the point x_{fo} down close to the principal point, x_{ca} , or rather the image plane, and x_{fi} will move up. This defines the position of the camera in a point in the centre of the optics, as for the CCM, see the right plot in Figure 2.7. Note that the mathematical image plane is not in the same position as the real detector. This is common also for other camera models where sometimes the image plane is in front of the optics.

From (2.7-2.14) the object line, corresponding to an image point x_r^i can be calculated according to

$$x_{fo}^w + \lambda(x_{fi}^w - x_r^w) \quad (2.15)$$

which also can be expressed as

$$x_{ca}^w + f_o(r)e_z^w + \lambda(f_i(r)e_z^w - x_{r1}^i e_x^w - x_{r2}^i e_y^w) \quad (2.16)$$

This is the object line in a parametric form. The parameter, λ , can be varied to move along the line. (2.7-2.16) represent the conversion from an image point x_r^i to the object line. The same camera model is used when there is a need to go in the opposite direction, from a 3D point observed by the camera to the point where it will end up in the 2D image. First the object point's position x_o^w is transformed to the 3D camera coordinate system x_o^c . Then the following equation is used, derived using the similarity of triangles in Figure 2.7:

$$\frac{f_i(r)}{r} = \frac{x_{o3}^c - f_o(r)}{\sqrt{x_{o1}^c{}^2 + x_{o2}^c{}^2}} \quad (2.17)$$

The equation is solved for r . Then the fact is used that the ratio between x_{r1}^i and x_{r2}^i is the same as for x_{o1}^c and x_{o2}^c , although having opposite signs. This uniquely defines the 2D point as

$$x_{rj}^i = -r \frac{x_{oj}^c}{\sqrt{x_{o1}^c{}^2 + x_{o2}^c{}^2}}, \quad j = 1, 2 \quad (2.18)$$

If $f_i(r)$ and $f_o(r)$ are polynomials or quotients between polynomials, then (2.17) for projecting to the image will be a polynomial equation. The GCM illustrated in Figure 2.7 and

the equations (2.7-2.18) can thus project both from the image to object space and vice versa. These projections are indicated in Figure 2.2. The GCM can convert efficiently in the opposite direction of an arrow in Figure 2.2 since that requires solving a polynomial equation of normally a low degree. In Paper III it is also presented how to model catadioptric cameras using the GCM.

The model by Gennery [31], developed independently of the GCM, can also include a varying entrance pupil point. It can also model both conventional cameras and fisheye cameras and has the PCM as a special case. However, it is more complicated and needs several iterations to perform a camera model projection, while the GCM can do that in a single strike. Gennery [31] offers no simple way of converting an image to a non-distorted image even if the entrance pupil point is constant.

2.2.2 Decentering Distortion

There exist a large number of decentering distortion compensation methods, e.g. in [73]. Two new methods are presented here. One is specialised in compensating for a leaning detector. The other is more general and can be used for different decentering distortion effects. An advantage of the methods presented here is that it does not matter how the image coordinate axes are oriented in relation to the irregularities of the decentering distortion. The methods determine the relations between radially distorted image points x_r and decentering distorted image points, x_d , which is indicated in Figure 2.2.

Leaning Detector Surface

To handle leaning detector surface, first the image coordinate system is temporarily rotated at an angle β around the principal point, x_{ca} , so that the new x -axis is pointing in the direction of the steepest decent of the detector. After that the leaning detector compensation formulae below (2.19), (2.20) are applied, and finally the rotation is reversed. A rotated image point in the non-leaning detector is denoted x_{rr}^i . The corresponding rotated point in the leaning detector, i.e. the decentering distorted point, is called x_{dr}^i . r is just as previously the distance to the principal point in the non-leaning image plane. The distance from the principal point to an exit pupil point is $f_l(r)$, so this point can vary with r in the model, but it can also be set to a constant value. The function $f_l(r)$ is of the same form as the parametric functions for radial distortion $f_i(r)$ in the GCM, (2.10). δ is the leaning angle of the detector. The relations between x_{rr}^i and x_{dr}^i when compensating for the leaning detector are, derived from projections to the xz plane and the yz plane

$$\frac{\cos(\arctan(\frac{x_{rr1}^i}{f_l(r)}))}{x_{dr1}^i} = \frac{\cos(\delta + \arctan(\frac{x_{rr1}^i}{f_l(r)})}{x_{rr1}^i} \quad (2.19)$$

$$\frac{\cos(\arctan(\frac{x_{rr2}^i}{f_l(r)}))}{x_{dr2}^i} = \frac{\cos(\arctan(\frac{x_{rr2}^i}{f_l(r)}) + \arctan(\frac{x_{rr1}^i \tan \delta}{x_{rr2}^i}))}{x_{rr2}^i} \quad (2.20)$$

The relations are equivalent to the geometric construction illustrated in Figure 2.9. A line is extended from the exit pupil point on the optical axis to x_{rr} in the non-leaning plane to the leaning plane. The intersection of the leaning image plane is x_{dr} . The corresponding points of x_{rr}^i and x_{dr}^i in the not rotated coordinate system are denoted x_r^i and x_d^i respectively.

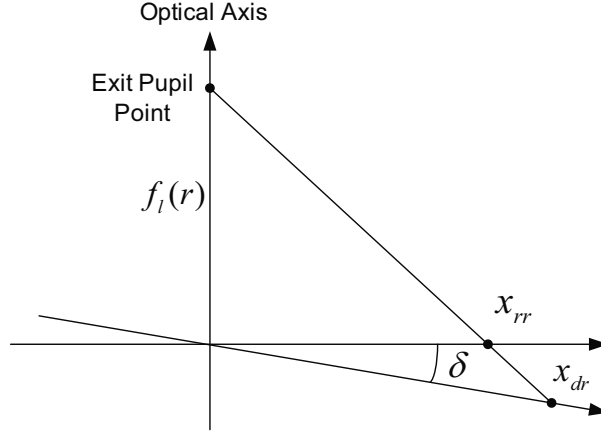


Figure 2.9: Simplified 2D illustration of the leaning detector compensation.

With these formulae it is easy to convert from the non-leaning to leaning detector coordinates, since x_{dr1}^i and x_{dr2}^i can easily be explicitly extracted, and only the values of functions need to be calculated. In the direction from x_{dr}^i to x_{rr}^i non linear equations need to be solved.

General Decentering Distortion

A general method for decentering distortion is also presented in Paper III. This is an improvement of what was given in Paper II. These calculations are performed in polar coordinates. The origin is still kept in the centre of the image. As before, index r denotes radially distorted points and index d denotes decentering distorted points. Once the polar coordinates, i.e. radius r and angle γ_r , for the points are calculated, the following formulae are applied for the non-rotationally symmetric distortion:

$$r_d(r, \gamma_r) = r + \sum_{\ell=1}^{N_r} \left(\sum_{j=1}^{n_{r\ell}} k_{g\ell j} r^{j+1} \right) \sin(\ell\gamma_r + \varphi_{q\ell}) \quad (2.21)$$

$$\gamma_d(r, \gamma_r) = \gamma_r + \sum_{\ell=1}^{N_g} \left(\sum_{j=1}^{n_{g\ell}} k_{s\ell j} r^j \right) \sin(\ell\gamma_r + \varphi_{u\ell}) \quad (2.22)$$

This is similar to Fourier expansions. The equation (2.21) provides a small radial correction and (2.22) a small angular correction for any point (r, γ_r) in the image. The expressions yield new polar coordinates that can be converted back to Cartesian coordinates. The constants k_i and φ_i are intrinsic camera parameters. N_r , N_g , $n_{r\ell}$ and $n_{g\ell}$ determine how many camera parameters are used in the compensation. This method is efficient for converting from non-decentering distorted coordinates to decentering distorted coordinates, just as for the leaning detector compensation. To efficiently convert in the opposite direction r and r_d can be exchanged together with γ_r and γ_d in (2.21) and (2.22).

An illustration of the distortion compensation is shown in Figure 2.10. It shows in an exaggerated way how the angle depending part $r_d(\gamma_r)$ in the trigonometric expression varies with γ_r for $\ell = 2$ and $\ell = 4$ in (2.21). The green dotted curve represent $\ell = 2$ and the blue dashed curve is for $\ell = 4$. Observe that using this method also the origin can be adjusted for decentering distortion if the method here is combined with the coordinate transformation that will be presented in the next subsection.

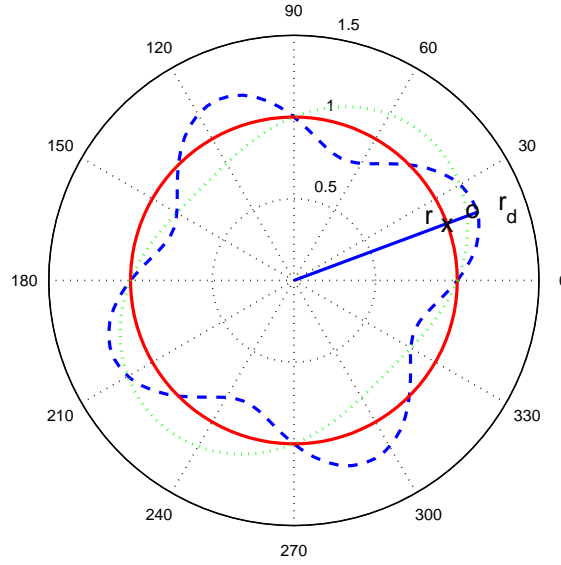


Figure 2.10: Illustration of the general decentring distortion method. Distortion curves of the angle depending part in (2.21) are shown for $\ell = 2$ (green dotted curve) and $\ell = 4$ (blue dashed curve).

A similar method was suggested by Kannala in [45], but the polar expressions in (2.21), (2.22) offer more freedom in describing the distortion than what is given in [45]. Further, the radial part is expressed in r in our model, which gives more direct calculations than in [45], where the angle α is used.

2.2.3 Detector Chip Coordinates

In the expressions so far, it has been assumed that the image coordinate system originates in the principal point. Also, the same coordinate axis units are used in the two image directions, and the image coordinate axes have been perpendicular. In a real camera however, this is usually not the case, but that problem is solved by a conversion between the real camera detector chip coordinate system and the simplified ones used above. This is needed both for the PCM, CCM, FCM and the GCM and is not a new contribution for this work. The transformation between the coordinate systems, i.e. between a possibly distorted image point x^i and a detector chip coordinate, x^d , is

$$x^{dc}(x^i) = \begin{bmatrix} \mu k & s \\ 0 & k \end{bmatrix} x^i + x_0^{dc} \triangleq Bx^i + x_0^{dc} \quad (2.23)$$

The equation (2.23) shifts the origin, with x_0^{dc} , and transforms the coordinate system, so that it coincides with the detector chip coordinate system. μ is the aspect ratio, compensating for different scaling in the image x - and y - direction, i.e. used if the pixel distances are not the same in the different image directions. If the detector image coordinate axes are not perpendicular to each other the parameter s is used, otherwise it is zero. The value of k scales the coordinate system. It can normally be set to unity. In that case it causes the focal distance, f , and the image coordinates of x_d^i , x_r^i and x_p^i to measure in units of pixel distances in the image y - direction. The coordinate transformation (2.23) is a conversion between the two lower boxes of Figure 2.2.

2.2.4 Alternative Camera Model Representation

A compact representation of the GCM will be derived. According to Figure 2.7 we have

$$\lambda(-x_r^i) = \begin{bmatrix} x_{o1}^c \\ x_{o2}^c \end{bmatrix} \quad (2.24)$$

$$\lambda f_i(r) = x_{o3}^c - f_o(r) \quad (2.25)$$

which can be expressed as

$$\begin{bmatrix} 0 \\ 0 \\ f_o(r) \end{bmatrix} + \lambda \begin{bmatrix} -x_r^i \\ f_i(r) \end{bmatrix} = M \begin{bmatrix} x_o^w \\ 1 \end{bmatrix} \quad (2.26)$$

where M is defined in (2.2), so that the right hand side of (2.26) is x_o^c . The PCM in (2.1) is a special case of (2.26) where $f_o(r) = 0$, $f_i(r) = f$ and $x_r^i = x_p^i$. The procedures for decentring distortion in Section 2.2.2 can be seen as a function $g(\cdot)$ such that

$$x_d^i = g(x_r^i) \quad (2.27)$$

By inverting (2.23) we obtain

$$x_d^i = B^{-1}x_d^{dc} - B^{-1}x_0^{dc} \quad (2.28)$$

If $g^{-1}(\cdot)$ is the inverse function of $g(\cdot)$, then

$$x_r^i = g^{-1}(x_d^i) = g^{-1}(B^{-1}x_d^{dc} - B^{-1}x_0^{dc}) \quad (2.29)$$

and (2.26) and (2.29) give a compact representation of the GCM model, where

$$\begin{bmatrix} 0 \\ 0 \\ f_o(r) \end{bmatrix} + \lambda \begin{bmatrix} -g^{-1}(B^{-1}x_d^{dc} - B^{-1}x_0^{dc}) \\ f_i(r) \end{bmatrix} = M \begin{bmatrix} x_o^w \\ 1 \end{bmatrix} \quad (2.30)$$

and

$$r = \|x_r^i\| = \|g^{-1}(B^{-1}x_d^{dc} - B^{-1}x_0^{dc})\| \quad (2.31)$$

This defines the GCM model together with $f_i(r)$ in (2.10) and $f_o(r)$ in (2.14). If there is no decentring distortion $g^{-1}(B^{-1}x_d^{dc} - B^{-1}x_0^{dc})$ is replaced by $(B^{-1}x_d^{dc} - B^{-1}x_0^{dc})$. If the projection is made to an image plane in front of the camera, which is commonly done, the minus sign in front of $g^{-1}(\cdot)$ will change to a plus sign. (2.30) is a generalisation of a model used by Micusik [61] and Scaramuzza [70] with constant entrance pupil point and no decentring distortion. Figure 2.7 together with the discussion in Section 2.2 give a geometric understanding of what the function $f_i(r)$ (as well as $f_o(r)$) implies.

In order to compute an ‘‘object line’’ using this procedure, (2.30), the vector x_o^w is re-deemed so that

$$R^{-1} \left(\begin{bmatrix} 0 \\ 0 \\ f_o(r) \end{bmatrix} - t \right) + \lambda R^{-1} \begin{bmatrix} -g^{-1}(B^{-1}x_d^{dc} - B^{-1}x_0^{dc}) \\ f_i(r) \end{bmatrix} \equiv u^w + \lambda v^w = x_o^w \quad (2.32)$$

Where R and t are the rotation and translation inside matrix M . $u = x_{fo}$ is a point on the object line, and $v = x_{fi} - x_r$ is the direction of the line. $\lambda > 0$ can be varied to move along the line. In order to project to the image, again equation (2.17) for r is used. Then x_r^i and λ are calculated using (2.26). Finally (2.27) in (2.23) gives the image x_d^{dc} .

2.2.5 Variable Focus and Zoom

Two different methods of including variable focus and zoom in the model as well as in the calibration are presented in Paper II.

2.3 Camera Calibration

The camera parameters are calculated in the camera calibration procedure, which is performed regardless of the model used. These calculations are usually based on images of reference points. By formulating the calibration as an optimisation problem the calibration can be solved by standard optimisation procedures.

Early developments of camera calibration is described in Clarke [21]. The 3D reference points' positions can be either known or not known in advance in the calibration. A calibration procedure in which the 3D reference positions are not known a priori is called self-calibration or auto-calibration. A procedure that simultaneously calculates the 3D coordinate positions of the reference points and the camera parameters from several images of these references is called bundle adjustment. A natural way of solving the bundle adjustment problem is to minimize a sum of squared errors. The 3D reference points can be projected to the image using the camera model. The sum of squared differences between the calculated and detected image points can then be used as the error criterion, see (2.33). This can be minimised with respect to intrinsic and extrinsic camera parameters and the 3D reference positions. Another method is using sums of squared errors in object space. In that case, the detected image coordinates of the references are projected to object space using the camera model, and the distance between the object line and the 3D positions is squared and summed as in (2.34).

$$\arg \min \sum_j \sum_k |x_{dj k}^{dc} - \hat{x}_{dj k}^{dc}|^2 \quad (2.33)$$

$$\arg \min \sum_j \sum_k \left(\frac{|(\hat{x}_{fijk}^w - \hat{x}_{rjk}^w) \times (\hat{x}_{fojk}^w - \hat{x}_{ok}^w)|}{|\hat{x}_{fijk}^w - \hat{x}_{rjk}^w|} \right)^2 \quad (2.34)$$

These expressions are the same whether the references x_o are known or not. j and k are numberings of images and references respectively. If the references x_o are unknown seven parameters must be given to the system in order to lock a world coordinate system. E.g. two of the references can be given their 3D positions and the z- coordinate of a third can be set to zero. If this is done the calibration calculations will not “drift away” and a unique solution can be obtained. The expression summed in (2.34) is a formula for the shortest distance between an object line and the corresponding 3D reference point x_o . \times denotes the cross product. Notations from the first formulation of the GCM as presented in the first parts of Section 2.2 are used in (2.34). Similar calculations can be performed for other models. The minimisation criteria (2.33) and (2.34) are minimisations in the first and last box in Figure 2.2. In fact minimisation can be performed in any of the stages in Figure 2.2 by projecting the 3D object space references, starting from the top, down to some other box, and by transforming the detected 2D pixel coordinates from the lowest box up to the same box, and there calculate their differences to square.

One problem in the calibration calculations is that the image references need to be recognised, in order for it to be known what point in object space corresponds to what point in image space, known as the correspondence problem. It can be solved by grouping references in unique patterns, which can be recognised using a pattern matching algorithm. The scale invariant feature transform, (SIFT), or speed up robust features, (SURF), algorithms described in [57] and [8] can be used; these are briefly explained also in Section 4.2.6. Another option is to use a regular pattern like a checkerboard to recognise which reference is which.

Many optimisation methods use a residual vector and a Jacobian matrix. The residual vector is a vector containing all the individual errors inside the squared expression in (2.33) or (2.34). The Jacobian matrix contains the derivatives of all the components in the residual vector with respect to all the unknown variables. The Jacobian can be calculated numerically by calculating how much the components in the residual vector changes when changing the unknown variables slightly. The Jacobian is a sparse matrix and calculations can be made faster by only calculating the components which are not always zero in every iteration step of the optimiser.

A calibration method that does not calculate the reference positions, but only uses the information of what image points correspond to the same 3D points in several images, can be considered a true self- or auto-calibration method. In the true auto-calibration methods the theory of projective geometry is useful as described in e.g. Faugeras [29] and Hartley [39]. In [39] abstract mathematical concepts are defined, e.g. the absolute conic, the dual absolute conic, the image of the absolute conic and the plane at infinity. They are used for camera model calculations. Calculating the image of the absolute conic is equivalent to calibrating a camera, since the intrinsic camera parameters can be calculated from the image of the absolute conic, see Hartley [39]. One true self-calibration method in [39] uses Kruppa's equations and the dual of the image of the absolute conic. Another self-calibration method in [39] uses modulus constraints. The modulus constraint method calculates the plane at infinity. The modulus constraint and the Kruppa equations methods are related to each other since if the dual of the image of the absolute conic is known the plane at infinity can be calculated and vice versa. One disadvantage of the Kruppa equations method is that ambiguities can occur, and the equations are difficult to solve. The bundle adjustment method (2.33) or (2.34) does, however, not contain that ambiguity. Another advantage of methods that calculate the 3D reference points at the same time as the camera parameters, like the bundle adjustment, is that these reference positions can be used for other purposes, e.g. for camera pose calculations or if a robot needs to reach these points.

A calibration method using information of measurements of the positions or poses of the camera is called active calibration, see Wei [82]. In [82], not only the camera parameters and the reference positions but also a hand-eye relationship is determined in the calibration. The hand-eye relationship is the relative position and orientation between a point in the lens system and a measured point on the hand. One camera calibration procedure implementation for MATLAB, developed by Bouguet, can be downloaded from [11]. It calibrates cameras by analysing images of checkerboard patterns. This program considers the references known a priori, and hence it is not a self-calibration method. Another camera calibration toolbox in MATLAB, for omnidirectional cameras, is [71] developed by Scaramuzza. A calibration method that uses the fact that straight lines in object space are mapped to straight lines in image space for non-distorted images is presented in Devernay [24].

Camera calibration for radially symmetric distortion, for ordinary dioptric (i.e cameras

with only refraction lenses and no mirrors) and for catadioptric omnidirectional cameras are described by Tardif in [79]. Another method for calibration of central catadioptric cameras, i.e. catadioptric cameras with a single viewpoint, is described by Dunne in [28].

2.3.1 Pre-Processing Algorithms

Since the bundle adjustment calibration includes a large number of unknown parameters to be optimised, the calibration needs reliable approximate start data to converge. This thesis proposes useful pre-processing algorithms, presented in detail in Paper II. These are based on images of flat references, and are most efficient for circular references. By analysing the shape, size and positions of the references in the images, the intrinsic and extrinsic camera parameters as well as the 3D reference positions are estimated.

Better 2D image reference coordinates are also calculated, since the centres of gravity do not exactly correspond to the centre of the 3D references. First 2D image coordinates around the edge of the 2D image references are extracted. An optimisation procedure tries to find 3D reference positions together with camera parameters that match the 2D points corresponding object lines. When a good match is found the corresponding parameters are used as start data. The centres of the 3D references are projected back to the image using the camera model, providing the image coordinates. These procedures need both the forward and reverse camera model projections. Methods for obtaining starting values for the pre-processing algorithms are also presented in Paper II.

2.3.2 Non-Trivial Null Spaces

Non-trivial null spaces are ambiguities that occur e.g. if the calibration criterion (2.33) or (2.34) does not give unique parameter combinations, i.e. if several different combinations of parameters give equally low error residuals. Paper II identifies such ambiguities and presents methods of how to solve them. To solve the problems of ambiguities, constraints between the different parameters are needed, or the calibration images need to be captured from different angles or positions.

One example of a non-trivial null space is a possible constant offset in $f_o(r)$ of (2.14) for the entrance pupil point of the GCM. If a constant offset is added to the expression its value could not be calculated in a calibration based on only images, because of an ambiguity between the position of the camera and $f_o(r)$. The constraint $f_o(0) = 0$ is chosen for simplicity.

In the calibration a relation between the focal distance and the pixel distances can be determined, so that e.g. it is possible to calculate the focal distance in pixel units. However it is not possible to calculate the focal distance or pixel distances in e.g. m or mm based on only images. This is because if the focal distance is increased and the pixel distance is increased with the same proportion the same image coordinates will be obtained. In practice however this is not a big concern since the camera model transformation between the image space and object space can still be determined.

Another example of a non-trivial null space considers the definition of the world coordinate system. For a fixed camera position relative to the workspace the same images appear independent of the world coordinate system. That is why some extra information is needed, e.g. some of the reference points' positions could be given before the bundle adjustment calibration calculation (2.33) or (2.34), which will give a fixed world coordinate system, see

details in Paper II.

Hartley et al. in [38] give a deep mathematical analysis of when there is a unique projective reconstruction of the 3D geometry of the scene points and the camera positions. They consider two, three and more views of the scene. Guilbert et al. in [35] discuss ambiguities between intrinsic and extrinsic parameters in uncalibrated vision by analysing the Jacobian of reprojection errors. Åström et al. in [2] have investigated and pointed out ambiguities occurring for one dimensional images.

2.3.3 Total Calibration Process

A block diagram of the calibration process is presented in Figure 2.11. First image processing is needed in order to find image coordinates of the reference points. These reference points need to be recognised to solve the correspondence problem. Starting values of parameters can be achieved e.g. using the pre-processing algorithms. A calculation program is needed that calculates the residual vector as well as the Jacobian matrix, based on approximate camera parameters and reference positions together with as accurate 2D image coordinates as possible. Once the criterion is minimised the camera parameters are known, as well as the reference positions.

2.4 Experimental Verification of Camera Models

In order to verify and compare the accuracy of camera models they have been applied to real cameras. In Paper III error norms calculated after camera calibrations were compared, using different camera model configurations on two different camera types, one conventional lens and one fisheye lens.

2.4.1 Error Norms

The intrinsic camera parameters are now collected in a vector denoted θ . The error norms for comparing the models are (expressed for the GCM notation)

$$\bar{\varepsilon}^i(\theta) = \frac{1}{mn} \sum_j^m \sum_k^n |x_{djk}^{dc} - \hat{x}_{djk}^{dc}| \quad (2.35)$$

$$\bar{\varepsilon}^w(\theta) = \frac{1}{mn} \sum_j^m \sum_k^n \frac{|(\hat{x}_{fijk}^w - \hat{x}_{rjk}^w) \times (\hat{x}_{fojk}^w - \hat{x}_{ok}^w)|}{|\hat{x}_{fijk}^w - \hat{x}_{rjk}^w|} \quad (2.36)$$

$$\max(\varepsilon^w(\theta)) = \max_{jk} \frac{|(\hat{x}_{fijk}^w - \hat{x}_{rjk}^w) \times (\hat{x}_{fojk}^w - \hat{x}_{ok}^w)|}{|\hat{x}_{fijk}^w - \hat{x}_{rjk}^w|} \quad (2.37)$$

where j and k numerate images and references respectively. $\bar{\varepsilon}^i$ is the average distance between measured and calculated image points. $\bar{\varepsilon}^w$ is an object space error norm, where all the image points are projected out to object space using the camera model. $\max(\varepsilon^w)$ is the largest distance between the object line and the 3D reference points for all the references in all the images. Note that the expressions in the norms (2.35-2.37) are not squared as they are in the calibration optimisation expressions (2.33-2.34), to give a more natural understanding of the size of the errors.

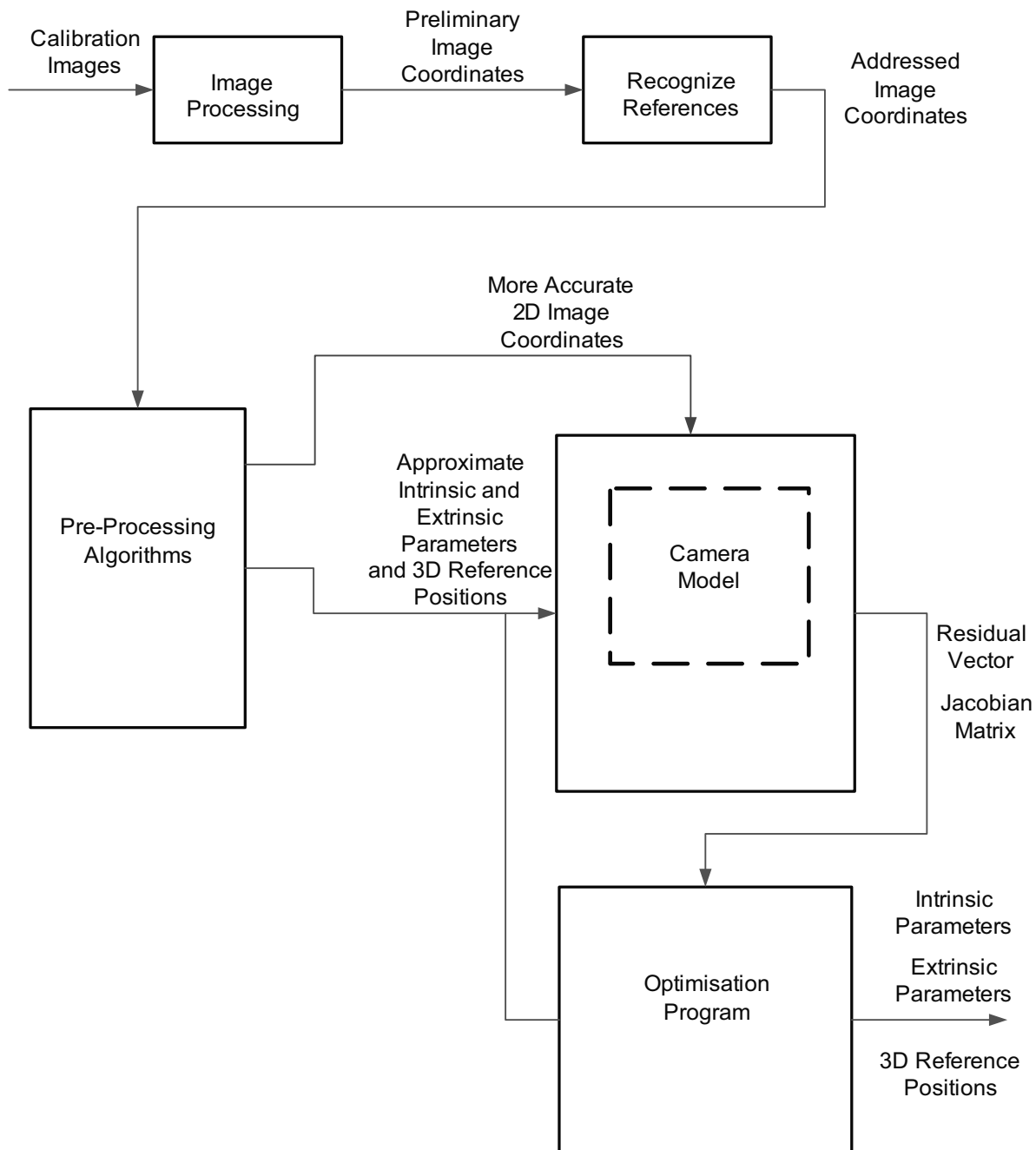


Figure 2.11: Block diagram of proposed camera calibration process.

2.4.2 Experimental Setup for Comparison

A checkerboard pattern was used as reference for the calibration. The camera Calibration Toolbox in MATLAB [11] performed the image processing and calculated preliminary data for the refined calibration. All the reference points on the checkerboard were seen in all images. After the toolbox calculations, an image space bundle adjustment calibration (2.33) calculates both the intrinsic and extrinsic camera parameters together with the 3D reference positions. The calibration optimisation was performed with the Knitro solver by TOMLAB running under MATLAB using an error residual vector and its Jacobian matrix. The Jacobian matrix was calculated numerically in each iteration.

The camera used is a Canon EOS 5D, with two different lenses, one fisheye lens, a Sigma 8mm 1:3.5 ExDG Fisheye, and one conventional zoom lens, Canon Zoom EF 24-105 1:4 IS USM.

The conventional zoom lens was maximally zoomed out giving a maximum viewing angle $\alpha_{\max} \approx 35^\circ$. The number of pixels in the images were reduced from 21 megapixels to about five megapixels before the calculations, since the calibration toolbox handles limited sized images. 38 calibration images were taken of a 8x10 checkerboard reference pattern. The average distance to the pattern was about 70 cm.

For the fisheye lens the number of pixels was reduced from 21 megapixels to about 2.5 megapixels. 23 calibration images were taken of a 6x8 reference pattern. The distance to the pattern was between 10 and 30 cm.

2.4.3 Model Comparison

Table 2.1 shows the different camera models implemented and compared and what they are called in the text and in the result Tables 2.2 and 2.3. Odd powers of the CCM and FCM correspond to even powers for the GCM. The number of intrinsic camera parameters and the degree of the corresponding polynomial equation are measures of the complexity of the models. The number of intrinsic camera parameters are shown in the tables as “number of θ ”. The “degree n_p ” is the degree of the polynomial equation for converting in the opposite direction of the arrows in Figure 2.2. Two of the intrinsic parameters are the coordinates of the principal point, one is the aspect ratio and one is the focal distance. The rest are radial distortion parameters. Lening detector compensation and varying entrance pupil point are discussed later. In the tables the object space errors $\bar{\varepsilon}^w(\theta)$ are shown. This is since the models are mainly used to calculate geometric information, and the image errors were almost proportional to the object space errors for a given camera. Also $\max(\bar{\varepsilon}^w(\theta))$ is shown which gives a measure of the deviations from the average errors.

Table 2.2 shows results for the conventional lens. Comparing models using 7 intrinsic camera parameters *CCMr, odd powers*, *CCMf, odd powers*, *GCM, even powers*, *GCM, ev pow quotient* and *GCM, constrained* were most accurate. They show similar results. The GCM models achieve the low errors with a lower degree of the corresponding polynomial equations, especially *GCM, constrained* with a 3rd degree polynomial. Comparing camera models using the same degree of the polynomial equations it is clear that the GCM models, especially models using a quotient, have the best accuracy. Another advantage with the GCM models is that they can be used also for omnidirectional cameras. For the best models an object space error norm $\bar{\varepsilon}^w$ of 0.003 mm corresponded to an image space error norm of about $\bar{\varepsilon}^i = 0.02$ pixels for the calibration images of the conventional lens.

Table 2.1: Notations used for radial distortion models compared

Model name	Definition
CCMr, regular	Both odd and even powers in $r_p(r)$ polynomial
CCMr, odd powers	Only odd powers in $r_p(r)$ polynomial
CCMf, regular	Both odd and even powers in $r(r_p)$ polynomial
CCMf, odd powers	Only odd powers in $r(r_p)$ polynomial
FCM, regular	Both odd and even powers in $r(\alpha)$ polynomial
FCM, odd powers	Only odd powers in $r(\alpha)$ polynomial
GCM, regular	Both odd and even powers polynomial in $f_i(r)$
GCM, quotient	Polynomial divided by first degree polynomial in $f_i(r)$
GCM, constrained	Polynomial divided by first degree polynomial and constraint (2.12)
GCM, even powers	Polynomial with only even powers in $f_i(r)$
GCM, ev pow quotient	Polynomial with only even powers divided by $1 + cr^2$

Table 2.3 compares accuracy for the GCM and FCM models for the fisheye lens. Only one CCM model is shown since they can not model large angle cameras. Comparing models using 7 intrinsic camera parameters *GCM, constrained*, *FCM, odd powers* and *GCM, ev pow quotient* had the lowest errors, and the *GCM, constrained* has a low degree of the polynomial equation, degree 3 instead of 7 for *FCM, odd powers*. Again, comparing models with the same degree of the polynomial equations the GCM models has very good results, especially quotient models with both odd and even powers of the polynomials.

The best over all performance turned out to be achieved with the *GCM, constrained* and *GCM, ev pow quotient* models, achieving a high accuracy for both of the lenses and yielding a low degree of the polynomial equation especially for *GCM, constrained*. An even lower degree of the polynomial equation is achieved if denominator has a higher degree than one in $f_i(r)$. Lowest degree of the equation (in relation to the number of intrinsic camera parameters) is achieved when the degree of the numerator is one larger than for the denominator and both odd and even powers are used.

When using varying entrance pupil point only less than 1% was gained in the accuracy performance for the conventional camera and nothing was gained for the fisheye camera. This is explained by Micusik [62] arguing that in recent years fisheye lenses have been built to satisfy the single viewpoint property, which means that the entrance pupil point is constant.

Including leaning detector compensation (2.19) and (2.20) increased the accuracy. For GCM using a quotient in $f_i(r)$ with five radial distortion parameters, using the conventional zoom lens, the image error decreased from 0.017 to 0.013 pixels for a constant exit pupil point, $f_i(r)$. When also varying exit pupil point was included the image error was only 0.0096 pixels. So the error decreased by 44%. The leaning angle δ of the detector was calculated to 0.044° . Decentering distortion is usually larger for low quality cameras, so since leaning detector compensation was useful for a professional camera it should be even more significant for cheap low quality cameras.

The results in this section can be compared to error results in Paper I and Paper II. There camera poses were calculated based on calibrated cameras (see Section 3.1). The vision pose output was compared with camera positions measured by a coordinate measurement machine. In Paper I only a conventional camera model was used. The accuracy was not high

Table 2.2: Comparison of the errors for conventional lens, only radial distortion.

Model	number of θ	degree n_p	$\bar{\varepsilon}^w$ [mm]	$\max(\varepsilon^w)$ [mm]
CCMr, regular	7	4	0.005	0.10
CCMr, odd powers	6	5	0.005	0.09
CCMr, odd powers	7	7	0.0026	0.018
CCMf, regular	7	4	0.004	0.06
CCMf, odd powers	6	5	0.043	0.67
CCMf, odd powers	7	7	0.0025	0.019
FCM, regular	7	4	0.032	0.62
FCM, odd powers	7	7	0.012	0.37
GCM, regular	7	3	0.0052	0.091
GCM, regular	8	4	0.0025	0.019
GCM, quotient	7	2	0.012	0.28
GCM, quotient	8	3	0.0025	0.019
GCM, constrained	7	3	0.0028	0.019
GCM, even powers	7	6	0.0025	0.019
GCM, ev pow quotient	7	4	0.0025	0.019

enough for welding applications, which is the reason that the GCM was developed. The main aim of it was to increase accuracy by including more types of distortion. Another aim was to make it more general, so that it could model a wider range of camera types. The data from the measurements with the camera on the coordinate measurement machine was then used to determine the accuracy of the GCM using a polynomial as $f_i(r)$, in Paper II. The error results turned out to be much larger than the results shown in this section and in Paper III. The main reason for this is the differences in image processing. In Paper I and II the references were infrared light emitting diodes, LEDs. The image processing found the centrepoints of these in the image in a very simple way, while the image processing from the camera calibration toolbox used in this section and Paper III found crossings in the checkerboard patterns in a very accurate sub pixel level.

Table 2.3: Comparison of the errors after calibration for fisheye lens, only radial distortion.

Model	number of θ	degree n_p	$\bar{\varepsilon}^w$ [mm]	$\max(\varepsilon^w)$ [mm]
CCMr, regular	7	4	0.390	2.23
FCM, regular	7	4	0.024	0.28
FCM, odd powers	7	7	0.016	0.17
GCM, regular	7	3	0.099	0.83
GCM, quotient	7	2	0.054	0.56
GCM, quotient	8	3	0.0067	0.075
GCM, constrained	7	3	0.015	0.17
GCM, constrained	8	4	0.0010	0.0087
GCM, even powers	7	6	0.024	0.28
GCM, ev pow quotient	7	4	0.018	0.19

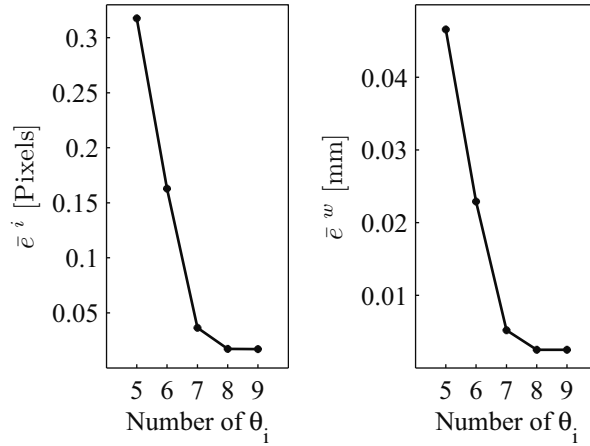


Figure 2.12: Error decreasing with increasing number of intrinsic camera parameters for the zoom lens for *GCM, regular*. Left plot shows image error $\bar{\epsilon}^i$ in pixel units, right plot shows object space error $\bar{\epsilon}^w$ in mm.

Influence of Number of Parameters

Figures 2.12 and 2.13 show how the error decreases with the number of intrinsic camera parameters for certain GCM models. Figure 2.12 considers the conventional lens system with a polynomial as $f_i(r)$ in (2.10). The left plot shows $\bar{\epsilon}^i$ (an image error in pixels), the right plot shows $\bar{\epsilon}^w$ (an object space error in mm). By increasing the number of parameters to more than eight the accuracy does not increase much for this camera. Figure 2.13 shows data for the fisheye lens where $\bar{\epsilon}^i$ decreases using *GCM, quotient* and *GCM, regular*. The accuracy for nine and ten parameters are very high for the *GCM, quotient* model.

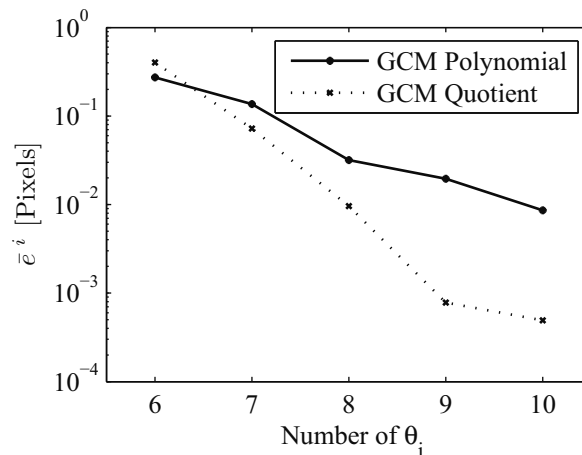


Figure 2.13: Image error $\bar{\epsilon}^i$ decreasing with increasing number of intrinsic camera parameters for the fisheye camera and GCM-models. Note the logarithmic y-axis.

2.4.4 Discussion

The GCM has been presented and compared to other radial distortion models. The method for including radial distortion makes it general. A straightforward analysis of the asymptotic

behaviour of the models at $\alpha = 90^\circ$ and $\alpha = 180^\circ$ in Paper III leads to the proposal of using quotients between polynomials in the radial distortion functions, which improves both the GCM and the CCM models. Varying entrance pupil point is included in the model in an efficient way and new methods for decentring distortion are proposed. There are a number of desirable properties of a camera model that should be considered, e.g. it should be general, accurate, fast, simple and natural. The models are analysed with respect to these criteria.

General and Natural

Usually different camera models are used for wide angle fisheye and conventional cameras. The reason is that in (2.3) and (2.4) r_p can not approach infinity for limited r , so special FCMs are used for these situations. These FCM's are usually not used for conventional cameras since they do not have the PCM as a simple special case. This property is important for the well developed mathematical theory, called projective geometry, which is applied to non distorted images, see the books by Faugeras [8] and especially Hartley [1]. Also the model should be able to convert between distorted and non distorted image for this reason. The conventional FCM (2.5) with degree one forces a radial distortion, which is not suitable for low distortion cameras.

The conventional models are made more general by the proposed change of the function dependency between r_p and r in Paper III. Letting r asymptotically approach a constant value when $r_p \rightarrow \infty$, and allowing for negative r_p accomplishes this. Therefore using this adjusted conventional camera model, or using the GCM, there is no longer a need to use different models for different cameras. The GCM can even model catadioptric cameras. It models these cameras in a more natural way, since the function of r in the GCM does not need to approach infinity for angles $\alpha = 90^\circ$ as is the case of the CCM. The GCM can efficiently convert a radially distorted image to a non distorted image, even for catadioptric cameras, if $f_o(r)$ is constant, i.e. if it has a single viewpoint, see Paper III. Further, the GCM also handles varying entrance pupil point as well as decentring distortion.

Accuracy

Comparing camera models with the same number of radial distortion parameters used for the two lenses the so called *GCM, constrained* and *GCM, ev pow quotient* models had the best accuracy. At the same time, *GCM, constrained* had a low degree of the corresponding polynomial equation.

The influence on the accuracy for varying entrance pupil point was negligible for the camera tested, while leaning detector compensation increased the accuracy of the model significantly. Compensation for decentring distortion is of even higher importance for simple and cheap cameras.

Simple and Fast

Most models can project easier either from image to object space or vice versa. Projecting in the opposite direction means that a polynomial equation needs to be solved. The improved models have a lower n_p , i.e. degree of polynomial equation, in relation to the number of camera parameters θ used, making the transformation in the opposite direction easier and faster. This applies both to the GCM with a quotient and both odd and even power, and the improved CCMf and CCMr in Paper III.

If varying entrance pupil point is needed, the GCM includes this using the $f_o(r)$ function, and accomplishes this in a simple and efficient way. [9] also has varying entrance pupil point, but needs several iterations in the calculations, while the GCM achieves this in a single strike.

Geometric Calculations from Camera Models

Once a camera model is calibrated for a camera, the model can be used for calculating geometric information from the images. Two examples of geometric calculations from images in this thesis are camera pose calculation and stereo vision, which are presented below. The pose calculation is similar to the camera calibration calculation. It assumes that references with known 3D positions are available and that these can be detected in the images e.g. using image processing techniques. The stereo vision method instead calculates positions of objects seen in images. Here at least two camera views are needed in order to determine depth. The 2D image coordinates of the points to be determined are needed in the calculations as well as the camera poses for the camera views and the intrinsic camera parameters. To find the 2D image coordinates image processing techniques are needed also in the stereo vision calculations. In the paper [42], by Heyden et al, it is shown how a 3D euclidian space can be reconstructed for a PCM given sufficiently many point correspondences in a sufficient number of images, even if the focal distance and the principal point are not known, and even varying. Both stereo vision and camera pose estimation are used for a visually guided mobile platform in [10] by Björkman et al.

3.1 Pose Calculation

If a camera model is calibrated for a camera and the 3D references are known and seen in the image, the pose of the camera can be calculated from a single image, see e.g. Dornaika [25], Lu [58]. The pose calculation is similar to the calibration problem, but it is less complicated. Just as the calibration, the pose calculation can be formulated as an optimisation problem. The same optimisation criteria as for the calibration can be used, except that the sum of images is omitted and the reference coordinates and intrinsic camera parameters need to be known, so the only unknown parameters are the 6D pose of the camera. As in the calibration the correspondance problem needs to be solved, so that it is known which reference in the images corresponds to which 3D object space reference. An example of a vision based pose measurement system mounted on a coordinate measurement machine is presented in Figure 3.1. There the correspondance problem was solved using unique groupings of IR-LEDs in combination with a pattern matching algorithm. In order to obtain a unique solution, normally at least four references need to be seen in the image. If three references are used the number of unknown parameters are the same as the number of equations, but it yields

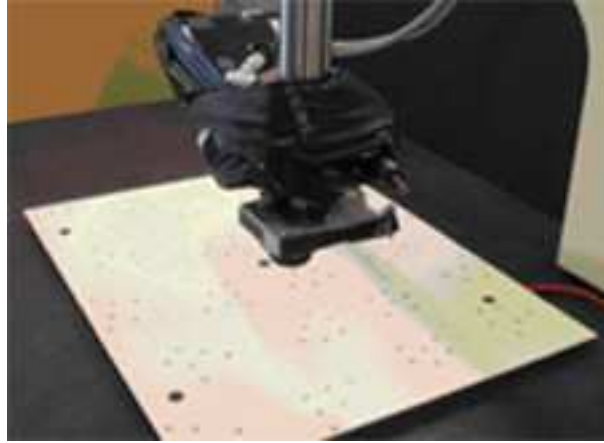


Figure 3.1: A camera sensor mounted on a coordinate measurement machine arm. The camera is looking downwards at the reference plate, with unique groups of 6 IR-LEDs.

several solutions. If one more reference is used normally only one solution is obtained. The accuracy of a pose measurement system using a conventional camera model is presented in Paper I and the usage of the GCM is presented in Paper II. If a camera is mounted on the hand of a robot this can be used for measuring the robot pose. This requires a hand-eye transformation that can be determined in a hand-eye calibration procedure, see Remy [68] and Lenz [52].

3.2 Stereo Vision

Stereo vision methods calculate positions of points seen from at least two directions. One direction is not sufficient, since then depth cannot be determined. Two basic methods exist, one minimising object space errors, see Paper II, and another minimising image space errors, see e.g. Hartley [39]. The image minimisation method is more often used, since the errors usually originate from the image capturing. In the image space minimisation method in [39], a 6th degree polynomial equation is used to calculate the 3D point, but the method does not include varying entrance pupil point camera models. The object space minimising procedure easily includes varying entrance pupil point cameras, or non-single viewpoint cameras by using the GCM. An analysis of different stereo vision algorithms can be found in [72]. The object space calculation stereo vision method suitable for the GCM is described in Paper II. For determining 3D points first calculate object lines based on the image points and the calibrated GCM according to

$$\text{Line 1: } x_{0a} + av_a; \quad \text{Line 2: } x_{0b} + bv_b \quad (3.1)$$

In perfect conditions with no errors these lines should cross each other. In practice this will however not be the case. But the closest points between them can be calculated if they do not exactly intersect. Solve the following system of equations for a and b yielding the closest points between the two lines.

$$\begin{bmatrix} v_a^T v_a & -v_a^T v_b \\ v_a^T v_b & -v_b^T v_b \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -v_a^T x_{0b} + v_a^T x_{0a} \\ -v_b^T x_{0b} + v_b^T x_{0a} \end{bmatrix} \quad (3.2)$$

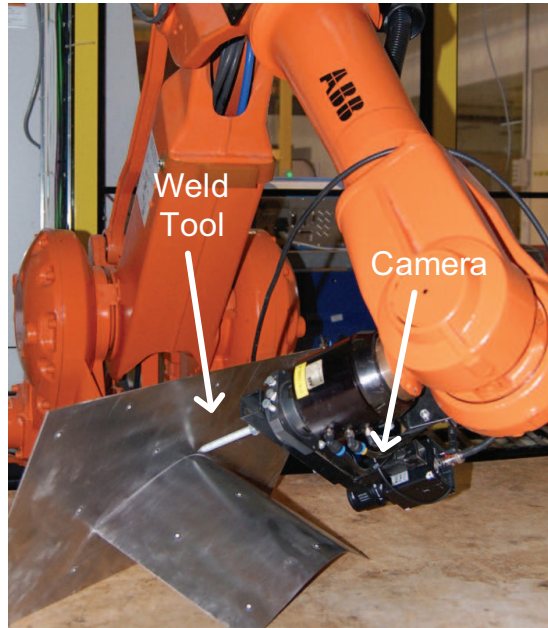


Figure 3.2: A robot following a 3D weld joint. The camera is mounted on the robot hand.

This system of equations is derived using the fact that a line between two points representing the shortest distance between the two lines, Line 1 and Line 2, is perpendicular to both of the two lines. If the solution a and b of (3.2) is inserted into the expressions (3.1) the points closest to each other are achieved. E.g. the centrepoints between these can be used as the 3D point searched for. When calculating 3D positions using vision the point needs to be detected in both images. If the point is found in one image, its corresponding object line can be projected to the other image. The resulting line in the other image is called the epipolar line. The point in the other image lies along this epipolar line. This procedure makes it easier to find the matching points for the stereo vision.

A stereo vision system for determining a weld path for robot welding is presented in Paper IV, see Figure 3.2. Here the stereo vision method for points is modified to calculate the 3D geometry of curves. One camera on the robot hand was used. The camera was calibrated using the GCM. After calibration, the camera was positioned by the robot onto two different positions to capture images of the weld joint, see Figure 3.3. The image processing found the weld seam in each image, see Figure 3.4, and a stereo vision algorithm calculated the 3D geometry using the GCM. For the image processing determining the edges in the images the Canny edge detection method was used, see Section 4.1. The 3D geometry was sent from the calculation computer to the robot control system. After that the robot followed the joint with the tool, see Figure 3.2.

A stereo vision system for robot navigation, inspired by method (3.1), (3.2) has already been published in [55] by Lidbom et al. Another stereo vision system for determining robot weld paths can be found in Chen et al [19]. They however use parallel cameras to simplify the calculations, which is not needed in Paper IV and Paper V since they use (3.2). Again, if such a system uses a camera mounted on the robot hand, hand-eye calibration is needed as in Remy [68]. Stereo vision from a single planar catadioptric camera is presented in [83].

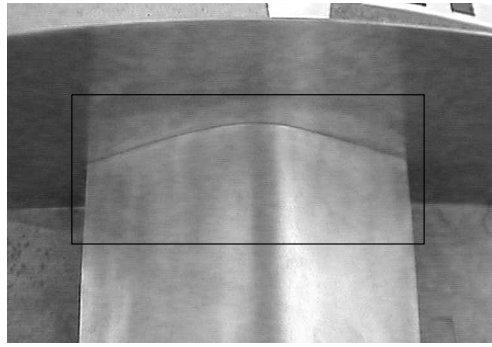


Figure 3.3: Image captured of metal sheets after preprocessing.

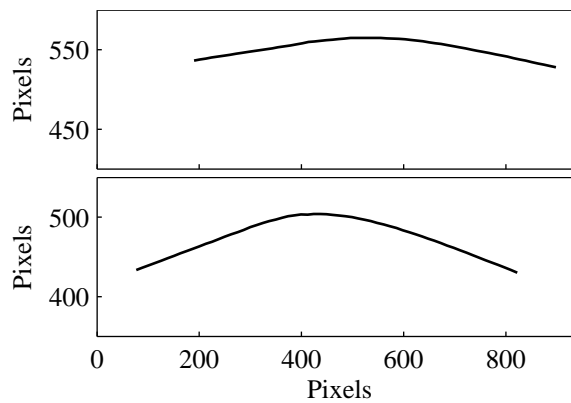


Figure 3.4: Edges found in the 2D-images of the weld joint. The images are taken from different poses.

3.3 Experimental Verification of Pose Measurement and Stereo Vision

The accuracy of the pose calculations is presented in Papers I and II. The experimental setup is presented in Figure 3.1. As references a plate with IR-LEDs was used. The LEDs were grouped in unique patterns of six LEDs. A pattern matching procedure recognised the patterns in the images to solve the correspondence problem. Electronic circuits triggered flashing of the LEDs at the same time as images were captured. The camera was calibrated using bundle adjustment (2). After the calibration the camera was mounted on the arm of a coordinate measurement machine, CMM. The CMM was programmed to stop at certain positions above the reference plate in order to measure with the CMM at the same time as an image was captured. Poses were calculated based on each image. Then the output of the CMM and the vision system were compared. They measured in different coordinate systems. Therefore the vision system coordinates were transformed to the CMM coordinate system before the comparison. The rotation and translation parameters of the transformation were calculated using optimisation methods. In Paper I, the results are shown for a conventional camera model, showing that the accuracy was not high enough for robot welding applications. In Paper II, conventional camera models are compared to the GCM based on the data from the CMM and the corresponding images. There the GCM had a slightly better accuracy than the

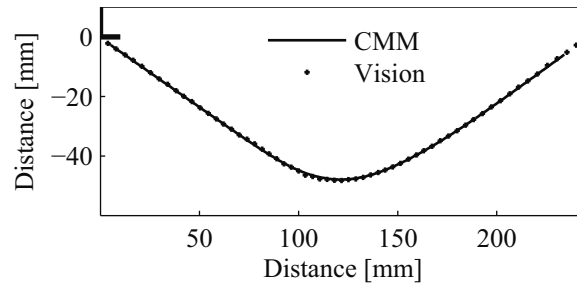


Figure 3.5: The calculated values for the joint from the vision system and the CMM-measures, both projected to a plane.

CCM, but due to primitive image processing the errors were mostly due to image errors and not the precision of the camera model chosen.

The 3D geometry from the stereo vision welding system in Paper IV was compared to measurements of the curve using a coordinate measurement machine, CMM, see Figure 3.5. The average error between the CMM and the calculated 3D curve was 0.23 mm, and the maximum error was 0.7 mm. This is acceptable for the welding application and promising for future work. The similar approach in [19] also use one camera mounted on the robot hand. They discuss e.g. sub-pixel methods for finding the weld joint, and their maximum position error was 3 mm.

Paper V verified the stereo vision robot system in a simulated environment. It shows that the camera calibration procedures, the image processing and the stereo vision calculations could be verified without access to a physical robot, camera and a prototype weld joint.

Related Computer Vision

Machine vision and computer vision are scientific areas in which information from camera images is extracted and used in a variety of applications, e.g. for guiding robots. Humans have an excellent ability to interpret and understand images, while computer perception from images is a difficult and challenging scientific field. In this section more computer vision methods are presented and related to the focus of this thesis on camera modelling and calibration and geometric calculations based on vision.

4.1 Image Processing

Image processing techniques process an image in order to obtain a new image that is enhanced in some way, e.g. yielding sharper edges. Edge detection techniques calculate where in an image the gradients of the intensity are large. The Sobel edge detection method [26] uses convolution filters that scan through the image. It provides a measure of how large the gradients of the intensity are locally around each pixel. It uses one convolution pattern for the x -direction and another for the y -direction. The output of the Sobel filter is two new images showing the components of the image intensity gradients in the two directions. The two new images can be combined into one image again by e.g. using the Pythagorean relation.

The Canny edge detection [14] also finds edges. In the Canny method, the image is first smoothed using a Gaussian filter in order to make the method less sensitive to noise. Then it looks through the image again, and when an edge is found it follows it in the two directions until the size of the gradient is lower than a certain threshold. The Canny edge detection algorithm does not only provide information of where the gradient is large, but also groups the edge segments to show which pixels belong to the same edge. Hence it finds segments of curves in the images, which is useful for applications in which the edge coordinates are searched for. The Canny edge detection method is used in Paper IV and V. For the images in the real physical system in Paper IV the Canny edge detection method found a large number of edges in the stereo images, caused by reflections and shadows. The edges not representing the welding joints needed to be filtered out. One method was to take away edge segments that were short. Another method of extracting the right edges was to analyse the direction of the edge curves. In the simulated environment the image processing of finding edges was easier, since there were no sharp reflections or shadows.

Corner detection methods find corners in the images, i.e. they find distinct points instead of curves in the images. A combined corner and edge detector is the Harris detector [37]. Other corner detectors are described by Moravec in [75], and Shi and Tomasi detector in [47].

Corner detection is relevant to camera calibration since it allows to calibrate a camera using images of an environment where only natural corners are used, so there is no need of adding extra references.

4.2 Image Part Recognition

Algorithms for automatically recognising parts in images can vary for different situations, but they usually follow the steps: Feature Selection → Segmentation → Feature Extraction → Classification, see Duda [27]. These procedures are discussed below and can be applied to grey scale images or any linear, or suitable non-linear, combination of the colour pixel intensities. Recognition of objects with known geometry is discussed in Section 4.2.5, and point feature detectors and descriptors are presented in Section 4.2.6. Image Part Recognition methods are related to camera calibration since they provide methods of solving the correspondence problem, making it possible to know which object space reference points or lines correspond to which image points or lines.

4.2.1 Feature Selection

The first step includes selection of features for classification of recognised parts. The features should be chosen so that they can distinguish between different objects that need to be recognised in the image, and they should also be possible to extract from the available data in the images. Examples of features can be found in Section 4.2.3.

4.2.2 Image Segmentation

The second step includes the segmentation of an image into different regions to facilitate the finding of objects in the image, see e.g. Sonka [75] and Pan [66]. Once it is known in which region of the image an object is located, features for part recognition can be extracted. Usually, it is assumed that there are sharp edges around an object in the image. This is similar to edge detection, but here an area, or a closed boundary of an object is searched for. The edges are boundaries between areas with darker and brighter pixels, i.e. places with large pixel gradients. If the image is noisy, a non-linear diffusion process can first be applied to the image as in Perona [67]. The diffusion is such that it smoothes out small variations in the image, but it keeps the more sharp edges. One simple method for segmentation is thresholding, see e.g. [75].

“Snakes” is one method to segment images as described in Ballerini [5] and Loizou [56]. A snake is usually a closed boundary in the image that automatically moves to find edges. A type of energy is defined for the snake. The energy is lower where the snake is aligned with a sharp edge, and also where it is not bending too sharply. The snake moves to find lower energy. It stops when it finds an optimal lowest energy, and then a segment is found, enclosed by the snake.

Watershed algorithms mimic running water to segment images, as in Sonka [75] and Karantzalos [46]. The image can be seen as a 3D topographic map, in which the pixel intensity represents the height. Using gradients, it can be concluded in which direction the water would flow in a point of the image. The water ends up in basins called catchment basins. Areas that make the water end up in the same catchment basin are considered to

belong to the same segment. The segments are divided by watersheds, so that on the other side of a watershed the water would flow into another catchment basin.

Region growing methods as in [75] start from a seed in a segment that grows to define the segment. The seed grows by sequentially trying to add pixels around it. A new pixel is added to the segment if it is inside the same segment, but not if it is considered to be outside. It stops when no more pixels can be added.

4.2.3 Feature Extraction

Features can be extracted when the image is segmented. Features can be defined in several ways. The shape of an object, i.e. the shape of its boundary, can be used as a feature, see e.g [75] where Fourier descriptors are used to analyse shapes. The shape of the object is found by the segmentation algorithm. Different kinds of moments can be used as features in the classification of an object. Moments define e.g. size, location, orientation and moments of inertia, see e.g. [75]. The texture describes the surface structure appearance, e.g. a square pattern, striped or dotted. The texture can be distinguished by using Fourier analysis, co-occurrence matrices or autocorrelation as in [75]. The colour of the object can also be used as a feature. The relative positions of corners and edges of the image can be used as features. They can be extracted with a corner and edge detection algorithm described in Section 4.1.

4.2.4 Classification

The fourth step is to use the features to classify what is seen in the image, e.g. Yau [84], Duda [27]. Usually a collection of objects are known, which possibly can be seen in a certain situation. By using the features, it should be possible to determine which of these objects is seen. The feature space is the mathematical space spanned by feature unit vectors representing different feature values, e.g. size and brightness. If values of all the features are known, this corresponds to a point in that space. Discrimination functions define where in the feature space the different objects are, see [75] or [27]. The discrimination functions are boundaries between the different parts; if the feature point is on one side of a discrimination function curve it is classified as one object, but if it is on the other side it is classified as another. Using support vector machines (SVM) is a method to find the discrimination functions, see Duda [27], Tang [78], Chen [20], Jeng [44] and Kumar [50]. The SVM non-linearly transforms the feature space into a usually higher dimensional space. In that space, linear discrimination functions are applied to classify the objects. Neural networks can be trained to classify objects, see Haykin [40] or Song [74]. The input to the network is the features that are found, and the output should be which object is observed, or the object class. The network can be trained by using known objects, where the feature values are extracted and the object class is known, so called supervised learning. Genetic algorithms can also be used to classify objects in Banzhaf [6], Tohka [80], Maulik [60]. The algorithm can be trained in a similar way as the neural networks.

4.2.5 Recognition of Objects with Known Geometries

If an object with a known geometry, e.g. from a CAD model, is to be found in the image, a slightly different method can be used. A camera model can be used to match a CAD model with an image. The corners and edges of an object can be projected to an image using the

camera model, and be matched with the geometries of corners and edges in the image. In that procedure the CAD model can be rotated and translated mathematically, to get close to the real image. If it is possible to find a good match, the object is recognised and the relative pose between camera and object is also known. Therefore, the procedure can be used for pose-control, by using the pose information as feedback. If the relative pose between object and camera is approximately known before the calculations, these approximations can be used as a starting point in the matching procedure, resulting in faster calculations. The matching can be formulated as an optimisation problem to be solved by an optimisation procedure. To perform the projections, the camera has to be calibrated, in order for the parameters of the camera model to be known.

4.2.6 Point Feature Detectors and Descriptors

The scale invariant feature transform, (SIFT), [57] by Lowe, is an algorithm for extracting certain points in images and then matching and recognising them to see if the same point appears in other images. When taking out the points, the images are smoothed using Gaussian filters with two different scales of the filters. Then the difference of these Gaussian (DOG) filtered images, is calculated. Local extremes of this DOG image are extracted. These are used as detected points in the images. Histograms are calculated locally around these points, describing how the image is varying around that point. The histogram is formed in a way that makes it invariant to orientation, distance, illumination and scale. If the histograms of two points of different images are equal enough, they are classified as images of the same point. Speeded up robust features, (SURF), described by Bay in [8], is an attempt to improve the SIFT algorithm. SIFT and similar methods are used in a large variety of applications, for example camera calibration, stereo matching, image stitching and object tracking.

4.3 Tracking

One common problem in vision systems is to track objects in image sequences. In doing that Kalman filters, see e.g. Grewal [34], are useful. They are methods for estimating the state of a dynamic system subject to random noise. The estimates are recursively calculated based on sequences of noisy measurements. The Kalman filter is divided into an update step and a prediction step. The particle filter, see e.g. Ristic [69], is a generalisation to non-linear systems. There are many different versions of particle filters, which are used in different situations. One example using catadioptric vision is described in [77] by Taiana. Another example of a particle filter is developed by Chen in [18].

Tracking is also related to camera calibration since it can solve the correspondence problem. If a video sequence is captured and reference points are tracked through the sequence, certain snapshots of the sequence can be used for the calibration calculation and the reference identification from the tracking can be used.

4.4 Visual Servoing

A visual servoing system is a system in which information from images is used for feedback control of e.g. robot motion. Chaumette [16], [17], Kragic [49] and Hutchinson [43] give

introductions to visual servoing. There are image-based and position-based visual servoing methods as well as hybrid approaches. The image-based methods try to minimise an error directly in the image by moving the robot. Position-based methods first process the image information to calculate the position or pose of an object which then is used as feedback to the robot. This is related to this thesis since many of the methods presented here can be used for calculating geometric information for especially position-based visual servoing. The calibrated camera model can also be used for calculating a reference image for image-based visual servoing in case the reference image can not be shown to the system using teach by showing as discussed in [49]. This can be the case e.g. if the desirable reference image changes in time. A seam tracker e.g. detects how well a welding robot follows a seam on-line while performing the weld. Vision can be used for this as in Kim [48], Yu [85] and [86]. Usually tracking is one part of a visual servoing system as in [30].

4.5 Virtual Vision

In order to develop robot vision systems off-line a simulated environment is useful. The simulated environment can be used for testing vision and robot programmes, which is presented in Paper V. The robot vision welding system developed for a real robot vision system in Paper IV was also implemented virtually. Calibration images are then captured in the virtual environment and then the system is calibrated. After that two images are captured of the weld seam to be used for the stereo vision calculations, and the stereo vision result is used for letting the welding robot weld the seam in the virtual environment. Hence both the camera and the robot are simulated in the same programme. The robot simulation makes it possible to test off-line programmed robot sequences, and in this case also in combination with testing camera systems and vision programs. The cameras are non distorted, and light sources can be added in suitable locations around a CAD geometry, to achieve similar conditions as in the real world. Shadows and reflections are calculated automatically and the image is calculated by the system.

4.6 Other Computer Vision Applications

Cameras can be used in a variety of industrial applications. Calibrated cameras are especially useful. This thesis has mentioned some, but far from all, application areas. Other examples of interesting camera-based applications are:

- A vision system which also can be used for quality control, as in [1] by Armingol. If methods of this thesis are used there is a camera available that at the same time can check quality or look for other errors in the working environment.
- Solving the problem of simultaneous localisation and mapping, called SLAM, see e.g. Milford [63]. The idea is to use some "agent" moving around in a world and from its sensors create a 3D map of the world and also localise the agent. The system can use e.g. vision, laser range finders and/or radar. It is hard to solve in a robust way using only vision. Images of flat surfaces without structure can e.g. not be used to calculate depth. A SLAM system using stereo vision and particle filters is described in [64] by Moreno. Another SLAM system using vision is presented in [76], by Spampinato et al. If vision is part of the SLAM system normally a calibrated camera model is needed.

- Robot calibration by optical methods can be performed using vision, as described in Kyle [51]. Dynamic photogrammetry calibration of industrial robots can be found in Maas [59] and Guo [36]. These systems also need calibrated camera models.

Conclusions and Future work

The GCM shows that it is possible to create a camera model that combines the advantages of fisheye camera models for fisheye cameras and the conventional camera models for low distortion cameras. The GCM can also model catadioptric cameras. It is a well-known problem in vision that different models used to be applied for different camera types, something which hereby is solved.

New types of distortion compensations are included in the GCM. Varying entrance pupil point is introduced in a natural and efficient way. Two types of decentring distortion are also presented. Ordinary radial distortion is included in a new way, that makes it more general. Using quotients between polynomials as radial distortion functions are motivated and turned out to increase the performance of the camera models. Methods for converting both from the 2D image to the 3D object space and vice versa are presented for the GCM.

A relation between the GCM and a model using homogeneous coordinates could be derived. The homogeneous coordinate radial distortion model is generalised to model the GCM with also varying entrance pupil point and decentring distortion. A relation between the GCM and the CCM is also presented, and methods for handling variable focus and zoom in the camera model as well as in the calibration are developed.

Non-trivial null spaces, or ambiguities, are identified, which improve the calibration. The awareness of them both improve the calculations as well as guide in how to capture the calibration images.

Pre-processing algorithms can calculate the starting values of both the intrinsic and extrinsic camera parameters as well as the 3D reference positions based on the shape, size and location of the image references. These are needed as initial data for the calibration calculation. Also, the pre-processing algorithms calculate more accurate image centre points of the references than the commonly used centres of gravity. The pre-processing algorithms are based on flat reference markers.

A simple expression for stereo vision calculations in object space has been suggested. The method is especially suitable for the new camera model, which efficiently projects from image to object space, and can include varying entrance pupil point. Commonly used methods minimise an image error, and need a non-distorted image for the stereo calculation. If the camera has a varying entrance pupil point, it is not possible to calculate the non-distorted image based on only the distorted image and the camera parameters.

The problem of detecting and determining the geometry of a weld joint, in order for a robot to follow it, could be solved using stereo vision from images captured by a camera mounted on the robot hand. The stereo vision calculations and camera calibrations are per-

formed using the GCM. This procedure is performed both in the real world and in a virtual environment. The virtual vision facilitates the development of vision systems off-line.

The accuracy of the camera models was investigated, by comparing camera poses calculated using the vision system, and corresponding position data obtained from a coordinate measurement machine. Further, error norms after calibration were calculated, showing that the overall performance and properties are favourable for the GCM compared to other models. The accuracy of a stereo vision system determining the geometry of weld joints was also investigated with satisfactory results.

For future research more types of decentring distortion, like the one in Slama [73], should be tested and compared to the methods proposed here. The GCM, CCM and FCM should also be compared for more real camera types.

The pre-processing algorithms should be validated, using circular and flat reference markers. This can be done either by simulations or with a real camera and real references. The accuracy of the parameters, robustness and calculation time should also be further analysed.

More work needs to be done in order to make the calculations more computationally efficient. Instead of MATLAB the calculations can be implemented in e.g. C which should make the calculations faster. They can also be implemented in hardware, e.g. FPGA circuits, which make the calculations much faster.

To industrialise the suggested methods, the computer programmes should finally be made more robust and more user friendly.

Summary of Appended Papers

This chapter presents a brief summary of the papers that the thesis is based on. They have been reformatted for uniformity and increased readability.

Paper I

Anders Ryberg, Anna-Karin Christiansson and Kenneth Eriksson. Accuracy Investigation of a Vision Based System for Pose Measurements. In *The 2006 9th IEEE International Conference on Automation, Robotics, Control and Vision*, Singapore. Dec 2006.

The accuracy of a vision based system for pose measurement was investigated. A camera was mounted on the end effector of a coordinate measurement machine and the outputs from the vision system and the coordinate measurement machine were compared in order to determine the accuracy of the vision system. Since they measured in different coordinate systems the position data from the vision system was transformed to the coordinate system of the coordinate measurement machine. The translation and rotation parameters of the transformation were calculated using a least squares optimisation matching method. The results showed that the accuracy was too low for robotic welding. One identified issue was related to the camera model, which motivates the development of a new camera model.

Paper II

Anders Ryberg, Anna-Karin Christiansson, Bengt Lennartson and Kenneth Eriksson. Camera Modelling and Calibration - with Applications. In book *Computer Vision*, Publisher: I-Tech Education and Publishing, Vienna, Austria, 2008.

This extensive book chapter presents a new generic camera model, camera calibration, non trivial null spaces that occur during calibration, calculations of approximate starting values of calibration parameters, calculation of centre points of references, applications for calibrated cameras and a new accuracy investigation. The new camera model includes new types of distortion compensations, such as varying entrance pupil point and two types of decentring distortion compensations. It also handles radial distortion in a new, more general, way. The new way of handling radial distortion makes it able to model both conventional low distor-

tion cameras as well as fisheye cameras. Because of the good properties of the new camera model, there is no longer a need for using certain models for certain camera types. The awareness of nontrivial null spaces that occur during calibration improves the camera calibration. Since the camera calibration is an extensive calculation, it is important to have good, reliable starting values of the parameters. These are calculated in a pre-processing algorithm by analysing the shape, size and position of the references in the images. The pre-processing algorithm also calculates better centre points of references than the centre of gravity. This also increases the accuracy of the calibration calculations.

Paper III

Anders Ryberg, Bengt Lennartson, Anna-Karin Christiansson, Lars Asplund and Mikael Ericsson. Analysis and Evaluation of a Generic Camera Model. Submitted to *Journal of Computer Vision and Image Understanding* 2010

In this paper the camera models are improved even further. A new way of analysing some camera models made it possible to refine them in a straight forward way. By using quotients between polynomials instead of just polynomials in the distortion compensation functions, the models are made more general, accurate and more computationally efficient. The GCM is also described in a slightly different way compared to in Paper II. It is shown how the GCM can model even catadioptric cameras in addition to fisheye and conventional cameras. Different camera model types are analysed and compared. In addition to the leaning detector compensation method, another general decentring distortion method is also introduced which is an improvement of a method in Paper II. Calibrations are performed for the different models, and the accuracy concerning two real cameras with conventional and fisheye lenses are compared.

Paper IV

Anders Ryberg, Mikael Ericsson, Anna-Karin Christiansson, Kenneth Eriksson, Jim Nilsson and Mathias Larsson. Stereo Vision for Path Correction in Off-Line Programmed Robot Welding. In *The 2010 IEEE International Conference on Industrial Technology*, Viña del Mar - Valparaiso, Chile. March 2010.

A stereo vision system for robot welding is developed. It uses one camera mounted on the welding robot's hand. The robot goes to two different positions to take photos of a weld joint. Based on the 2D images, the 3D welding curve is determined in order for the robot to weld it. The 3D path is calculated from the new GCM camera model. If a camera model with varying entrance pupil point such as the GCM is used, a stereo method minimizing the object space error is more convenient. The accuracy of the system is measured using a coordinate measurement machine, and an average error of 0.23 and maximum error of 0.7 mm has been achieved.

Paper V

Mikael Ericsson, Anders Ryberg, Jim Nilsson, Anna-Karin Christiansson and Bengt Lennartson. Off-line Simulation of Advanced Stereo Vision Welding Application. Submitted to *Journal of Machine Vision and Applications* 2010

In a virtual environment, robot vision systems can be simulated. This makes it possible to develop programs off-line, without shutting down the robotized production. The same robot and computer vision programmes described in Paper IV are implemented in the virtual environment. The images are captured virtually, and the calibration and the stereo vision calculations, as well as the image processing, are based on methods presented in this thesis work.

Bibliography

Bibliography

- [1] J. M. Armingol, J. Otamendi, A. De La Escalera, J. M. Pastor, and F. J. Rodriguez. Statistical pattern modeling in vision-based quality control systems. *Journal of Intelligent and Robotic Systems: Theory and Applications*, 37(3):321–336, 2003.
- [2] K. Åström and M. Oskarsson. Solutions and ambiguities of the structure and motion problem for 1D retinal vision. *Journal of Mathematical Imaging and Vision*, 12(2):121–135, 2000.
- [3] S. Baker and S. K. Nayar. Theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35(2):175–196, 1999.
- [4] H. Bakstein and T. Pajdla. Panoramic mosaicing with a 180° field of view lens. In T. Pajdla, editor, *Third Workshop on Omnidirectional Vision, 2002.*, pages 60–67, 2002.
- [5] L. Ballerini. Medical image segmentation using genetic snakes. *Proceedings of SPIE - The International Society for Optical Engineering*, 3812:13–23, 1999.
- [6] W. Banzhaf. *Genetic programming : an introduction : on the automatic evolution of computer programs and its applications*. M. Kaufmann, San Francisco, 1998.
- [7] A. Basu and S. Licardie. Alternative models for fish-eye lenses. *Pattern Recognition Letters*, 16(4):433–441, 1995.
- [8] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.
- [9] R. Benosman and S. B. Kang. *Panoramic vision: sensors, theory, and applications*. Springer, 2001.
- [10] Marten Björkman and Jan-Olof Eklundh. Real-time system for epipolar geometry and ego-motion estimation. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 506–513, 2000.
- [11] J-Y. Bouguet. Camera calibration toolbox, <http://www.vision.caltech.edu/bouguetj>. 2009, 2009.
- [12] C. Brauer-Burchardt and K. Voss. A new algorithm to correct fish-eye- and strong wide-angle-lens-distortion from single images. In *IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 225–228, Thessaloniki, 2001.
- [13] D. C. Brown. Close- range camera calibration. *Photogramm. Eng.*, 37(8):855–66, 1971.
- [14] J. Canny. Computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6):679–698, 1986.

- [15] J. S. Chahl and M. V. Srinivasan. Reflective surfaces for panoramic imaging. *Applied Optics*, 36(31):8275–8285, 1997.
- [16] F. Chaumette and S. Hutchinson. Visual servo control. i. basic approaches. *IEEE Robotics and Automation Magazine*, 13(4):82–90, 2006.
- [17] F. Chaumette and S. Hutchinson. Visual servo control. ii. advanced approaches [tutorial]. *IEEE Robotics and Automation Magazine*, 14(1):109–118, 2007.
- [18] H. Chen and Y. Li. Dynamic view planning by effective particles for three-dimensional tracking. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 39(1):242–253, 2009.
- [19] S.B. Chen, X.Z. Chen, and T. Qiu. Acquisition of weld seam dimensional position information for arc welding robot based on vision computing. *Journal of Intelligent and Robotic Systems*, 43:77–97, 2005.
- [20] Z. Chen, J. Li, and L. Wei. A multiple kernel support vector machine scheme for feature selection and rule extraction from gene expression data of cancer tissue. *Artificial Intelligence in Medicine*, 41(2):161–175, 2007.
- [21] T. A. Clarke and J. G. Fryer. The development of camera calibration methods and models. Technical report, Department of Electrical, Electronic, and Information Engineering, City University, London, UK. Department of Civil Engineering and Surveying, The University of Newcastle, Australia., 1998.
- [22] D. Claus and A. W. Fitzgibbon. A rational function lens distortion model for general cameras. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 213–219, 2005.
- [23] J. Courbon, Y. Mezouar, L. Eck, and P. Martinet. A generic fisheye camera model for robotic applications. In *IEEE International Conference on Intelligent Robots and Systems*, pages 1683–1688, San Diego, CA, 2007.
- [24] F. Devernay and O. Faugeras. Straight lines have to be straight. *Machine Vision and Applications*, 13(1):14–24, 2001.
- [25] F. Dornaika and C. Garcia. Pose estimation using point and line correspondences. *Real-Time Imaging*, 5(3):215–230, 1999.
- [26] R. O. Duda and P. E. Hart. *Pattern classification and scene analysis*. Wiley-Interscience, New York, 1973.
- [27] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern classification*. Wiley, New York, 2001.
- [28] A. K. Dunne, J. Mallon, and P. F. Whelan. Efficient generic calibration method for general cameras with single centre of projection. In *Proceedings of the IEEE International Conference on Computer Vision*, 2007.
- [29] O. D. Faugeras. *Three-dimensional computer vision : a geometric viewpoint*. MIT Press series in artificial intelligence, 99-0171310-2. MIT, Cambridge, Mass., 1993.
- [30] G. J. Garcia, J. Pomares, and F. Torres. Automatic robotic tasks in unstructured environments using an image path tracker. *Control Engineering Practice*, 17(5):597–608, 2009.

- [31] D. B. Gennery. Generalized camera calibration including fish-eye lenses. *International Journal of Computer Vision*, 68(3):239–266, 2006.
- [32] C. Geyer and K. Daniilidis. A unifying theory for central panoramic systems and practical implications. In *Computer Vision ECCV 2000*, pages 445–461. 2000.
- [33] C. Geyer and K. Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45(3):223–243, 2001.
- [34] M. S. Grewal and A. P. Andrews. *Kalman filtering : theory and practice*. Prentice-Hall information and system sciences series, 99-0139904-1. Prentice-Hall, Englewood Cliffs, N.J., 1993.
- [35] N. Guilbert, F. Kahl, and A. Heyden. Pose disambiguation in uncalibrated structure from motion. In *Proceedings of the 7th International Conference on Control, Automation, Robotics and Vision, ICARCV*, pages 132–137, 2002.
- [36] J-Y. Guo, J-L. Zhang, and T-S. Lu. Robot calibration using active vision-based measurement. *Journal of Dong Hua University (English Edition)*, 21(1):7–12, 2004.
- [37] C Harris. A combined corner and edge detector. *Plessey Research Roke Manor*, 1988.
- [38] R. Hartley and F. Kahl. Critical configurations for projective reconstruction from multiple views. *International Journal of Computer Vision*, 71(1):5–47, 2007.
- [39] R. Hartley and A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, Cambridge, 2003.
- [40] S. Haykin. *Neural networks : a comprehensive foundation*. Prentice Hall, Upper Saddle River, N.J., 1999. 2. ed.
- [41] J. Heikkilä. Geometric camera calibration using circular control points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1066–1077, 2000.
- [42] Anders Heyden and Kalle Astrom. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 438–443, 1997.
- [43] S. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12(5):651–670, 1996.
- [44] J-T. Jeng. Hybrid approach of selecting hyperparameters of support vector machine for regression. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(3):699–709, 2006.
- [45] J. Kannala and S. Brandt. 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, 2006.
- [46] K. Karantzalos and D. Argialas. Improving edge detection and watershed segmentation with anisotropic diffusion and morphological levellings. *International Journal of Remote Sensing*, 27(24):5427–5434, 2006.
- [47] C. S. Kenney, M. Zuliani, and B. S. Manjunath. An axiomatic approach to corner detection. volume I of *IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, pages 191–197, 2005.

- [48] J. S. Kim, Y. T. Son, H. S. Cho, and K. I. Koh. A robust visual seam tracking system for robotic arc welding. *Mechatronics*, 6(2):141–163, 1996.
- [49] D. Kragic and H. I. Christensen. Robust visual servoing. *International Journal of Robotics Research*, 22(10-11):923–939, 2003.
- [50] M. Kumar, R. Verma, and G. P. S. Raghava. Prediction of mitochondrial proteins using support vector machine and hidden markov model. *Journal of Biological Chemistry*, 281(9):5357–5363, 2006.
- [51] S. Kyle, R. Meyer, and G.D. van Albada. Robot calibration by optical methods. In *Proceedings of the IEE Colloquium on Next Steps for Industrial Robotics, May 17 1994*, IEE Colloquium (Digest), pages 9–1, London, UK, 1994.
- [52] R.K. Lenz and R.Y. Tsai. Calibrating a cartesian robot with eye-on-hand configuration independent of eye-to-hand relationship. In *Computer Vision and Pattern Recognition, 1988. Proceedings CVPR '88., Computer Society Conference on*, pages 67–75, 1988.
- [53] M. Lhuillier. Automatic scene structure and camera motion using a catadioptric system. *Computer Vision and Image Understanding*, 109(2):186–203, 2008.
- [54] M. Lhuillier. Toward automatic 3D modeling of scenes using a generic camera model. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, Anchorage, AK, 2008.
- [55] J. Lidholm, G. Spampinato, and L. Asplund. Validation of stereo matching for robot navigation. In *IEEE Conference on Emerging Technologies & Factory Automation*, pages 1–8, Mallorca, Spain, 2009.
- [56] C P. Loizou, C. S. Pattichis, M. Pantziaris, T. Tyllis, and A. Nicolaidis. Snakes based segmentation of the common carotid artery intima media. *Medical and Biological Engineering and Computing*, 45(1):35–49, 2007.
- [57] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [58] C.-P. Lu, G. D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(6):610–622, 2000.
- [59] H.-G. Maas. Dynamic photogrammetric calibration of industrial robots. *Proceedings of the SPIE - The International Society for Optical Engineering Videometrics V, 30-31 July 1997*, 3174:106–12, 1997.
- [60] U. Maulik. Unsupervised pattern classification using genetic algorithms. *Institution of Electronics and Telecommunication Engineers Journal of Research*, 48(5 SPEC):397–402, 2002.
- [61] B. Micusik. *Two-View Geometry of Omnidirectional Cameras*. PhD thesis, Czech Technical University in Prague, 2004.
- [62] B. Micusik and T. Pajdla. Estimation of omnidirectional camera model from epipolar geometry. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, 2003.

- [63] M. J. Milford. *Robot Navigation from Nature: Simultaneous Localisation, Mapping, and Path Planning based on Hippocampal Models*. Springer Tracts in Advanced Robotics, 1610-7438 ; 41. Springer-Verlag Berlin Heidelberg, Berlin, Heidelberg, 2008.
- [64] F. A. Moreno, J. L. Blanco, and J. Gonzalez. Stereo vision specific models for particle filter-based SLAM. *Robotics and Autonomous Systems*, 57(9):955–970.
- [65] A. Nowakowski and W. Skarbek. Lens radial distortion calibration using homography of central points. In W. Skarbek, editor, *EUROCON, 2007. The International Conference on "Computer as a Tool"*, pages 340–343, 2007.
- [66] Z. Pan, H. Wu, and M. Zhang. New method for fabric image segmentation and pattern recognition. *Journal of Computational Information Systems*, 2(1):415–420, 2006.
- [67] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(7):629–639, 1990.
- [68] S. Remy, M. Dhome, J. M. Lavest, and N. Daucher. Hand-eye calibration. volume 2 of *IEEE International Conference on Intelligent Robots and Systems*, pages 1057–1065, Piscataway, NJ, USA, 1997.
- [69] B. Ristic, S. Arulampalam, and N. Gordon. *Beyond the Kalman filter : particle filters for tracking applications*. Artech House, Boston, Mass. ;, 2004.
- [70] D. Scaramuzza. *Omnidirectional Vision: from Calibration to Robot Motion Estimation*. Phd thesis, ETH Zurich, 2008.
- [71] D. Scaramuzza. Omnidirectional camera calibration toolbox, <http://asl.epfl.ch/scaramuz/research>. 2009.
- [72] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002.
- [73] C. C. Slama. *Manual of photogrammetry*. American Society of photogrammetry, Falls Church, Va., 1980. 4. ed.
- [74] Y. Song, Y. Wan, P. Chen, and X. Chen. Simulation model for classification of remote sensing images by SOM neural networks. *Journal of Computational Information Systems*, 1(4):827–833, 2005.
- [75] M. Sonka, V. Hlavac, and R. Boyle. *Image processing, analysis, and machine vision*. PWS Publishing, Pacific Grove, Calif., 1999.
- [76] G. Spampinato, J. Lidholm, L. Asplund, and F. Ekstrand. Stereo vision based navigation for automated vehicles in industry. In *IEEE Conference on Emerging Technologies & Factory Automation*, Mallorca, Spain, 2009.
- [77] M. Taiana, J. Gaspar, J. Nascimento, A. Bernardino, and P. Lima. 3D tracking by catadioptric vision based on particle filters, 2008.
- [78] F. Tang, M. Chen, and Z. Wang. Approach to training support vector machine. *Journal of Systems Engineering and Electronics*, 17(1):200–205, 2006.

- [79] J. P. Tardif, P. Sturm, M. Trudeau, and S. Roy. Calibration of cameras with radially symmetric distortion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1552–1566, 2009.
- [80] J. Tohka, E. Krestyannikov, I. D. Dinov, A. MacKenzie Graham, D. W. Shattuck, U. Ruotsalainen, and A. W. Toga. Genetic algorithms for finite mixture model based voxel classification in neuroimaging. *IEEE Transactions on Medical Imaging*, 26(5):696–711, 2007.
- [81] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *Robotics and Automation*, 3(4):323 – 344, 1987.
- [82] G-Q. Wei, K. Arbter, and G. Hirzinger. Active self-calibration of robotic eyes and hand-eye relationships with model identification. *Robotics and Automation, IEEE Transactions on*, 14(1):158–166, 1998.
- [83] H. H. P. Wu, M. T. Lee, P. K. Weng, and S. L. Chen. Epipolar geometry of catadioptric stereo systems with planar mirrors. *Image and Vision Computing*, 27(8):1047–1061, 2009.
- [84] H-F. Yau, Y. O. Yang, and P-W. Chen. Optical demonstration of shift-, rotation- and limited size-invariant pattern recognition using a circular harmonic-synthetic discrimination function filter. *Optical and Quantum Electronics*, 28(10):1543–1550, 1996.
- [85] J. Y. Yu and S. J. Na. A study on vision sensors for seam tracking of height-varying weldment. part 1: Mathematical model. *Mechatronics*, 7(7):599–612, 1997.
- [86] J. Y. Yu and S. J. Na. A study on vision sensors for seam tracking of height-varying weldment. part 2: Applications. *Mechatronics*, 8(1):21–36, 1998.