Graz University of Technology

Institute for Computer Graphics and Vision, Graz University of Technology

Institute for Automation, University of Leoben

Dissertation

Integrating Vision and Robotics: A Highly Flexible Measurement Cell for Industrial Quality Inspection

Matthias Rüther

Graz, Austria, December 2009

Thesis supervisors

Prof. Dr. Horst Bischof

Prof. Dr. Paul O'Leary

Abstract

This thesis describes the combination of a projector-camera system and a robot manipulator to form a three-dimensional measurement system for industrial applications. One of the central observations is, that this combination is more than just a movable vision sensor. In an integrated system, the cameras support the robot in automated calibration, the projectors support the cameras through automated illumination optimization and the robot supports the cameras by effectively enhancing the measurement range.

A prototype of such a system is constructed and evaluated against real-world requirements in terms of flexibility, robustness and accuracy. In contrast to previous work in industrial quality inspection, the resulting measurement cell can solve a large variety of measurement problems without changing the underlying hardware and mechanical configuration. Applications range from three-dimensional (3D) measurement over two-dimensional (2D) image processing to recognition problems.

The methodological focus lies on optimization of geometrical accuracy and radiometric image quality. State-of-the-art calibration methods for cameras and projectors are integrated in a fully automatic procedure, which is enhanced by a novel method for robot calibration. The problem of structure estimation from multi-view measurements is addressed and a novel method is proposed, which is more accurate than traditional triangulation. Radiometric limitations of cameras and projectors are improved by exploiting the principle of high-dynamic range imaging and automatic illumination adaption on the projector side.

The underlying methods and the integrated system are extensively evaluated to prove the practical applicability and the increase in accuracy and robustness. It is demonstrated that geometrically complex measurement problems can be solved, even on objects with difficult surface properties, which would be infeasible or very expensive to solve with state-of-the-art systems.

Keywords. computer vision, industrial metrology, structured light, pattern adaption, robot calibration

Kurzfassung

Durch die Kombination von Projektor-Kamera Systemen und Industrierobotern ist es möglich, eine optische 3D Messzelle zu konstruieren. Die zentrale Beobachtung dabei ist, dass diese Kombination mehr darstellt, als nur ein bewegliches Bildverarbeitungssystem. In einem integrierten System ist es einerseits möglich, die kinematische Kette des Roboters mit Hilfe des Kamerasystems automatisiert zu kalibrieren, den Messbereich des Bildverarbeitungssystems auf den Bewegungsbereich des Roboters zu erweitern, sowie die Aufnahmesituation sowohl geometrisch durch Bewegen des Roboters, als auch radiometrisch, durch Adaptieren des Projektionsmusters, zu beeinflussen.

In dieser Arbeit wird der Prototyp eines solchen Systems vorgestellt und auf Genauigkeit, Robustheit und Flexibilität evaluiert. Die so entstandene Messzelle erlaubt es, eine große Bandbreite von messtechnischen Problemstellungen zu lösen, ohne die darunterliegende Hardware des Systems zu verändern. Die Bandbreite der lösbaren Messaufgaben reicht dabei von 3D-Oberflächenmessungen über "klassische" 2D-Bildverarbeitungsaufgaben bis hin zu Aufgaben der optischen Objekterkennung.

Der Fokus dieser Arbeit liegt aus methodischer Sicht auf der Optimierung von geometrischer Genauigkeit und radiometrischer Bildqualität. Methoden zur geometrischen Kalibrierung von Kameras, Projektoren und Industrierobotern werden in einer vollautomatischen Prozedur kombiniert und um eine neue Methode zur kinematischen Kalibrierung von Industrierobotern erweitert. Eine neue Methode zur Berechnung von 3D-Struktur aus mehreren Bildmessungen wird vorgestellt, die, speziell unter teilweisen Verdeckungen, genauer als die klassische optische Triangulierung ist. Die radiometrische Beschränkung von bestehenden Projektor-Kamerasystemen wurde durch Anwendung von radiometrischer Optimierung auf Projektorseite, und High-Dynamic-Range-Imaging auf Kameraseite verbessert.

Die Methoden werden in einem Forschungsprototyp extensiv evaluiert, um die praktische Anwendbarkeit, Genauigkeit und Robustheit zu zeigen. Die Lösung von geometrisch

komplexen Messaufgaben wird auf optisch schwierigen Oberflächen demonstriert.

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am ……………………………   …………………………………………………..
                                                                                        (Unterschrift)

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

……………………………   …………………………………………………..
        date   (signature)

Acknowledgments

This work would not have been possible without the support of many people. Especially, I wish to thank my supervisor Horst Bischof for his valuable and inspiring comments, and for always having time for discussion. I further wish to thank him for providing a fruitful and challenging working environment and supporting me in my day to day work. Further, I wish to thank Paul O'Leary for agreeing to act as my second thesis supervisor, and for the fruitful discussion.

This work can also be seen as one of the first cornerstones of a still growing research group at the ICG, namely the robot vision lab (http://rvlab.icg.tugraz.at). I hereby wish to express my gratitude to the diploma students who contributed, namely Wolfgang Heidl, Arno Knapitsch and Martin Lenz. Without them, it would not have been possible for me to cover the many different aspects of this work. Special thanks also to Franz Eberhart from M&R Automation, who provided me with valuable support from side of the company.

I am also grateful to my other diploma students and now fellow coworkers Markus Heber, Michael Maurer, Katrin Pirker, Jakob Santner and Georg Pacher for their commitment and enthusiasm in their projects. Especially, I want to thank all coworkers in the robot vision lab for providing me the freedom for writing this thesis over the last year.

My colleagues and alumni from the ICG should also be mentioned, who contributed to a unique and extremely fruitful atmosphere at the institute.

Finally, I am grateful for the love and support of my family and the mental support and patience of my girlfriend Elisabeth.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Contents

In the last decades, industrial automation has seen a strong trend towards outsourced manufacturing and production. This process, accompanied by the ever increasing complexity of manufactured goods, led to an increasing demand in quality inspection. Especially parts critical for an end-product's functionality require 100% inspection, which is a time consuming process. In this field of automated quality control, vision-based measurement systems are becoming key components. They provide the ability of non-tactile measurement, and are able to solve a large variety of measurement problems with one hardware device. The traditional domain of computer-vision devices are hereby two-dimensional inspection problems, e.g. reading characters or codes, measuring geometric properties like distances and hole diameters, or analyzing object color and texture.

The problem of measuring three-dimensional properties of machined objects is classical the domain of tactile sensors and point- or line-scanners, based on the laser triangulation principle. In the first case, the probe of a tactile sensor is moved towards the object until it makes contact. The nominal probe position, in combination with the local probe displacement, gives an accurate and very robust measurement of the contact point's location. The advantage of tactile methods is their outstanding robustness. Measurement quality is independent of optical properties like specularity or transparency. Measurements are easy to perform, because the probe position relative to the object can be directly observed and controlled. On the downside, accuracy requirements demand precision engineering of the

probe and motion system, which makes these methods expensive. The method returns point measurements only, so surface scanning at high resolution is time consuming and points not reachable by the probe cannot be measured.

A triangulation based sensor on the other side, projects a laser line or laser point on the object surface. The reflection is captured by a camera system which is rigidly connected to the light source. Three-dimensional information is calculated by optical triangulation. This method is non-tactile, which allows the measurement of non-rigid objects. If laser lines are projected, a profile scanner can be realized, which reduces measurement time for surface areas. The disadvantage of this approach is its inherent dependence on optical surface properties. Specular reflecting objects as well as transparent objects can not be measured by off-the-shelf triangulation sensors.

Vision systems, which provide three-dimensional measurements, are fundamentally based on the multi-view principle: if several (central-) perspective images of an object from different viewpoints are available and surface point correspondences and camera geometry are known, the three-dimensional location of features can be determined by optical triangulation. This measurement principle is inherently area-based, which makes object surface scanning fast. High flexibility and applicability to a large variety of measurement problems make this method a viable alternative to classical methods in industrial metrology, but no real replacement by itself, due to several shortcomings. It is arguably the least robust method and needs to be fine-tuned for most applications. Hardware requirements are similar to laser triangulation methods, but assembled in custom setups, which makes the approach relatively expensive. Measurement volume is limited by camera resolution, spatial constraints and the requirement that each measured point must be observable by at least two cameras from different viewpoints.

In this work, three principal drawbacks of vision based three-dimensional (3D) measurement systems are addressed: the need for customization to specific measurement tasks, the necessity to fine-tune illumination and camera parameters to specific measurement problems, and the limited measurement volume. This work contributes to the evolution of a "Plug and Play" industrial sensor, capable of performing a wide range of optical inspection tasks within a large measurement volume, by mounting the vision sensor on a mobile platform. The ability of automatic system calibration and the adaption of illumination, viewpoint and camera parameters, increases the flexibility of the system beyond that known from state-of-the-art vision solutions.

1.1 Problem Domain
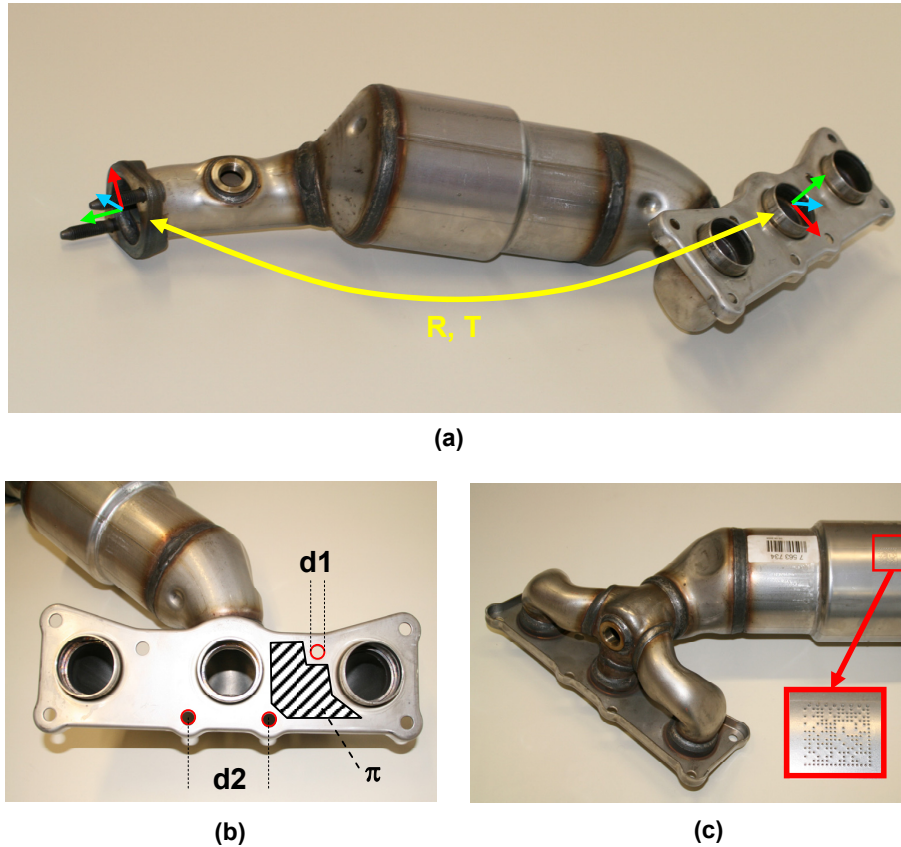
Figure 1.1: Typical measurement problem in industrial quality inspection. An exhaust pipe needs to be checked for dimensional accuracy of inlet and outlet flange (a), borehole accuracy (diameter d_1 and distance d_2) need to be checked, as well as flange planarity π (b). Measurement results need to be assigned to a code embossed in the pipe (c).

Quality inspection of medium scale machined parts (e.g. exhaust pipes) often requires multiple accurate geometric measurements spread across a large workspace. A typical example is shown in Figure 1.1. According to the user specification, a number of geometrical properties need to be checked:

- Borehole accuracy on inlet and outlet flanges. Distances between borehole centers and borehole diameter to an accuracy of 0.05mm.

- Flange planarity. The planarity of inlet and outlet flanges to an accuracy of 0.05mm.

- Dimensional accuracy of inlet and outlet flange. The relative position of inlet and outlet flange to an accuracy of 0.5mm.

- Part Identification. Read the embossed DataMatrix code.

To fulfill the requirements, a measurement solution has to be constructed which is able to perform geometric measurements in a volume of $1000 \times 500 \times 500\text{mm}^3$. The type of measurement reaches from two-dimensional image processing tasks for code reading, to three-dimensional measurements like surface planarity and position of feature points. The required measurement accuracy is up to 0.05mm for distinct tasks.

In a purely theoretical consideration, a single camera would require a resolution of 800megapixels to resolve an area of 0.5m^2 at $0.025 \times 0.025\text{mm}^2$ per pixel, and would not be able to handle the occlusions on a geometrically complex object.

A state-of-the-art solution would incorporate a number of specialized components, assembled in a measurement cell. For example, one laser triangulation sensor per flange, with motorized linear drive, one camera system per flange and one DataMatrix code reader. The relative position of all sensors has to be known, either by means of precise construction or some kind of calibration method. Furthermore, the sensors must be rigidly mounted, to avoid degradation of geometric calibration over time.

The main drawbacks of such a solution are costs and the lack of flexibility. Even a slight change in the measurement procedure or type of object would require reconstruction of the measurement cell, probably requiring additional sensor components and making others obsolete. Applying different sensors in a custom setup is expensive, especially when they have to be maintained over several years.

In the proposed solution, a single sensor head, which is capable of performing all required tasks in the previous example, is designed and evaluated. A vision system, consisting of multiple cameras and projection systems, is mounted on a moving platform, in this case an industrial manipulator. The construction resembles a measurement cell, where a single sensor is able to perform a variety of measurement tasks in a large workspace. The measurement volume is limited only by the robot workspace. Within this workspace, the sensor can be placed at various positions and perform a measurement within its local range. Several such measurements can be combined in a global coordinate frame (CF) using accurate knowledge of the manipulator motion.

While this idea is not new, few such systems were actually applied, mainly because:

- Retrieval of sensor motion requires accurate calibration and high stability of the manipulator. Usually, manipulator calibration is hardware intensive and time consuming.

- For most types of non-visual 3D sensors, accurate tool-hand calibration is not straightforward.

- Recalibration is part of the system maintenance process, often incurring prohibitive costs.

- Planning and implementation of inspection tasks is still hard, and often specific problems can not be solved with a single sensor device.

The proposed sensor is capable of measuring three-dimensional geometric features (surface planarity, angle, area, borehole diameter, position, circularity, edge length, linearity, etc.) in a volume of $180 \times 120 \times 80 \mathrm{mm}^3$. Implicitly, the sensor also acts as a position measurement device, enabling it to perform self-calibration, tool-hand calibration and kinematic calibration in a single framework. Only one accurate reference target ($150 \times 150 \mathrm{mm}^2$) is required. For kinematic calibration, a novel procedure based on self-inspection through a mirror is proposed, which only requires a planar mirror as an external reference.

1.2   Contributions

Compared to the state-of-the-art in vision based industrial metrology, the proposed system combines a variety of methods, which contribute to more precise, more robust, and easier to implement measurement tasks.

At first, the concept of a robot mounted structured light sensor, based on multiple projectors and multiple cameras is proposed. The sensor by itself allows to perform 3D and two-dimensional (2D) measurement in a limited volume, but with high accuracy. State-of-the-art micro-projectors with a modified light source are hereby combined with multiple industrial cameras to form a small, lightweight and rugged sensor, which can be mounted on an industrial manipulator.

Projector-camera systems are capable of performing 3D measurements by applying the well known structured light principle, a method which suffers in robustness, if metallic, shiny objects are measured. The following enhancements are discussed and evaluated: enhancement of sensor dynamic range through high dynamic range (HDR) imaging, and alternatively through automatic adaption of the projected illumination pattern. Scene interreflections cannot be eliminated by controlling only the light source and camera. Here, a method to robustly identify and mask out interreflection regions is introduced. All these methods are based on a simple radiometric self-calibration method, which can be conducted automatically by the sensor system.

After radiometry, the second most important portion of a 3D sensor is geometry. It is shown that the sensor by itself can be automatically calibrated, including camera, projector and multi-view calibration. Structure estimation, i.e. the problem of calculating 3D structure from known correspondences in image space, is hereby solved by a novel method, which increases accuracy at least by a factor of two. Robot geometry is only reliable, if it can be re-calibrated after installation on a frequent basis. Eye-hand calibration and kinematic optimization can hereby also be performed automatically. A novel method for kinematic calibration using only an external mirror is proposed in this course.

1.3 Outline

The thesis is organized as follows. In Chapter 2, the theoretical foundations for the proposed methods are discussed. Topics hereby range from projective geometry and geometrical image formation to barycentric interpolation, robot kinematics and parameter estimation. The overall concept of the proposed measurement system, and a possible hardware layout, are described and evaluated in Chapter 3. Chapter 4 contains the proposed radiometric image optimization techniques using active vision systems. Redundancy in the number of views, and number of images per view, as well as automatically controllable illumination are used to optimize illumination conditions for further measurement problems. Geometrical concepts, like the sensor model and methods for calibration and structure estimation are covered by Chapter 5. The introduced methods allow to generate a robust 3D sensor based on the structured light principle, which is discussed in Chapter 6. Flexibility in the measurement range and poses comes from the combination with an accurately calibrated robot. A mirror-based calibration method is proposed in Chapter 7. In Chapter 8, the complete sensor system is applied to the measurement problem introduced in this chapter, and evaluated. A summary of the presented work is given in Chapter 9.

Chapter 2

Theory and Background

Contents

2.1 Introduction

Theoretical foundations which are used throughout the work are reviewed in this chapter. Most of the topics, ranging from image-based metrology over three-dimensional metrology to robot calibration, are based on a closely related set of fundamental problems. All operate on geometric entities and many tasks finally result in inverse problems which need to be solved in an efficient way, so it is sensible to summarize the concepts of Euclidean geometry, parameter estimation and error modeling under a common notation to facilitate further reading.

First, the fundamentals of projective geometry are reviewed. The notation used in this work is closely related to the book of Hartley & Zisserman [42]. Because the notation used by photogrammetrists (see Atkinson [3] and Luhmann [71]) and robotics engineers (see Zhuang [118] and Sciavicco [95]) differs to a large extent, it was tried to find a common framework to express the basic concepts of geometry.

The following section is concerned with advanced projection models, especially those describing the deviation of real image sensors from the ideal pinhole model.

Kinematic manipulator models allow to describe robot geometry, pose and motion in a Euclidean world coordinate frame (`WCF`). The most popular models are reviewed, with a focus on the suitability for robot calibration.

Finally, the concepts of parameter estimation in Euclidean geometry are reviewed. Calibration procedures in computer vision and robotics almost inevitably lead to optimization problems, which are traditionally solved in a least squares manner. The problems of formulating the error metric, calculating the estimate and issues regarding parametrization and evaluation are discussed.

2.2   Projective Geometry

In dealing with geometric entities, rigid transformations and perspective projections, projective geometry provides a useful and elegant algebraic formulation. Purpose of this chapter is to introduce the basic concepts and define the notation. For an in-depth treatment, refer to the introductory chapters of [42], [26], [71] and [95].

In the first section geometrical entities and their representations in `2D` and `3D` are presented, followed by geometrical transformations and their hierarchy. The basic central perspective camera model is introduced in the final section.

2.2.1   Geometric Primitives

(a)   (b)   (c)
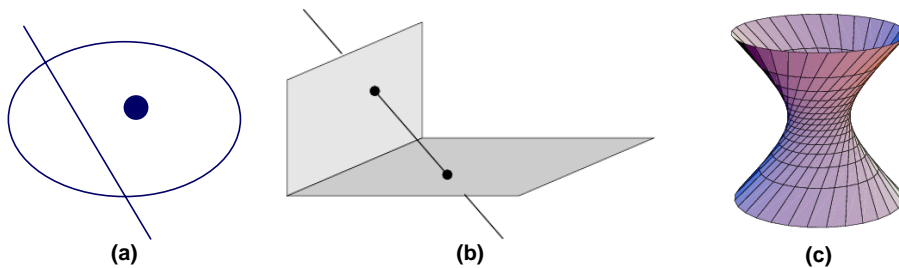
Figure 2.1: Geometric Primitives. Point, lines and conics (a) can be expressed as linear entities in $\mathbb{P}^2$. Planes, points and lines (b), as well as quadrics (c) are linear entities in $\mathbb{P}^3$.

Geometric primitives are elementary entities like points, lines or planes. Algebraically, these can be represented as vectors or matrices in projective space. Together with geometric transformations, they form the basis of projective geometry.

2.2.1.1   Points and Lines in 2D

Points and lines in the two-dimensional Euclidean space $\mathbb{R}^2$ (see Figure 2.1 (a)) are represented by column vectors $\mathbf{x}_{Euc} = (x_1, x_2)^T$. Note that boldface symbols always indicate vectors or matrices. Consequently, $\mathbf{x}_{Euc}$ has two degrees of freedom (DOF), namely its Cartesian coordinates. A line in $\mathbb{R}^2$ is naturally represented by the coefficient vector of the line equation $l_1 x_1 + l_2 x_2 + l_3 = 0$: $\mathbf{l} = (l_1, l_2, l_3)^T$. With three coefficients, the line is overparametrized and every vector $\mathbf{l}_k = k(l_1, l_2, l_3)^T$ with non-zero scalar k is a representation of the same line. Hence the set of all $\mathbf{l}_k$ forms an equivalence class. A point $\mathbf{x}$, defined as the intersection of two non-identical lines $\mathbf{l}_a$ and $\mathbf{l}_b$, must satisfy $\mathbf{Mx} = \mathbf{0}$, where

$$\mathbf{M} = \left[\begin{array}{c} \mathbf{l}_a^T \\ \mathbf{l}_b^T \end{array} \right].   \tag{2.1}$$

Because $\mathbf{M}$ has rank two, the right null-space of $\mathbf{M}$ is also an algebraic representation of points, with three coefficients and two DOF. Due to the freedom of scale, all $\mathbf{x}_k = k\mathbf{x}$ form an equivalence class, representing a single point. This representation is commonly denoted as a homogeneous coordinate vector. Division by the last coefficient gives again the Euclidean representation (see Table 2.1).

Vectors $\mathbf{l}$ and $\mathbf{x}$ are commonly denoted as homogeneous representations of lines and points, defined in the projective space $\mathbb{P}^2 = \mathbb{R}^3 \backslash (0, 0, 0)^T$. Algebraic properties are summarized in Table 2.1.

An important property is the duality principle, stating that in any theorem regarding points and lines in $\mathbb{P}^2$, another theorem can be derived by interchanging points and lines. This principle also holds for points in $\mathbb{P}^3$, conics, quadrics and their respective dual representations.

2.2.1.2   Points and Planes in 3D

Points and planes in $\mathbb{P}^3$ are a straightforward generalization from points and lines in $\mathbb{P}^2$. The role of lines is taken over by planes, which are algebraically represented as four dimensional vectors $\mathbf{p} = (p_1, p_2, p_3, p_4)^T$, and, dual to planes, points are represented as $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$. Algebraic properties are summarized in Table 2.2.

Relation	Algebraic formulation
Homogeneous point	$\mathbf{x} = (x_1, x_2, x_3)^T$
Homogeneous line	$\mathbf{l} = (l_1, l_2, l_3)^T$
Euclidean point	$\mathbf{x}_{Euc} = (\frac{x_1}{x_3}, \frac{x_2}{x_3})^T$
Intersection of two lines	$\mathbf{x} = \mathbf{l}_a \times \mathbf{l}_b$
Line joining two points	$\mathbf{l} = \mathbf{x}_a \times \mathbf{x}_b$
Line normal	$\mathbf{n} = (l_1, l_2)^T$
Points at infinity (ideal points)	$\mathbf{x}_\infty = (x_1, x_2, 0)^T$
Line at infinity	$\mathbf{l}_\infty = (0, 0, 1)^T$

Table 2.1: Algebraic relations of points and lines in $\mathbb{P}^2$.

Relation	Algebraic formulation
Homogeneous point	$\mathbf{x} = (x_1, x_2, x_3, x_4)^T$
Homogeneous plane	$\mathbf{p} = (p_1, p_2, p_3, p_4)^T$
Euclidean point	$\mathbf{x}_{Euc} = (\frac{x_1}{x_4}, \frac{x_2}{x_4}, \frac{x_3}{x_4})^T$
Intersection of three planes	$\begin{bmatrix} \mathbf{p_a}^T \\ \mathbf{p_b}^T \\ \mathbf{p_c}^T \end{bmatrix} \mathbf{x} = \mathbf{0}$
Plane joining three points	$\begin{bmatrix} \mathbf{x_a}^T \\ \mathbf{x_b}^T \\ \mathbf{x_c}^T \end{bmatrix} \mathbf{p} = \mathbf{0}$
Plane normal	$\mathbf{n} = (p_1, p_2, p_3)^T$
Points at infinity (ideal points)	$\mathbf{x}_\infty = (x_1, x_2, x_3, 0)^T$
Plane at infinity	$\mathbf{p}_\infty = (0, 0, 0, 1)^T$

Table 2.2: Algebraic relations of points and planes in $\mathbb{P}^3$.

2.2.1.3   Conics and Quadrics

Conic sections, such as ellipses, hyperbola and parabola, are expressed in $\mathbb{P}^2$ as second order polynomials in three parameters:

$$ax_1^2 + bx_1x_2 + cx_2^2 + dx_1x_3 + ex_2x_3 + fx_3^2 = 0. \tag{2.2}$$

The equation holds for all points lying on the conic. Combining coefficients a to f in a symmetric 3×3 matrix $\mathbf{C}$, (2.2) can be formulated in matrix notation:

$$\mathbf{x}^T\mathbf{C}\mathbf{x} = 0, \tag{2.3}$$

where

$$\mathbf{C} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}. \tag{2.4}$$

For a tangent line $\mathbf{l}$ to conic $\mathbf{C}$ in conic point $\mathbf{x}$, the following relation holds:

$$\mathbf{l} = \mathbf{C}\mathbf{x}. \tag{2.5}$$

If $\mathbf{x}_{pole}$ is a point lying outside $\mathbf{C}$, (2.5) is generalized to the pole-polar relationship

$$\mathbf{l}_{polar} = \mathbf{C}\mathbf{x}_{pole}. \tag{2.6}$$

Here, line $\mathbf{l}_{polar}$ is denoted polar line with respect to the pole $\mathbf{x}_{pole}$. An intuitive explanation of this relation is shown in Figure 2.2. If $\mathbf{x}_{pole}$ is outside $\mathbf{C}$, there exist exactly two lines through $\mathbf{x}_{pole}$ and tangent to $\mathbf{C}$. The two resulting tangent points uniquely define the polar line $\mathbf{l}_{polar}$.

In equivalence to points and lines in $\mathbb{P}^2$, there exists a dual entity for a point conic, namely the dual conic $\mathbf{C}^*$. In the non-degenerate case $\mathbf{C}^*$ may be computed as the inverse of $\mathbf{C}$:

$$\mathbf{C}^* = \mathbf{C}^{-1}. \tag{2.7}$$

The dual conic imposes a constraint on tangent lines $\mathbf{l}$, and defines a conic as the set of all lines satisfying
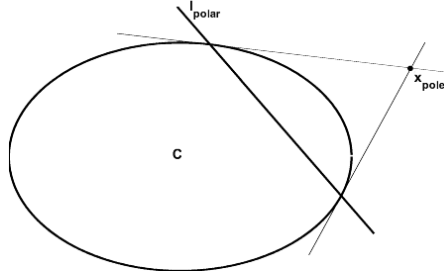
Figure 2.2: Pole-polar relationship. For a point $\mathbf{x}_{pole}$ outside conic $\mathbf{C}$, there exist exactly two tangent lines going through $\mathbf{x}_{pole}$. The line $\mathbf{l}_{polar}$ going through both tangent points is called polar to $\mathbf{x}_{pole}$.
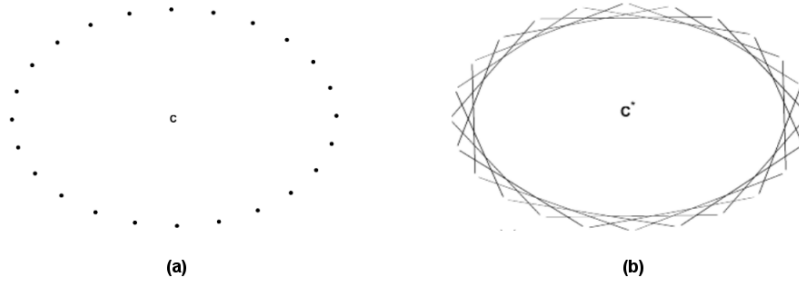
Figure 2.3: Conic representations. A conic may either be defined as the set of points satisfying $\mathbf{x^T C x} = 0$, or as the set of lines satisfying $\mathbf{l^T C^* l} = 0$. C^* is hereby the dual conic.

$$\mathbf{l^T C^* l} = 0. \tag{2.8}$$

In $\mathbb{P}^3$, quadrics are the three-dimensional counterparts to conics, for example ellipsoids, cones or hyperboloids. Algebraically they are represented by a symmetric 4×4 matrix $\mathbf{Q}$. A point lying on a quadric satisfies

$$\mathbf{x}^T \mathbf{Q} \mathbf{x} = 0, \tag{2.9}$$

and a tangent plane $\mathbf{p}$ satisfies

$$\mathbf{p}^T \mathbf{Q}^* \mathbf{p} = 0, \tag{2.10}$$

where $\mathbf{Q}^*$ is the dual quadric. In analogy to the two-dimensional case, constraints on tangent planes and the pole-polar relationship also hold:

$$\mathbf{p} = \mathbf{Q}\mathbf{x}, \tag{2.11}$$

where $\mathbf{p}$ is the plane tangent to $\mathbf{Q}$ in $\mathbf{x}$, if $\mathbf{x}$ lies on $\mathbf{Q}$, and the polar plane, if $\mathbf{x}$ is outside $\mathbf{Q}$. Note that the polar plane intersects $\mathbf{Q}$ in a conic, and each line through a conic point and the pole gives a line tangent to $\mathbf{Q}$.

2.2.2 Geometric Transformations

A special class of geometric transformations $\mathcal{F} : \mathbb{P}^n \to \mathbb{P}^n$ can be conveniently expressed as linear mappings. Here, linear transformations in $\mathbb{P}^3$ are reviewed, including their geometric invariants and `DOF`. Special emphasis is on the set of Euclidean, or rigid transformations and their parametrization, because these play a central role in multi-view geometry as well as robotics.

Generally, a mapping $\mathcal{F} : \mathbb{P}^3 \to \mathbb{P}^3$ is a projective transformation, if it is invertible and preserves collinearity. The algebraic representation is a left-multiplication with an invertible 4×4 matrix $\mathbf{H}$.

Under this mapping, a transformed point is given by $\mathcal{F}(\mathbf{x}) = \mathbf{H}\mathbf{x}$. Because the scale of $\mathbf{x}$ is arbitrary, so is the scale of $\mathbf{H}$, and consequently the most general projective mapping has 16 `DOF`. There exist a number of specializations of this mapping which have fewer `DOF` and leave certain geometrical properties unchanged. As shown in Figure 2.4, the specializations form a hierarchy, with the metric transform as the most special one, leaving only rotational and translational freedom (6 `DOF`), whereas the most general transformation has 15 `DOF`.

2.2.2.1 Projective Transformations

In the most general case, transformation matrix $\mathbf{H}$ has fifteen `DOF`. One parameter is free due to the scale freedom (see Figure 2.4). A collinear set of points is always mapped to a collinear set of points. Furthermore, the cross ratio, i.e. the ratio of length ratios of any four collinear points is preserved:

$$\frac{d(\mathbf{x}_a, \mathbf{x}_b)}{d(\mathbf{x}_a, \mathbf{x}_d)} : \frac{d(\mathbf{x}_b, \mathbf{x}_c)}{d(\mathbf{x}_b, \mathbf{x}_d)} = \frac{d(\mathbf{x}'_a, \mathbf{x}'_b)}{d(\mathbf{x}'_a, \mathbf{x}'_d)} : \frac{d(\mathbf{x}'_b, \mathbf{x}'_c)}{d(\mathbf{x}'_b, \mathbf{x}'_d)}. \tag{2.12}$$

Here, $d(\mathbf{x}, \mathbf{y})$ denotes the Euclidean distance between points $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{x}_i$ are original points and $\mathbf{x}'_i$ are transformed points.
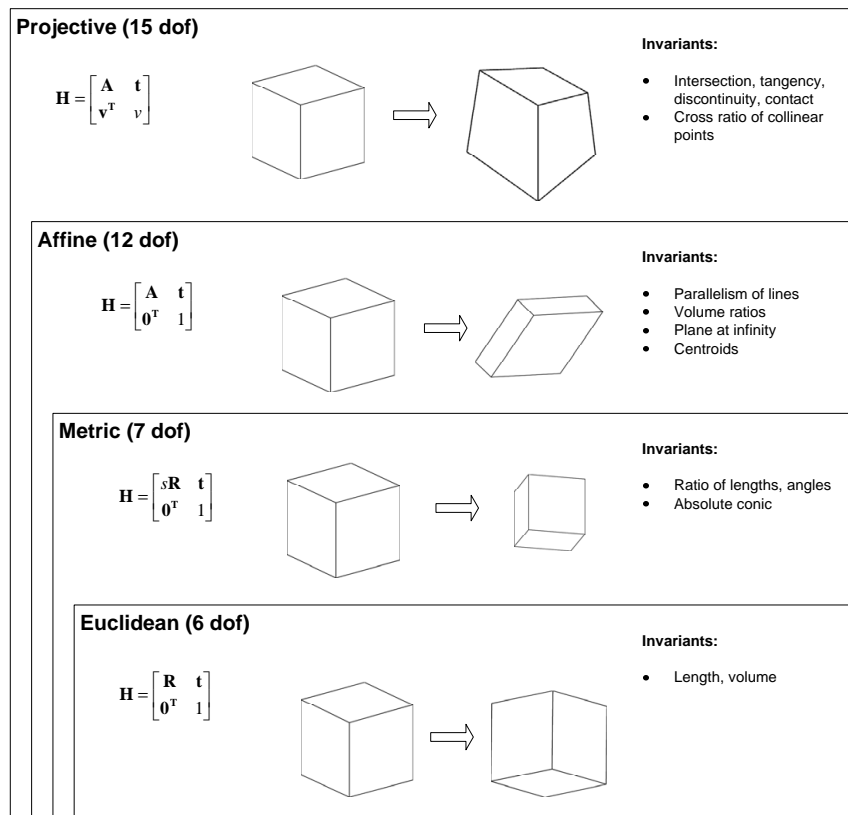
Figure 2.4: Hierarchy of Projective Transformations. Depending on the number of DOF in the transformation matrix **H**, certain geometric properties are left unchanged, resulting in a hierarchy of increasingly specialized transformations.

2.2.2.2 Affine Transformations

Affine transformations restrict the projective transform by setting the last row of **H** to $(0, 0, 0, \lambda)$, where λ is an arbitrary scalar. As shown in Figure 2.4, this leaves twelve DOF, representing anisotropic scaling (6 DOF), rotation (3 DOF) and translation (3 DOF). The most important invariant geometric property is parallelity of lines. As a consequence, the point of intersection of parallel lines, lying at infinity, remains at infinity after transformation. Further invariants include volume ratios, centroids and length ratios on parallel lines.

2.2.2.3 Similarity Transformations

Similarity transforms only change object rotation, translation and isotropic scale (7 DOF). The upper left 3×3 matrix becomes an orthogonal matrix, multiplied with the scale factor.

Angles and length ratios are preserved under affine transformations.

2.2.2.4 Metric Transformations

A metric transformation with six `DOF` changes rotation and translation. As a specialization of similarities, the scale factor is one. This class of transformations naturally describes the relative position of objects and `CF`, as well as relative movements in $\mathbb{P}^3$. As depicted in Figure 2.4, the rotation is represented by a 3×3 orthogonal matrix $\mathbf{R}$, the translation by a 3×1 translation vector $\mathbf{t}$.

Elementary rotations about the three coordinate axes take on special forms:

$$\mathbf{R}_x = \begin{bmatrix} 1 & 0 & 0 \\ 0 & c_\varphi & -s_\varphi \\ 0 & s_\varphi & c_\varphi \end{bmatrix} \quad \mathbf{R}_y = \begin{bmatrix} c_\varphi & 0 & s_\varphi \\ 0 & 1 & 0 \\ -s_\varphi & 0 & c_\varphi \end{bmatrix} \quad \mathbf{R}_z = \begin{bmatrix} c_\varphi & -s_\varphi & 0 \\ s_\varphi & c_\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad (2.13)$$

where s_φ, c_φ are sine and cosine of rotation angle φ.

2.2.2.5 Representation of Rotations

As shown in the previous section, a rotation in projective geometry is represented by a 3×3 orthogonal matrix, so three `DOF` are encoded in nine parameters with orthogonality and norm constraints. This representation has disadvantages in several applications, like parameter estimation, as will be shown in Section 2.5.

It would be favorable to have a representation which is more intuitive, locally continuous, differentiable and one-to-one. For parameter estimation, it should be locally approximately linear. Popular representations in this respect are Euler angles, unit quaternions and axis-angle representations.

Euler Angles   Euler angles are a minimal representation of rotation with three angles. Based on the observation that any rotation in `3D` can be represented by three consecutive rotations about non-parallel axes, one defines a vector $\boldsymbol{\Phi} = (\varphi, \vartheta, \psi)^T$, where each angle specifies a rotation about one coordinate axis. Since the exact sequence of axis rotations is not defined, and the only constraint is that consecutive rotations are not about the same axis, twelve distinct Euler representations are possible. Although no common agreement exists on what representation to use, two of the more popular examples are the ZYZ representation and the roll-pitch-yaw (`RSC`) angles.

The ZYZ Euler vector $\mathbf{\Phi} = (\varphi, \vartheta, \psi)^T$ contains three angles, where φ represents a rotation angle around the z-axis, ϑ is a rotation angle around the current y-axis (i.e. y-axis after the first rotation) and ψ is a rotation angle about the current z-axis (after the first two rotations):

$$\mathbf{R}_\Phi = \mathbf{R}_z(\varphi)\mathbf{R}_{y'}(\vartheta)\mathbf{R}_{z''}(\psi). \tag{2.14}$$

A formula for direct computation of $\mathbf{R}_\Phi$ is given by

$$\mathbf{R}_\Phi = \begin{bmatrix} c_\varphi c_\vartheta c_\psi - s_\varphi s_\psi & -c_\varphi c_\vartheta s_\psi & c_\varphi s_\vartheta \\ s_\varphi c_\vartheta c_\psi - c_\varphi s_\psi & -s_\varphi c_\vartheta s_\psi & s_\varphi s_\vartheta \\ -s_\vartheta c_\psi & s_\vartheta s_\psi & c_\vartheta \end{bmatrix}, \tag{2.15}$$

and for the inverse mapping:

$$\begin{aligned} \varphi &= \arctan\left(\frac{r_{23}}{r_{13}}\right) \\ \vartheta &= \arctan\left(\frac{\sqrt{r_{13}^2 + r_{23}^2}}{r_{33}}\right) \\ \psi &= \arctan\left(\frac{r_{32}}{-r_{31}}\right). \end{aligned}$$

The representation has a singularity if ϑ is zero or a multiple of π, which would lead to a double rotation about the identical z-axis.

The second representation, RPY angles, is commonly chosen for ease of understanding. Here, similar to the motion of an airplane, the angle vector $\mathbf{\Phi} = (\varphi, \vartheta, \psi)^T$ represents a sequence of rotations about fixed reference axes:

$$\mathbf{R}_\Phi = \mathbf{R}_x(\psi)\mathbf{R}_y(\vartheta)\mathbf{R}_z(\varphi), \tag{2.16}$$

or equivalently about current axes

$$\mathbf{R}_\Phi = \mathbf{R}_z(\varphi)\mathbf{R}_{y'}(\vartheta)\mathbf{R}_{x''}(\psi). \tag{2.17}$$

The forward mapping is hereby given by:

$$\mathbf{R}_\Phi = \begin{bmatrix} c_\varphi c_\vartheta & c_\varphi s_\vartheta s_\psi - s_\varphi c_\psi & c_\varphi s_\vartheta c_\psi + s_\varphi s_\psi \\ s_\varphi c_\vartheta & s_\varphi s_\vartheta s_\psi - c_\varphi c_\vartheta & s_\varphi s_\vartheta c_\psi - c_\varphi s_\psi \\ -s_\vartheta & c_\vartheta s_\psi & c_\vartheta c_\psi \end{bmatrix}. \tag{2.18}$$

For the inverse mapping, the following equations hold:

$$
\begin{aligned}
\varphi &= \arctan\left(\frac{r_{21}}{r_{11}}\right) \\
\vartheta &= \arctan\left(\frac{-r_{31}}{\sqrt{r_{32}^2 + r_{33}^2}}\right) \\
\psi &= \arctan\left(\frac{r_{32}}{r_{33}}\right).
\end{aligned}
\tag{2.19}
$$

This representation degenerates if $c_\vartheta = 0$, leading again to a double rotation about identical axes.

Unit Quaternions   Unit quaternions are a four parameter representation of rotations, also denoted as Euler parameters. A quaternion vector $\mathbf{Q} = (\eta, \epsilon_x, \epsilon_y, \epsilon_z)^T$ consists of a scalar part and a three element vector:

$$
\begin{aligned}
\eta &= \cos\frac{\vartheta}{2} \\
\epsilon &= \sin\frac{\vartheta}{2}t,
\end{aligned}
\tag{2.20}
$$

where ϑ is the rotation angle and t is the rotation axis. While the parameter space is essentially four-dimensional, the unit quaternion is constrained to lie on the unit sphere: $\|\mathbf{Q}\|_2 = 1$. The corresponding rotation matrix is of the form:

$$\mathbf{R}(\eta, \epsilon) = \begin{bmatrix} 2(\eta^2 + \epsilon_x^2) - 1 & 2(\epsilon_x \epsilon_y - \eta \epsilon_z) & 2(\epsilon_x \epsilon_z + \eta \epsilon_y) \\ 2(\epsilon_x \epsilon_y + \eta \epsilon_z) & 2(\eta^2 + \epsilon_y^2) - 1 & 2(\epsilon_y \epsilon_z - \eta \epsilon_x) \\ 2(\epsilon_x \epsilon_z) - \eta \epsilon_y & 2(\epsilon_y \epsilon_z + \eta \epsilon_x) & 2(\eta^2 + \epsilon_z^2) - 1 \end{bmatrix}. \tag{2.21}$$

The following result allows to compute a unit quaternion from a given rotation matrix:

$$\eta \;=\; \frac{1}{2}\sqrt{r_{11} + r_{22} + r_{33} + 1} \tag{2.22}$$

$$\epsilon \;=\; \frac{1}{2}\begin{bmatrix} \mathrm{sgn}(r_{32} - r_{23})\sqrt{r_{11} - r_{22} - r_{33} + 1} \\ \mathrm{sgn}(r_{13} - r_{31})\sqrt{r_{22} - r_{33} - r_{11} + 1} \\ \mathrm{sgn}(r_{21} - r_{12})\sqrt{r_{33} - r_{11} - r_{22} + 1} \end{bmatrix}$$

Quaternions as rotation representations have several nice properties. They are easy to construct from rotation matrices or axis-angle representations, do not have singularities or discontinuities and behave approximately linear for small changes in rotation. Additional algebraic properties are rotation inversion, which is computed as

$$\mathbf{Q}^{-1} = (\eta, -\epsilon), \tag{2.23}$$

the composition of two rotations $\mathbf{R}_1 \cdot \mathbf{R}_2$ as

$$\mathbf{Q}_1 * \mathbf{Q}_2 = (\eta_1\eta_2 - \epsilon_1^T\epsilon_2, \;\; \eta_1\epsilon_2 + \eta_2\epsilon_1 + \epsilon_1 \times \epsilon_2). \tag{2.24}$$

The identity element for quaternion multiplication $*$ is given by $\mathbf{Q}_i = (1, \mathbf{0})$.

Axis-Angle Representation Clearly, a rotation in 3D can also be represented by the rotation axis $\mathbf{t}_{3\times 1}$ and the rotation angle ϑ around this axis, leading to a four parameter representation. A minimal parametrization however can be reached by encoding ϑ in the norm of $\mathbf{t}$: $\mathbf{t} = \vartheta\hat{\mathbf{t}}$, where $\hat{\mathbf{t}}$ is the unit vector in direction of the rotation axis.

Calculation of the rotation matrix is done via the Rodrigues formula for a rotation matrix:

$$\mathbf{R}(\vartheta, \hat{\mathbf{t}}) = \mathbf{I} + \sin\vartheta[\hat{\mathbf{t}}]_\times + (1 + \cos\vartheta)[\hat{\mathbf{t}}]_\times^2. \tag{2.25}$$

Conversion of $\mathbf{t}$ to a unit quaternion is given by

$$\mathbf{Q} = (\mathrm{sinc}(\|\mathbf{t}\|/2)\mathbf{t}, \;\; \cos(\|\mathbf{t}\|)/2)^T. \tag{2.26}$$

2.2.3 Central Perspective Camera Model

In addition to geometric primitives and their transformations, projective geometry also allows an elegant formulation of image formation by central perspective projection. As-

suming the ideal pinhole camera model, a `2D` image is formed by intersection of a ray through a world point $\mathbf{X}$ and the center of projection $\mathbf{C}$ with an image plane π_i (see also Figure 2.5). This mapping from $\mathbb{P}^3$ to $\mathbb{P}^2$ has a total of eleven `DOF`, six for the relative orientation between camera and scene, and five for transformations in the image space.
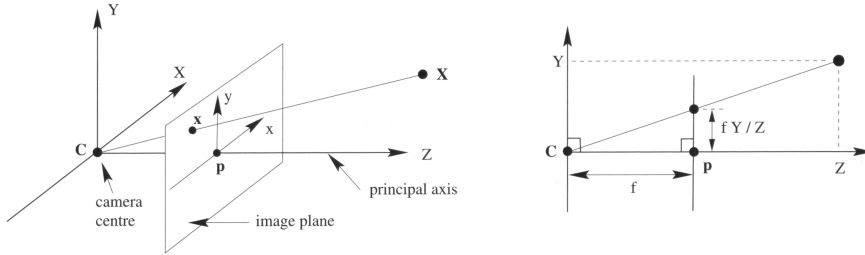
Figure 2.5: Central perspective image formation. The mapping from world point $\mathbf{X}$ to image point $\mathbf{x}$ is algebraically described by eleven parameters, six of which describe the pose of the `CCF`, and five describe transformations in the image plane [42].

At first, one defines a local `CCF`, which coincides with the `WCF`. The `CCF` is defined such that the origin lies in the projection center, the z-axis is perpendicular to the image plane and the camera looks in positive z-direction. To fix the remaining rotation ambiguity, the xz-plane intersects the image plane in the image x-axis, where world x-axis and image x-axis look in the same direction. Such a configuration allows for a simple algebraic formulation of central perspective projection:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} f\frac{X}{Z} \\ f\frac{Y}{Z} \end{bmatrix}, \tag{2.27}$$

where the image point $\mathbf{x} = (x, y)^T$, world point $\mathbf{X} = (X, Y, Z)^T$ (compare also Figure 2.5). The scalar f, commonly denoted as the focal length, defines the perpendicular distance between image plane an projection center. It is an intrinsic camera parameter, because it specifies an isotropic scaling in image space.

A projective formulation of (2.27) is even simpler, leading to a linear relationship:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} f & & 0 \\ & f & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}, \tag{2.28}$$

or, equivalently:

$$\mathbf{x} = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix}_{3\times3} \begin{bmatrix} \mathbf{I}_{3\times3} & | & \mathbf{0}_{3\times1} \end{bmatrix}_{3\times4} \mathbf{X}. \tag{2.29}$$

The first 3×3 matrix, currently containing only f, will further be denoted as the calibration matrix $\mathbf{K}$. Multiplying $\mathbf{K}$ with the 3×4 matrix $[\mathbf{I}|\mathbf{0}]$, where $\mathbf{I}$ is the 3×3 identity matrix and $\mathbf{0}$ a three element null vector, gives the projection matrix $\mathbf{P}_{3\times4}$:

$$\mathbf{x} = \mathbf{PX}, \qquad \mathbf{P} = \mathbf{K} \begin{bmatrix} \mathbf{I}_{3\times3} & | & \mathbf{0}_{3\times1} \end{bmatrix}. \tag{2.30}$$

All intrinsic camera parameters, describing transformations in image space, are contained in $\mathbf{K}$, which is upper triangular and defined up to scale. So a total of five intrinsics is possible:

$$\mathbf{K} = \begin{bmatrix} f_x & s & p_x \\ & f_y & p_y \\ & & 1 \end{bmatrix}. \tag{2.31}$$

According to Figure 2.6, the five intrinsic parameters have a geometric interpretation. The focal length is a scaling of the image coordinate frame (ICF) relative to the WCF. This can be achieved by leaving the ICF constant and changing f, which moves the image plane relative to the projection center. So f, or f_x in (2.31), is the perpendicular distance between image plane and projection center. If the scaling is different along the coordinate axes, e.g. by using rectangular pixels or in scanned images, f_y differs from f_x by an aspect ratio $a = f_y/f_x$.

The ICF origin lies at the principal point $\mathbf{p} = (p_x, p_y)^T$, which is the orthogonal projection of $\mathbf{C}$ onto the image plane π_i (see Figure 2.5). The origin is translated to some other place for convenience reasons, e.g. to the upper-left pixel position of a charge coupled device (CCD) array. This translation is modeled by the principal point offset $(p_x, p_y)^T$. In the rare case when one has to deal with non-rectangular pixels, a non-zero skew factor may be applied, leading to a pixel shear of angle α, as shown in Figure 2.6:

$$s = f_x/p_y \tan \alpha. \tag{2.32}$$

The projection model from (2.30) can also be interpreted in a different way. First, $\mathbf{X}$ is mapped to a canonical image point
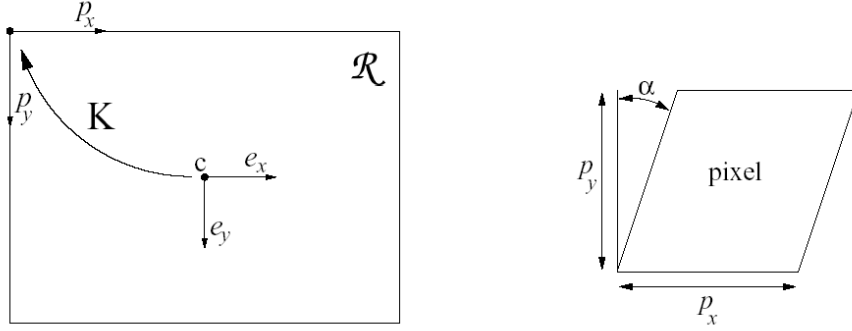
Figure 2.6: Intrinsic camera parameters [87]. The calibration matrix $\mathbf{K}$ contains a set of five linear transformations in image space $\mathbb{P}^2$. Focal lengths f_x, f_y correspond to anisotropic scaling, the principal point coordinates p_x, p_y define a translational offset and the skew parameter s specifies a pixel shear.

$$\mathbf{x}_c = [\mathbf{I}|\mathbf{0}]\mathbf{X} = (X_1/X_3, X_2/X_3, 1)^T. \tag{2.33}$$

Final image coordinates are calculated as $\mathbf{x} = \mathbf{K}\mathbf{x}_c$, so $\mathbf{K}$ is equivalent to a 2D projective transformation in the image plane, containing a translation $(p_x, p_y)^T$, shearing s and anisotropic scaling $(f_x, f_y)^T$.

For notational reasons it is desirable to split $\mathbf{K}$ into two transformations $\mathbf{K} = \mathbf{K}_i\mathbf{K}_m$:

$$\mathbf{x} = \mathbf{K}_i\mathbf{K}_m\mathbf{x}_c = \begin{bmatrix} d_x^{-1} & s_i & p_x \\ & d_y^{-1} & p_y \\ & & 1 \end{bmatrix} \begin{bmatrix} f_m & & \\ & f_m & \\ & & 1 \end{bmatrix} \mathbf{x}_c. \tag{2.34}$$

As such, $\mathbf{K}_m$ is a transformation from canonical images space to metric image space

$$\mathbf{x}_m = \mathbf{K}_m\mathbf{x}_c, \tag{2.35}$$

where f_m is the camera focal length in metric units (e.g. millimeters). Point $\mathbf{x}_m$ is also given in metric units and may be further transformed to pixel units through $\mathbf{x}_i = \mathbf{K}_i\mathbf{x}_m$, where $(d_x, d_y)^T$ are the metric pixel dimensions and $s_i = s(fd_y)^{-1}$.

Until now, the WCF coincided with the CCF. To realize a different camera pose, a six DOF Euclidean transformation is defined, which aligns WCF and CCF before projection:

$$\mathbf{x} = \mathbf{K}\begin{bmatrix} \mathbf{I} & | & \mathbf{0} \end{bmatrix}\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}\mathbf{X} \quad \Leftrightarrow \quad \mathbf{x} = \mathbf{K}\begin{bmatrix} \mathbf{R} & | & \mathbf{t} \end{bmatrix}\mathbf{X}, \tag{2.36}$$

where $\mathbf{R}$ and $\mathbf{t}$ represent the rotation and translation necessary to transform points from the WCF to the CCF. The position of the projection center in world coordinates is given by inverse transformation of $(0,0,0,1)^T$: $\mathbf{C}_{world} = -\mathbf{R}^T\mathbf{t}$.

To summarize, the projection matrix

$$\mathbf{P} = \mathbf{K} \left[\begin{array}{c|c} \mathbf{R} & \mathbf{t} \end{array} \right] \tag{2.37}$$

describes a linear mapping from $\mathbb{P}^3$ to $\mathbb{P}^2$ and has eleven DOF. Inversely, every 3×4 matrix represents a finite projective camera, as long as the left 3×3 submatrix is non-singular. While the coefficients of $\mathbf{P}$ do not have a direct geometric interpretation, the matrix can be dissected in an upper triangular matrix $\mathbf{K}$ containing five intrinsic parameters, an orthogonal matrix $\mathbf{R}$ and a vector $\mathbf{t}$ containing six extrinsic or pose parameters.

2.3   Advanced Projection Models

Although the central perspective camera model described in Section 2.2.3 perfectly describes a pinhole camera, it is less useful for modeling real cameras. The reason is the presence of systematic deviations when using optical components like lenses and mechanical imperfections.

| (a)   (b) |
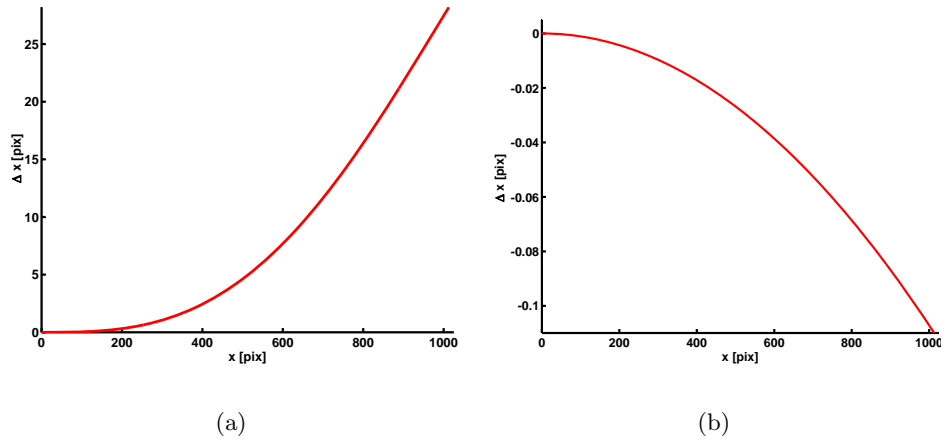
Figure 2.7: Typical lens distortion curves. The Figures show distortion curves of an 8mm C-mount lens mounted on a 1/1.8" CCD sensor. Displacement in x-direction (Δx) is given over x, with the origin at the center of distortion. Radial lens distortion (a) is larger by an order of magnitude than decentering distortion (b).

The largest amount of distortion is a radial symmetric displacement in radial direction

(Figure 2.7 (a)), with distortion center close to the principal point. Commonly known as barrel or pincushion distortion, this effect is caused by lenses in the optical path. Mechanical deficiencies, like inaccurate alignment of lenses and distortions of the image plane, cause additional displacements which are usually smaller by at least one order of magnitude. Lens misalignment, as shown in Figure 2.8, is modeled by decentering distortion, the other effects are usually neglected.
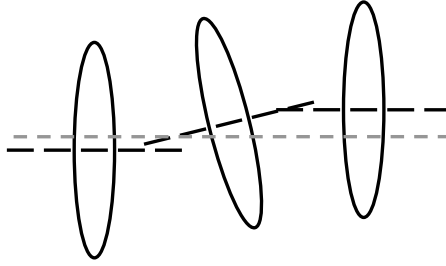
Figure 2.8: Reason for decentering lens distortion. Decentering distortion originates from misaligned lenses. The optical effect is comparable to inserting a thin prism in the optical path.

The gold standard in modeling these effects has been established by Brown [13] [14], who explicitly modeled distortion displacement by a radial component and a decentering component:

$$\mathbf{x}_{corr} = \mathbf{x}_{meas} + \boldsymbol{\Delta}_{rad}(\mathbf{x}_{meas}, \kappa_1, ..., \kappa_n) + \boldsymbol{\Delta}_{tan}(\mathbf{x}_{meas}, p_1, ..., p_n), \qquad (2.38)$$

where $\mathbf{x}_{meas}$ is the measured (i.e. distorted) image point and $\mathbf{x}_{corr}$ is the corrected point, corresponding to an ideal central perspective projection.

The radial displacement $\boldsymbol{\Delta}_{rad}$, specified by a set of radial distortion parameters $\kappa_1, ..., \kappa_n$, is given by

$$\boldsymbol{\Delta}_{rad} = \bar{\mathbf{x}}(\kappa_1 r^2 + \kappa_2 r^4 + ... + \kappa_n r^{2n}), \qquad \bar{\mathbf{x}} = \mathbf{x}_{meas} - \mathbf{p}. \qquad (2.39)$$

The displacement function is symmetric around $\mathbf{p}$ and a function of radius $r = \|\bar{\mathbf{x}}\|$, so the center of distortion coincides by definition with the principal point. In practice, the number of radial distortion parameters is at most three.

Tangential or decentering displacement Δ_{tan} is given by

$$\boldsymbol{\Delta}_{tan} = \begin{bmatrix} p_1(r^2 + 2\bar{x}^2) + 2p_2\bar{x}\bar{y} \\ p_2(r^2 + 2\bar{y}^2) + 2p_1\bar{x}\bar{y} \end{bmatrix} (1 + p_3 r^2 + ...). \qquad (2.40)$$

Only the first two distortion parameters are considered in practice. To model deficiencies of the sensor plane, Brown [16] later introduced two more additive displacement models, which are purely empirical, polynomial descriptions. First the in-plane distortion

$$\mathbf{\Delta}_{ip} = \left[\begin{array}{c} a_1\bar{x} + a_2\bar{y} + a_3\bar{x}\bar{y} + a_4\bar{y}^2 + a_5\bar{x}^2\bar{y} + a_6\bar{x}\bar{y}^2 + a_7\bar{x}^2\bar{y}^2 \\ a_8\bar{x}\bar{y} + a_9\bar{x}^2 + a_{10}\bar{x}^2\bar{y} + a_{11}\bar{x}\bar{y}^2 + a_{12}\bar{x}^2\bar{y}^2 \end{array} \right], \qquad (2.41)$$

to model displacements in the image plane, and second the out-of-plane distortion

$$\mathbf{\Delta}_{op} = \left[\begin{array}{c} a_{13}s_x(\bar{x}^2 - \bar{y}^2) + a_{14}s_x\bar{x}^2\bar{y}^2 + a_{15}s_x(\bar{x}^4 - \bar{y}^4) \\ a_{13}s_y(\bar{x}^2 - \bar{y}^2) + a_{14}s_y\bar{x}^2\bar{y}^2 + a_{15}s_y(\bar{x}^4 - \bar{y}^4) \end{array} \right], \qquad (2.42)$$

$$s_x = -\frac{\bar{x}}{f_m}, \qquad s_y = -\frac{\bar{y}}{f_m}$$

to model sensor plane unflatness. Theoretically, all distortion components can be added to form a very general distortion model. This approach is not feasible in practice, because of redundancy in the parameters and lack of observability. Exact values of the distortion parameters are typically identified in a calibration process, by minimizing the observation error of known 3D reference features. This is only possible if the parameters are sufficiently constrained by the error function, so a minimal set of parameters is desired which describes the distortion at hand. Although the presented distortion models have been originally designed for photogrammetric, analog film cameras, they have been adopted to digital pixel cameras with almost no change.

Radial lens distortion is reportedly the most dominant factor [30], but also behaves nicely during calibration, regarding robustness and repeatability. Usually one or two parameters are sufficient, three radial parameters are only required for wide angle lenses and high accuracy applications. It is a known fact that radial distortion varies with focusing and object depth [34] [31] [10]. With fixed focus setting, only the change of distortion with depth is of interest and a significant influence has been reported for object depths below thirty times the focal length.

Decentering distortion is small compared to the radial component, rarely exceeding one pixel. There is also an insignificant variation with focusing, which can be ignored [30]. Regarding calibration, decentering is highly correlated with the principal point offset, so for low to medium accuracy requirements it should be ignored for sake of robustness.

Out-of-plane distortion effects are due to focal plane unflatness and are a limiting factor for measurement accuracy [32]. Bending of the focal plane may be due to manufacturing defects, or due to elastic deformation during mounting (Figure 2.9). The amount of
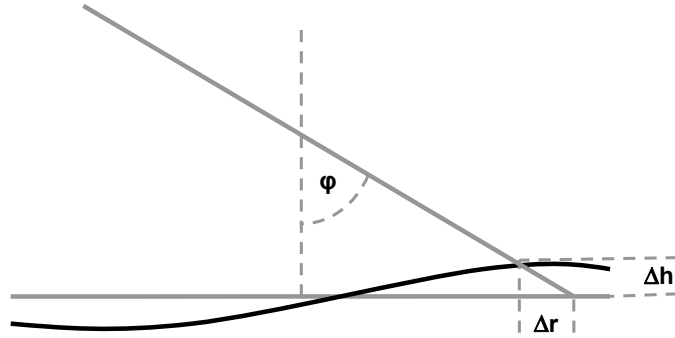
Figure 2.9: Image plane deformation. An uneven sensor plane causes a radial displacement Δr as a function of the ray incidence angle ϕ and height displacement Δr.

distortion for a single pixel is a function of the incidence angle. Clearly, the smaller the incidence angle between sensor plane and imaging ray, the larger the displacement on a deformed sensor plane, so the distortion is a function of the image radius. The problematic thing about this type of distortion is that it is hard to model and calibrate without explicit measurement of sensor plane topology. Explicit measurement of surface unflatness, like it is possible with analog film, is also impractical for digital cameras, since one would have to measure on the mounted sensor plane inside the camera housing. Measurements on the physical chip surface have shown a peak height difference of 1.7μm over a sensor area of 14×9mm^2 [30]. Assuming a relative height difference of 1.7μm between image center and the sensor border, and assuming further a wide angle lens where the outermost imaging rays have an incidence angle of $45°$, a distortion of 1.7μm or 0.12pixels would be the consequence, which is significant only in high accuracy measurements.

In-plane distortion is of lesser concern when using digital pixel sensors. The pixel grid layout on the semiconductor typically is accurate in the nanometer scale and scaling effects due to asynchronous clocks or line jitter, as observed in [7], are obsolete due to on-board A/D conversion. Even if such effects are prominent, they typically manifest as anisotropic pixel scaling and shear, which can be modeled by reducing (2.41) to

$$\mathbf{\Delta}_{ip} = \begin{bmatrix} a_1\bar{x} + a_2\bar{y} \\ 0 \end{bmatrix}. \tag{2.43}$$

The remaining coefficients are redundant to d_x, d_y and s_i in (2.34).

Although Brown's photogrammetric distortion model has been adopted by the computer vision community as the gold standard also for CCD and complementary metal oxide semiconductor (CMOS) cameras, there exists some confusion on how to apply the correction

function (2.38). Denoting the correction mapping $\mathbf{x}' = \mathcal{F}(\mathbf{x})$ as

$$\mathcal{F}(\mathbf{x}) = \mathbf{x} + \Delta_{rad}(\mathbf{x}, \kappa_1, ..., \kappa_n) + \Delta_{tan}(\mathbf{x}, p_1, ..., p_n) + \Delta_{ip}(\mathbf{x}, a_1, ..., a_{12}) + \Delta_{op}(\mathbf{x}, a_{13}, ..., a_{15}),$$
(2.44)

photogrammetrists typically apply the function in metric image space, with different subsets of distortion parameters, so (2.34) is modified to

$$\mathcal{F}(\mathbf{K}_i^{-1}\mathbf{x}) = \mathbf{K}_m\mathbf{x}_c.$$
(2.45)

Beyer [7] and Fraser [30] applied three radial parameters and two decentering parameters. The intrinsic parameters contained focal length, principal point, skew and a scale factor in x-direction. Tsai [102] modeled a `CCD` camera using two radial distortion components and fixed skew, whereas Heikkila [45] applied the opposite mapping

$$\mathbf{x} = \mathbf{K}_i\mathcal{F}(\mathbf{K}_m\mathbf{x_c}),$$
(2.46)

taking two radial and two decentering distortion components into account. Zhang [116] also applied the reversed distortion mapping, but in the canonical image space:

$$\mathbf{x} = \mathbf{K}_i\mathbf{K}_m\mathcal{F}(\mathbf{x_c}).$$
(2.47)

The direction of the correction mapping does not make much of a difference in accuracy. The main problem lies in its nonlinearity. It is hard to find the inverse of the correction mapping, so distortion parameters based on different models are practically incompatible.

In this work, the correction mapping from (2.46) is used, mainly for historical reasons and compatibility to existing software components.

2.4   Kinematic Manipulator Models

Industrial robot arms are modeled as serial kinematic chains. These consist of a number of joints with one degree of motion (`DOM`) each, which are connected by rigid links, as sketched in Figure 2.10. The base `CF` $\mathbf{C}_b$ stands at the beginning of the kinematic chain and is rigidly attached to the ground, whereas the hand `CF` $\mathbf{C}_h$ is moved by manipulating the joint excitations. The pose of $\mathbf{C}_h$ relative to $\mathbf{C}_b$ is given by a Euclidean point transformation

(a)   (b)
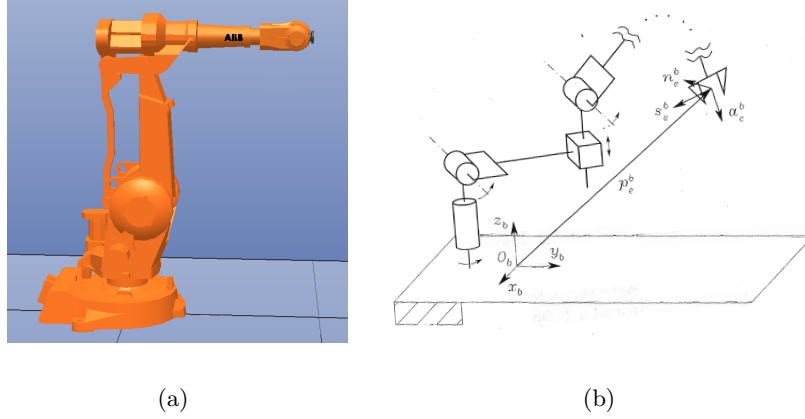
Figure 2.10: Serial kinematic chain. An industrial manipulator (example shown in (a)), is geometrically modeled by a series of joints, connected by rigid links (b) [95].

$\mathbf{H}_{h\to b,b}$ [i] with six DOF. Clearly, $\mathbf{H}_{h\to b}$ is constrained by the geometry of the links and parameterized by the position of the joints:

$$\mathbf{H}_{h\to b} = \mathcal{F}(\mathbf{p}, \mathbf{s}). \tag{2.48}$$

Vector $\mathbf{p} = [p_1, ..., p_m]$ contains the joint positions, whereas vector $\mathbf{s}$ contains the set of constant parameters used to describe link geometry. During operation only $\mathbf{p}$ changes, so $\mathcal{F} : \mathbb{R}^m \to \mathbb{R}^6$ constitutes a nonlinear mapping from m joint positions to the six parameters of $\mathbf{H}_{h\to b}$.

One central problem is parametrization of the links. For the purpose of calibration, the model should have the following properties:

- easy construction from robot blueprints,

- singularity-free,

- unambiguous, i.e. minimal parametrization,

- complete, i.e. can describe all possible manipulator structures.

Considering only $\mathbf{p}$ as the set of free parameters, these requirements are relatively easy met, except freedom of singularities. During calibration of a robotic manipulator, however,

---

[i]$\mathbf{H}_{a\to b,c}$ defines a Euclidean transformation from CF a to b, given in reference CF c. For convenience, if $c = a$, the reference frame is omitted.

p as well as **s** are free parameters. In this section, the three most popular manipulator models are reviewed and their properties regarding robot calibration discussed.

2.4.1 Denavit Hartenberg Notation
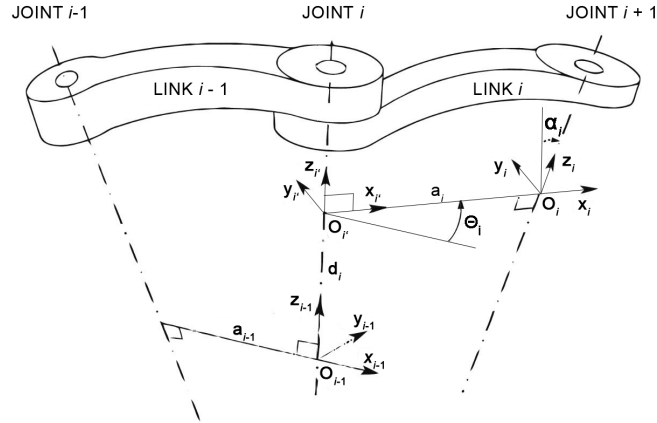
Figure 2.11: `DH` parameters. A set of four parameters and constraints on the joint `CFs` comprise the `DH` model.

`DH` parameters [24] are among the oldest and most popular methods for kinematic modeling. A total of four parameters are used to describe the Euclidean transformation between consecutive joints. Additional constraints on the `CF` pose for each joint account for the remaining two parameters. The `DH` model for one revolute joint is illustrated in Figure 2.11. Each joint comes with a unique axis, which is the axis of rotation for a revolute (rotational), or the axis of translation for a prismatic (translational) joint. As long as consecutive axes are not parallel, the line of shortest distance between them, also known as common perpendicular, is uniquely defined. The origin of the i-th joint `CF` $\mathbf{C}_i$ is constrained to lie at the intersection point of the i-th joint axis and the common perpendicular between joint axes i and $i + 1$. The axes directions are constrained such that $\mathbf{x}_i$ looks in direction of the common perpendicular and $\mathbf{z}_i$ is aligned with the joint axis. Rotation direction of the joint determines whether the z-axis looks in positive or negative direction along the joint axis.

Hereby, the relative pose of two consecutive joint `CF` is determined by four parameters:

1. The shortest distance between consecutive joint axes: a.

2. The angle between consecutive joint axes around the common perpendicular: α.

3. The perpendicular distance d between the previous common perpendicular and the coordinate origin.

4. The angle of rotation θ around the joint axis ($= \mathbf{z}_i$).

The rigid transformation from joint CF $i+1$ to frame i is consequently defined by two rotations and two translations:

$$
\mathbf{H}_{i+1 \rightarrow i} = \begin{pmatrix} \cos\theta_{i+1} & -\sin\theta_{i+1} & 0 & 0 \\ \sin\theta_{i+1} & \cos\theta_{i+1} & 0 & 0 \\ 0 & 0 & 1 & d_{i+1} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & a_i \\ 0 & \cos\alpha_i & -\sin\alpha_i & 0 \\ 0 & \sin\alpha_i & \cos\alpha_i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \qquad (2.49)
$$

or written in different notation

$$
\mathbf{H}_{i+1 \rightarrow i} = \mathrm{Trans}(0, 0, d_{i+1}) \mathrm{Rot}(\mathbf{z}_{i+1}, \theta_{i+1}) \mathrm{Trans}(a_i, 0, 0) \mathrm{Rot}(\mathbf{x}_i, \alpha_i), \qquad (2.50)
$$

where $\mathrm{Rot}(\mathbf{a}, b)$ denotes a rotation of angle b around axis $\mathbf{a}$, and $\mathrm{Trans}(a, b, c)$ denotes a translation by vector $(a, b, c)^T$.

The Euclidean transformation for a serial kinematic chain (i.e. a set of joints, connected by links, without loops) is finally created by concatenating the joint transformations to

$$
\mathbf{H}_{n \rightarrow 0} = \mathbf{H}_{n \rightarrow n-1} ... \mathbf{H}_{i \rightarrow i-1} ... \mathbf{H}_{1 \rightarrow 0}. \qquad (2.51)
$$

The result is a highly nonlinear function of the DH parameters. For revolute joints, the vector of free parameters $\mathbf{p}$ consists of the rotations $(\theta_0, ..., \theta_n)^T$, whereas for prismatic joints, $\mathbf{p} = (d_0, ..., d_n)^T$. Any combination of prismatic and revolute joints is possible. The remaining DH parameters specify link geometry and thereby are contained in $\mathbf{s}$.

The DH model is easily constructed from robot blueprint data, but suffers from singularities in case of parallel joints. Consecutive parallel joints leave the common perpendicular undefined and therefore allow an infinite set of consecutive displacements d_i and d_{i+1}. To overcome this problem, Hayati [43] proposed a different parametrization for parallel and close-to-parallel revolute joints:

$$
\mathbf{H}_{i+1 \rightarrow i} = \mathrm{Rot}(\mathbf{z}_{i+1}, \theta_{i+1}) \mathrm{Trans}(a_i, 0, 0) \mathrm{Rot}(\mathbf{x}_i, \alpha_i) \mathrm{Rot}(\mathbf{y}_i, \beta_i). \qquad (2.52)
$$

The common perpendicular is replaced by a plane, perpendicular to joint axis $i+1$ and

going through the origin of CF $\mathbf{C}i$. The plane intersects both joint axes and hereby defines a connector line, which is not necessarily perpendicular to joint axis i. To account for this angular offset, an additional rotation angle β around $\mathbf{y}_i$ is introduced. The displacement d is zero by definition and therefore omitted.

2.4.2   CPC Model

<div align="center">(a)   (b)</div>
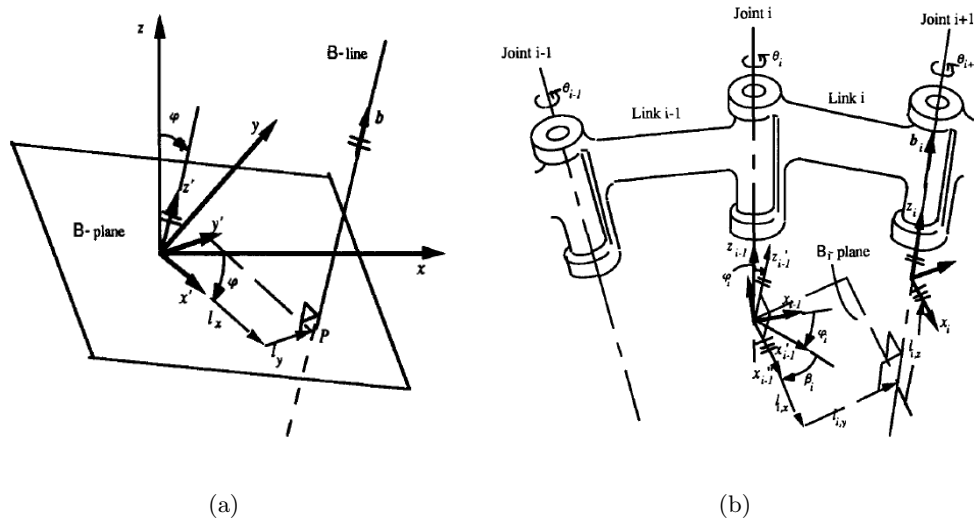
Figure 2.12: The CPC kinematic model [117]. The CPC model is based on a singularity-free line representation (a) to parameterize consecutive joints (b).

Zhuang and Roth introduced the CPC model [117] to improve numerical stability in robot calibration. According to the authors, completeness is the ability of a model to fully describe geometry and motion of a robot. The model has enough parameters to describe the slight geometric deviations caused by manufacturing inaccuracies. Continuity is given if continuous changes in the robot joint axes imply continuous changes in the robot joint parameters.

The central step towards continuity is the singularity free modeling of lines in 3D, which ultimately represent the joint axes. Zhuang adopted the line representation of Roberts [92], which is minimal (four parameters) and singularity-free. To make the model complete, three more parameters are added for each joint. For revolute joints, a total of seven parameters describe one joint, which is a redundant representation, but allows for arbitrary placement of the local joint CF. The only restriction is, that the z-axis must be parallel to the joint axis.

To discuss the particular joint parameters, one starts with a review of the underlying line representation (Figure 2.12(a)). The orientation of a line is given by the direction cosine vector $\mathbf{b} = (b_x, b_y, b_z)^T$. One parameter is redundant, as $b_z = \sqrt{1 - b_x^2 - b_y^2}$, and can be omitted in a minimal parametrization. The direction cosines specify a rotation matrix

$$\mathbf{R} = \text{Rot}(\mathbf{z} \times \mathbf{b}, \arccos(zb)) = \begin{pmatrix} 1 - \frac{b_x^2}{1+b_z} & \frac{-b_x b_y}{1+b_z} & b_X \\ \frac{-b_x b_y}{1+b_z} & 1 - \frac{b_y^2}{1+b_z} & b_y \\ -b_x & -b_y & b_z \end{pmatrix}, \qquad (2.53)$$

which rotates the reference CF $\mathbf{C}$ to a local frame $\mathbf{C}'$, where the $\mathbf{z}'$-axis is parallel to the line direction. Consequently the xy-plane π'_{xy} is perpendicular to it. The line position is specified by the coordinates $(l_x, l_y)^T$ of the intersection point of line and local xy-plane.

The four line parameters form the basis of the CPC parametrization. The only constraints on the link CFs are, that the z-axes must be parallel to the joint axes and that the CFs are right-handed. Based on the CF of one joint $\mathbf{C}_i$, the following CF $\mathbf{C}_{i+1}$ is specified by its joint axis (four parameters), and, in case of a revolute joint, by a rotation Θ around the joint axis. The origin of $\mathbf{C}_{i+1}$ lies at $(l_x, l_y, 0)^T$, relative to the local CF $\mathbf{C}'_i$.

To allow for arbitrary positioning of CFs, e.g. for modeling WCF, base and tool, two more parameters are added, which are zero for link frames. Rotation β_i around the z-axis allows to set the joint "zero" position, translation l_z allows arbitrary positioning along the joint axis.

2.4.3 Robinson's Notation

The kinematic modeling convention by Robinson [93], shown in Figure 2.13, incorporates six parameters for each link. A homogeneous transformation from joint i to $i+1$ is defined by parameter vector $\mathbf{p}_i = (p_1, p_2, ..., p_6)^T$ in the following way:

$$\mathbf{H}_{i \to (i-1)} = \text{Rot}(p_1, z)\text{Trans}(p_2, z)\text{Trans}(p_3, x)\text{Rot}(p_4, x)\text{Rot}(p_5, y)\text{Rot}(p_6, z). \qquad (2.54)$$

Six elementary transformations allow to form a complete and continuous model, but it is not minimal. During kinematic calibration, minimality is achieved by fixing redundant parameters and optimizing only a subset.
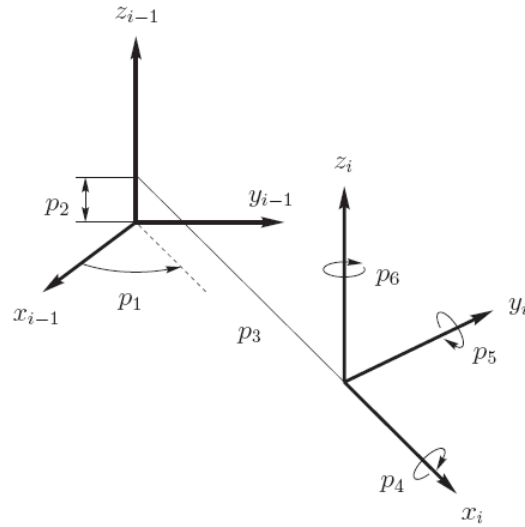
Figure 2.13: Robinson's kinematic model convention. A particularly simple joint parametrization was introduced by Robinson et al. [93].

2.5   Parameter Estimation

Photogrammetric measurement problems are traditionally ill-posed. Based on observations in image space and a functional relation, one seeks to estimate the measurement result. The exact parametrization of this functional is unknown, so one faces the problem of estimating the model parameters and the measurement result, either simultaneously, which is also termed bundle adjustment, or parameters only, which is termed calibration.

Camera calibration is a practical example in this context. A set of reference points with known world coordinates (Figure 2.14(a)) is observed by the camera. Features are measured in image space, which correspond to the reference points (Figure 2.14(b)). The functional relation between reference and measurements is given by central perspective projection, the unknowns are the camera intrinsic and extrinsic parameters (Figure 2.14(b)), which are identified in a parameter estimation procedure.

More generally, a measurement vector $\mathbf{x}$ is given in measurement space $\mathbb{R}^m$, and one searches a parameter vector $\mathbf{p}$, given in parameter space $\mathbb{R}^n$. The values of $\mathbf{p}$ are constrained by a functional relation $\mathcal{F} : \mathbb{R}^n \to \mathbb{R}^m$, such that $\mathcal{F}(\mathbf{p}) = \mathbf{x}$. The set of all measurement vectors $\mathbf{x} = \mathcal{F}(\mathbf{p})$ forms a subspace in $\mathbb{R}^m$ where all $\mathbf{x}$ satisfy the model. In most practical cases, a measurement vector will be corrupted by noise, so in an estimation procedure the goal is to find a vector $\hat{\mathbf{x}}$ which satisfies the model exactly and is as close to $\mathbf{x}$ as possible. Closeness is hereby defined as minimal distance with respect to some error

metric in measurement space.

Parameter estimation is governed by four problems: choosing the mathematical model, choosing the error metric to optimize, choosing the method for optimization and efficient implementation.

Providing an optimal solution to the wrong problem is of limited value, so the correct choice of the model to estimate forms the basis for a successful parameter estimation task. In most geometric estimation problems, the model is closely related to the real world (e.g. central perspective model for a real camera) and hence relatively easy to derive. Important issues in choosing a valid model are the number of parameters, observability of parameters and the degree of redundancy during estimation. The model parameters should be sufficiently general to describe all systematic variations, but it has to be specific enough to prevent over-fitting to noisy measurements. A practical example is shown in Figure 2.15, where the regression curve to a set of noisy measurements has to be found. The measured points are approximately collinear, except for the upper part, where a slight curvature is noticeable. Based on knowledge about the physical nature of these measurements, one has to decide whether this curvature is due to a systematic measurement error (e.g. saturation effects or similar), or whether it is part of the model. In the first case, a linear model would be preferred, coupled with a robust method to handle systematic outliers, while in the second case, using a more general model, e.g. a higher order polynomial, would be the better choice.

Observability of the parameters is another issue and is related to the nature of possible measurements. For example, to estimate a general 3D regression plane through a set of measured points, the points should not be collinear. The reachable degree of redundancy should also be taken in to account. A too general model coupled with sparse measurements easily leads to over-fitting of the data.

Error metrics and estimation methods differ in their physical motivation, numerical properties and computational complexity during estimation. In the following sections popular error metrics for geometric parameter estimation are reviewed, using conic fitting as a practical example. Following the argumentation of [114] and [42], conic fitting is sufficiently complex to demonstrate the properties of estimation methods, but still intuitive enough to visualize.

(a)

(b)

(c)
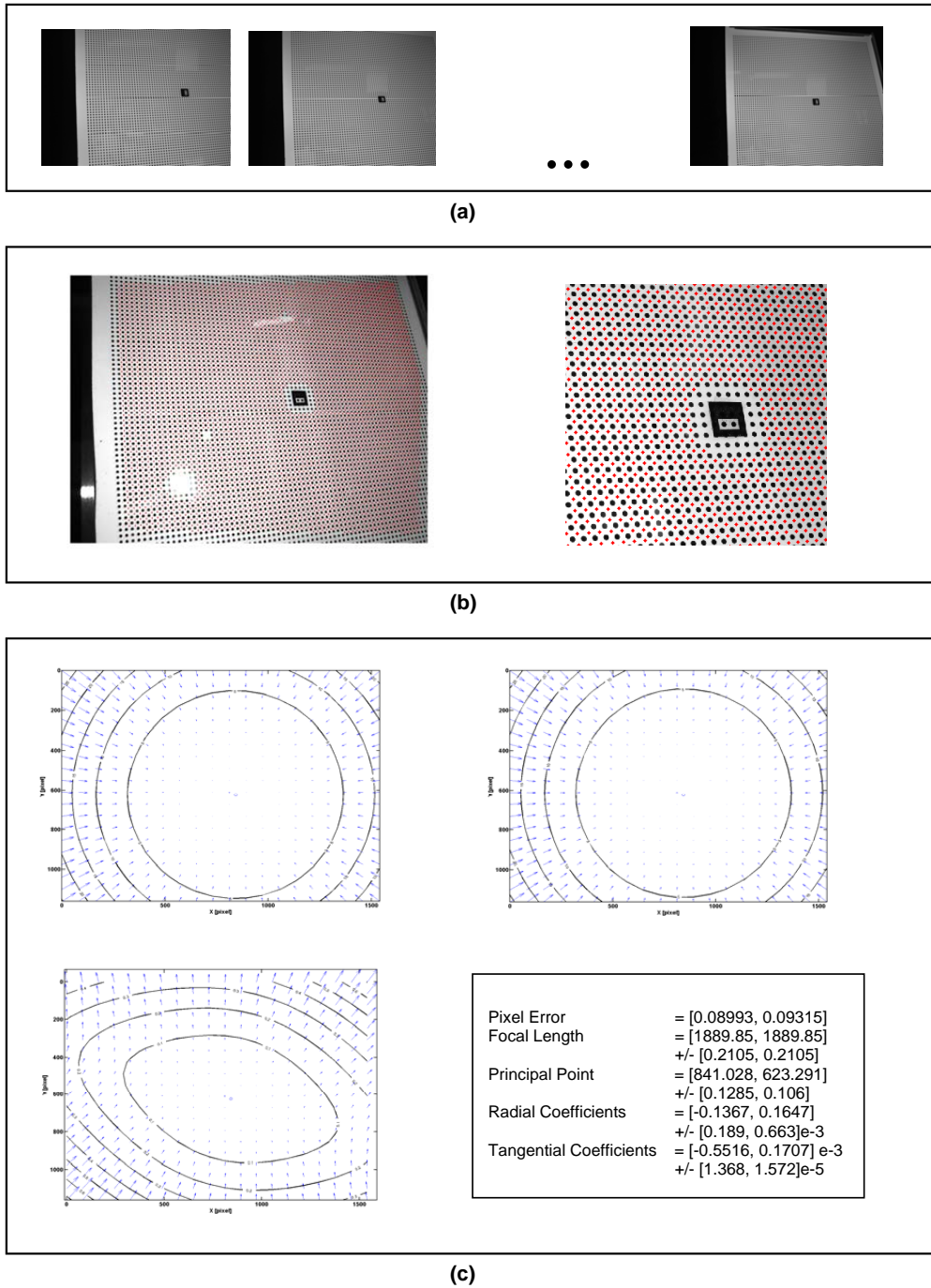
Figure 2.14: Parameter estimation example. Monocular camera calibration is performed by observing a reference target from a number of views (a). Measured image features (b) are related to the reference by a function $\mathcal{F}$. A number of model parameters (c) specify $\mathcal{F}$. These are determined in an estimation procedure.
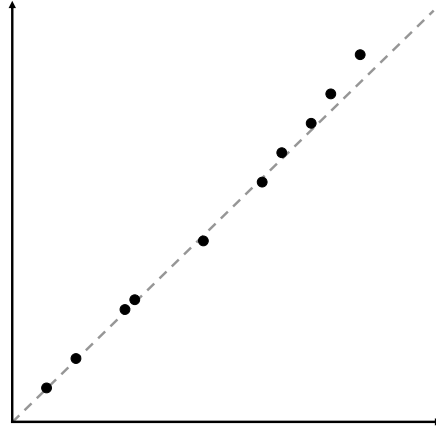
Figure 2.15: Line fit to noisy data. A set of measured points lies approximately on a line. In the upper right region, a systematic deviation can be noticed.

2.5.1   Error Metrics

Choice of the error metric ultimately determines the function to minimize. Considering a measurement vector $\mathbf{x}$ and an estimated vector $\hat{\mathbf{x}} = \mathcal{F}(\mathbf{p})$, one seeks to minimize the distance between them. The distance metric can be defined in several ways. One intuitive option would be to choose Euclidean distance. Another option would be the Mahalanobis distance, leading to a maximum likelihood estimation (`MLE`) of $\mathbf{p}$. The cross-product of vectors leads to a particularly simple, algebraic error metric. In homogeneous notation, this metric can also be used to express point-point distances by taking the cross product of homogeneous vectors.

In the context of conic fitting, measurements are given by a set of points $\mathbf{x} = \{\mathbf{x}_i\} = \{(x_{1i}, x_{2i}, x_{3i})^T\}$. The underlying model is a second order curve, as given in (2.3):

$$\mathbf{x}^T \mathbf{C} \mathbf{x} = 0, \qquad \mathbf{C} = \begin{bmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{bmatrix}, \tag{2.55}$$

with parameter vector $\mathbf{p} = (a, b, c, d, e, f)^T$. $\mathbf{p}$ is defined up to scale and has five `DOF`. Given five point measurements, it is possible to find an exact solution. For more than five points, measurement noise will prevent finding an exact solution and one will seek an approximation.

2.5.1.1   Algebraic Error

Directly substituting a point measurement into (2.55) leads to a particularly simple error metric which is linear in $\mathbf{p}$. This algebraic error measure however, has no geometric justification, and, although easy to minimize, causes a number of problems, like bad numerical conditioning and dependency on the underlying CF.

Stacking the measurements into a measurement matrix $\mathbf{A}$ leads to a homogeneous linear equation system of the form $\mathbf{Ax} = \mathbf{0}$:

$$
\begin{bmatrix}
x_{11}^2 & x_{11}x_{21} & x_{21}^2 & x_{11}x_{31} & x_{21}x_{31} & x_{31}^2 \\
 & & \vdots & & & \\
x_{1n}^2 & x_{1n}x_{2n} & x_{2n}^2 & x_{1n}x_{3n} & x_{2n}x_{3n} & x_{3n}^2
\end{bmatrix}_{n\times 6} \mathbf{p} = \mathbf{0}. \tag{2.56}
$$

Matrix $\mathbf{A}$ must be of rank five to give a non-trivial and non-degenerate solution. To obtain a non-trivial solution, the solution space is constrained, e.g. by providing a minimal parametrization (setting $f = 1$), or by constraining the norm of $\mathbf{p}$. The resulting algebraic error metric is $\mathcal{F}(\mathbf{p}) = |\mathbf{Ap}|$ under the constraint $|\mathbf{p}| = 1$.

One central problem of the algebraic error is its dependency on the underlying CF. To demonstrate this, the problem is transferred to a different CF by an arbitrary similarity transformation $\mathbf{H}$:

$$
\hat{\mathcal{F}}(\hat{\mathbf{p}}) = |\hat{\mathbf{x}}^T \hat{\mathbf{C}} \hat{\mathbf{x}}|, \text{ under } |\hat{\mathbf{p}}| = 1, \tag{2.57}
$$

where $\hat{\mathbf{x}} = \mathbf{Hx}$ and $\hat{\mathbf{C}} = \mathbf{H}^{-T}\mathbf{CH}^{-1}$. The values of $\mathcal{F}$ and $\hat{\mathcal{F}}$ are identical except for scale, as one easily computes:

$$
\hat{\mathcal{F}}(\hat{\mathbf{p}}) = |\hat{\mathbf{x}}^T \hat{\mathbf{C}} \hat{\mathbf{x}}| = |\mathbf{x}^T \mathbf{H}^T \mathbf{H}^{-T} \mathbf{CH}^{-1} \mathbf{Hx}| = |\mathbf{x}^T \mathbf{Cx}| = \mathcal{F}(\mathbf{p}). \tag{2.58}
$$

While the error vector remains identical up to scale under arbitrary homographies $\mathbf{H}$, the norms of $\mathbf{p}$ and $\hat{\mathbf{p}}$ are not simply related, even for similarity transformations, giving rise to different solutions. As a consequence, selection of the underlying CF is not arbitrary and some sort of normalization is required to bring the optimization problem into an 'optimal´ frame.

The algebraic distance between point and a normalized ellipse (center at origin, axes aligned with CF) can be rewritten to:

$$
\mathcal{F}(\mathbf{p}) = ax_1^2 + cx_2^2 + fx_3^2 = -f(d_x^2/c_x^2 - 1), \tag{2.59}
$$
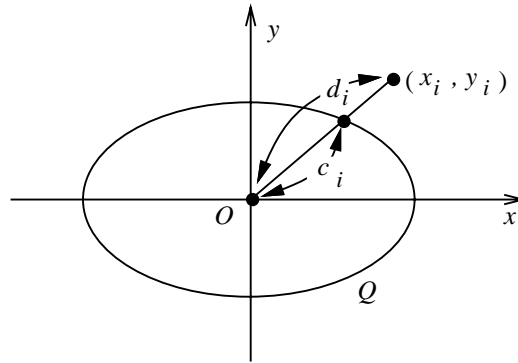
Figure 2.16: Algebraic distance of point and conic.

where d_x is the distance from point $\mathbf{x}$ to the ellipse center and c_x is the distance from intersection point $\mathbf{x}_c$ to the ellipse center, as shown in Figure 2.16. Clearly, for the same orthogonal distance from the ellipse, the ratio d_x/c_x will be larger in regions of small curvature, so measurements there will be weighted stronger than high-curvature points, a problem which is also addressed as high curvature bias.

To overcome the problem of high curvature bias, the cost function can be weighted with its gradient, as proposed by several authors [19] [68] [35] [115]:

$$\mathcal{F}'(\mathbf{p}) = \mathcal{F}(\mathbf{p})/\nabla\mathcal{F}(\mathbf{p}), \tag{2.60}$$

which leads to a nonlinear cost function and a good approximation of the orthogonal distance metric.

As stated before, algebraic error is not invariant to similarity transformations of the underlying CF, so the question arises, which frame is the 'best´ for estimation. Taking a closer look at (2.56), one notices that the coefficients of $\mathbf{A}$ are multilinear in the measured point coordinates. Since pixel-coordinates are roughly in the range of 500, and the homogeneous coordinate is usually one, the columns of $\mathbf{A}$ will differ by several orders of magnitude. Hartley investigated in [41] the effects of isotropic and anisotropic normalization. A normalizing transformation, which translates the CF origin to the centroid of all measured points, and scaling, which sets the average distance from the origin to $\sqrt{2}$ brought significantly better estimation results. Mühlich and Mester [80] proposed normalization based on statistical analysis of the error model, where inherently noise-free parts of $\mathbf{A}$, i.e. parts depending on the homogeneous coordinate only, should receive much higher weight than noise-corrupted parts. Slightly better results than anisotropic scaling have

been reported.

2.5.1.2   Euclidean Error
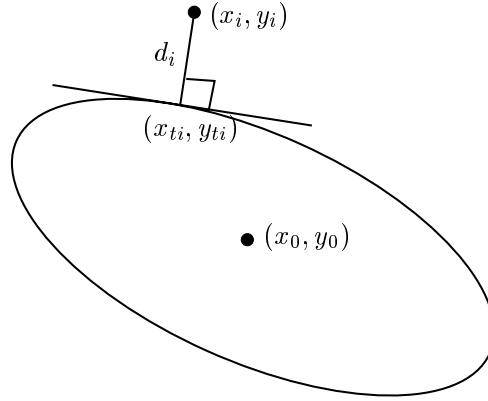
Figure 2.17: Orthogonal distance of point and conic.

From geometric intuition, a good error metric is Euclidean distance. In the conic fitting example, this would be the orthogonal distance between conic and measured point. Least-squares fitting of a conic requires minimization of the error function:

$$\mathcal{F}(\mathbf{p}) = \sum_i d_\perp(\mathbf{x}_i, \mathbf{C})^2 = \sum_i d(\mathbf{x}_i, \mathbf{x}_{ti})^2. \tag{2.61}$$

Orthogonal distance $d(\mathbf{x}_i, \mathbf{C})$ is equivalent to the Euclidean distance between measured point $\mathbf{x}_i$ and the closest point on the conic $\mathbf{x}_{ti}$ (see Figure 2.17). The line through $\mathbf{x}_i$ and $\mathbf{x}_{ti}$ must be orthogonal to the tangent of $\mathbf{C}$ in $\mathbf{x}_{ti}$.

Computation of orthogonal distance is a nonlinear problem and rather complex. A detailed derivation of $d_\perp(\mathbf{x}_i, \mathbf{C})$ for elliptic conics is given in [114]. The algorithm includes finding the roots of a fourth order polynomial to find $\mathbf{x}_{ti}$ and taking the solution with minimum distance to $\mathbf{x}_i$.

Orthogonal distance does not show the problems of algebraic error. It is independent of the underlying CF and does not show curvature bias. One possibility to avoid explicit computation of $d(\mathbf{x}_i, \mathbf{C})$ is to over-parameterize the problem and add the closest tangent points $\mathbf{x}_{ti}$ to the set of parameters. These points are constrained to lie on the conic to estimate and the resulting error metric is:

$$\mathcal{F}(\mathbf{p}) = \sum_i d(\mathbf{x}_i, \mathbf{x}_{ti})^2. \tag{2.62}$$

Since each point $\mathbf{x}_{ti}$ lies on the conic, it adds one free parameter - the position on the conic - to the set of unknowns. The number of parameters to estimate is thereby drastically increased, and requires some care during implementation to solve the problem efficiently.

2.5.1.3   Statistical Error

Parameter estimation is inherently based on noisy measurements, so it is favorable to incorporate knowledge on uncertainty. A standard assumption in least squares estimation is that measurement error follows a Gaussian distribution with zero mean and covariance $\mathbf{\Sigma}$. The probability density function (PDF) is given by

$$P(\mathbf{x}) = (2\pi)^{N-2}\det\left(\mathbf{\Sigma}^{-1}\right)^{1/2}\exp(-(\mathbf{x}-\bar{\mathbf{x}})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\bar{\mathbf{X}})/2), \tag{2.63}$$

where $\bar{\mathbf{x}}$ is the mean vector. In the simpler case of an isotropic Gaussian distribution, covariance $\mathbf{\Sigma}$ is diagonal with identical entries $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ and the PDF is given by

$$P(\mathbf{x}) = (\sqrt{2\pi}\sigma)^{-N}\exp(-\sum_{i=1}^{N}(x_i-\bar{x}_i)^2/(2\sigma^2)), \tag{2.64}$$

where, x_i are the elements of vector $\mathbf{x}$. An important observation is that $P(\mathbf{x})$ is a function of $(\sum_{i=1}^{N}(x_i-\bar{x}_i)^2)^{1/2}$, which is the Euclidean distance between $\mathbf{x}$ and $\bar{\mathbf{x}}$. A joint Gaussian distribution is equivalently a function of $(\mathbf{x}-\bar{\mathbf{x}})^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\bar{\mathbf{X}})$, which leads to the definition of the Mahalanobis distance metric between $\mathbf{x}$ and $\mathbf{y}$ [42]:

$$|\mathbf{x}-\mathbf{y}|_\Sigma = \sqrt{(\mathbf{x}-\mathbf{y})^T\Sigma^{-1}(\mathbf{x}-\mathbf{y})}. \tag{2.65}$$

The Gaussian PDF is given by:

$$P(\mathbf{x}) = (2\pi)^{N-2}\det\left(\mathbf{\Sigma}^{-1}\right)^{1/2}\exp(|\mathbf{x}-\bar{\mathbf{x}}|_\Sigma^2(-1/2)). \tag{2.66}$$

Thus, for a joint Gaussian distribution $P(\mathbf{x})$ is a function of the Mahalanobis distance between $\mathbf{x}$ and $\bar{\mathbf{x}}$.

We now return to the Euclidean error metric (2.61). Assuming that point measurements $\mathbf{x} = \{x_i\}$ are independently and uniformly distributed around the true ellipse points

$\bar{\mathbf{x}} = \{\mathbf{x}_{ti}\}$, the probability of $\mathbf{x}$, given the true points $\bar{\mathbf{x}}$, is

$$P(\mathbf{x}|\bar{\mathbf{x}}) = \prod (2\pi\sigma^2)^{-1} \exp\left(-d(x_i, x_{it})^2/(2\sigma^2)\right). \tag{2.67}$$

Maximizing $P(\mathbf{x}|\bar{\mathbf{x}})$ is equivalent to maximizing $\log P(\mathbf{x}|\bar{\mathbf{x}})$ and hence is equivalent to minimizing

$$\sum_i d(\mathbf{x}_i, \mathbf{x}_{ti})^2. \tag{2.68}$$

Minimal Euclidean distance is equivalent to the MLE of the true values, of course only under the assumption of Gaussian measurement error.

2.5.1.4   Robust Cost Functions

In the last section, Euclidean distance as a cost function was statistically motivated for Gaussian measurement noise with no gross outliers. This error model is reasonable only if explicit blunder detection has been carried out to remove outliers, e.g. by a random sample consensus (RANSAC) procedure. If the cost function itself should tolerate a certain amount of outliers, the Gaussian error model should be replaced by a more robust one. Examples are given in Figure 2.18. To handle outliers implicitly, the tails of robust PDFs are considerably wider than the in Gaussian case, i.e. largely wrong measurements are expected with a comparatively large probability.

To derive the cost function from a probability distribution $p(x)$, in case of independent measurements, one accumulates the probability of a set of measurements as $p(x_1, ..., x_n) = \prod_{i=1}^{n} p(x_i)$. Taking the negative logarithm leads to a cost function which is suitable for further minimization:

$$e(x_1, ..., x_n) = -\log(p(x_1, ..., x_n)) = -\sum_{i=1}^{n} \log(p(x_i)). \tag{2.69}$$

Several robust cost functions have been proposed in literature, some of them are shown in Figure 2.18, according to [42]. In contrast to Gaussian error distribution, which results in a quadratic cost function, robust PDFs attenuate the importance of gross outliers. As such, the cost function is much flatter for large values than the quadratic function.

A typical example is the Blake-Zisserman error distribution, where equally distributed outliers are assumed. Consequently, $p(x)$ is a Gaussian distribution with an added constant:
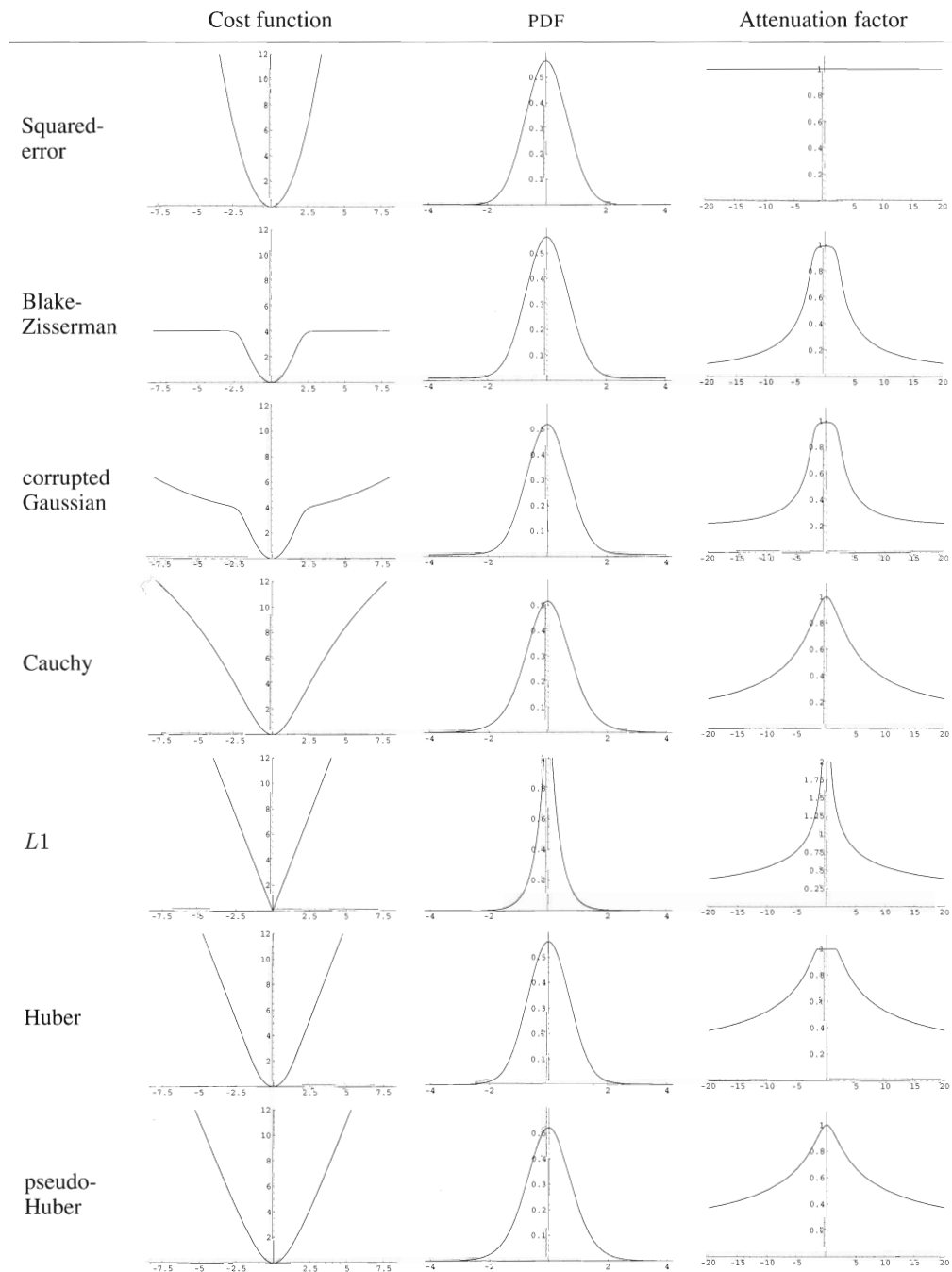
Figure 2.18: Robust cost functions (first column), corresponding PDFs (second column) and attenuation factors (last column) [42].

$$p(x) = \exp(-x^2) + \epsilon, \tag{2.70}$$

with

$$e(x) = -\log(\exp(-x^2) + \epsilon). \tag{2.71}$$

For large values of x, the error is approximately constant, whereas in close vicinity to 0, it has approximately quadratic behavior. It is important to note that (2.70) is not strictly a PDF, because its integral equals ∞.

Another statistically motivated PDF would be a mixture of two Gaussians, one with small variance to model measurement error, and the second with larger variance to cope with gross outliers.

Cost functions like the L_1 norm or total variation

$$e(x) = |x|, \tag{2.72}$$

or the Huber function

$$e(x) = \left\{ \begin{array}{l} x^2 \text{ for } |x| < b \\ 2b|x| - b^2 \text{ for } |x| \geq b \end{array} \right\} \tag{2.73}$$

are heuristic in nature and do not have a statistical or physical motivation, but have the property of being convex, which allows to find a global optimum of the problem.

2.5.2 Minimization

Apart from its statistical and geometrical motivation, least squares optimization has additional advantages, which make it applicable especially to photogrammetric problems. Among them are its flexibility and ease of use. The role of coefficients as model parameters, constants or measurements an easily be interchanged, an important property when estimation procedures need to be tuned to specific problems. Constraints on measurements can incorporated in the estimation process, and a large software basis exists for solving least squares problems. In this section, least squares estimation methods for linear and nonlinear problems are reviewed.

2.5.2.1   Linear Least Squares

Given is a linear equation system

$$\arg\min_{\mathbf{p}} \|\mathcal{F}(\mathbf{p})\|_2^2, \quad \mathcal{F}(\mathbf{p}) = \mathbf{A}\mathbf{p} - \mathbf{b}, \tag{2.74}$$

where matrix $\mathbf{A}_{m \times n}$ and vector $\mathbf{b}$ are specified by the measurement vector $\mathbf{x}$, with $m \geq n$. In case of rank-deficient systems, where $rank\mathbf{A} < m$, the function to minimize becomes $\mathbf{0}$ for an infinite number of possible parameter values, so one is interested in a family of minima, where the number of free parameters $c = n - rank\mathbf{A}$. In case of $rank\mathbf{A} = n$ there exists exactly one parameter vector for which $\mathcal{F}(\mathbf{P}) = 0$. In most practical cases, $rank\mathbf{A} = n$ if $m > n$, as measurement noise prohibits finding exact solutions of overdetermined problems. Here, a least squares solution can be found using singular value decomposition or by means of the normal equations.

Singular value decomposition allows to reformulate the original problem, according to

$$\|\mathbf{A}\mathbf{p} - \mathbf{b}\| = \|\mathbf{U}\mathbf{D}\mathbf{V}^T\mathbf{p} - \mathbf{b}\| = \|\mathbf{D}\mathbf{V}^T\mathbf{P} - \mathbf{U}^T\mathbf{b}\|. \tag{2.75}$$

The last equality is due to the orthogonal columns of $\mathbf{U}$. According to [42] one writes

$$\arg\min_{\mathbf{y}} \left\| \begin{bmatrix} d_1 & & & \\ & \ddots & & \\ & & d_n & \\ & \mathbf{0} & & \\ & \vdots & & \end{bmatrix} \mathbf{y} - \mathbf{b}' \right\| \Leftrightarrow \arg\min_{\mathbf{y}} \left\| \begin{bmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{bmatrix} \mathbf{y} - \mathbf{b}'_n \right\|. \tag{2.76}$$

The vector of parameters is now $\mathbf{y} = \mathbf{V}^T\mathbf{p}$, $\mathbf{b}' = \mathbf{U}^T\mathbf{b}$, and $\mathbf{b}'_n$ consists of the first n elements of $\mathbf{b}'$. Setting the elements of $\mathbf{y}$ to $y_i = b'_i/d_i$ gives the minimal solution to (2.76) in a least squares sense. The minimal solution for $\mathbf{p}$ is calculated using $\mathbf{p} = \mathbf{V}\mathbf{y}$.

A different approach to solving linear least squares problems is based on the normal equations

$$\mathbf{A}^T\mathbf{A}\mathbf{P} = \mathbf{A}^T\mathbf{b}. \tag{2.77}$$

The argumentation is as follows: if $\mathbf{p}$ is the correct solution, $\mathbf{A}\mathbf{p}$ must be the vector in the column space of $\mathbf{A}$, which is closest to $\mathbf{b}$. The residual vector $\mathbf{A}\mathbf{p} - \mathbf{b}$ must be orthogonal to the column space of $\mathbf{A}$, and consequently:

$$\mathbf{A}^T(\mathbf{A}\mathbf{p} - \mathbf{b}) = \mathbf{0}, \tag{2.78}$$

which is equivalent to 2.77. A solution for $\mathbf{p}$ is found by inverting $\mathbf{A}^T\mathbf{A}$, leading to

$$\mathbf{P} = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{b}, \tag{2.79}$$

where $\mathbf{A}^+ = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T$ is the pseudo-inverse of $\mathbf{A}$.

The presented method, also addressed as ordinary least squares estimator, gives an optimal estimate of $\mathbf{p}$, where 'optimal´ means minimum covariance of $\mathbf{p}$ if the coefficients of residual vector $\mathbf{e} = \mathbf{A}\mathbf{p} - \mathbf{b}$ are uncorrelated and have constant variance. If this is not the case, it is favorable to provide a means for weighting residuals according to their covariance, leading to a weighted linear least squares problem of the form

$$\mathbf{A}^T\mathbf{C}\mathbf{A}\mathbf{p} = \mathbf{A}^T\mathbf{C}\mathbf{b}. \tag{2.80}$$

Matrix $\mathbf{C}$ is symmetric and positive definite and applies a weight to the residuals in $\mathbf{e}$. Usually the weighting is proportional to the error covariance, which is probably known a priori. A new inner product is defined by $\mathbf{C}$, as $\mathbf{x} * \mathbf{y} = \mathbf{x}^T\mathbf{C}\mathbf{y}$, and a new vector norm $\|\mathbf{x}\|_C = \sqrt{\mathbf{x}^T\mathbf{C}\mathbf{x}}$. Again one argues that the residual vector must be orthogonal to the column space of $\mathbf{A}$ with respect to the new inner product, leading to

$$\mathbf{A}^T\mathbf{C}(\mathbf{A}\mathbf{P} - \mathbf{b}) = \mathbf{0}, \tag{2.81}$$

which is equivalent to (2.80).

2.5.2.2   Nonlinear Least Squares

The most popular least-squares iteration schemes, namely the Newton and Gauss-Newton method, as well as Levenberg-Marquardt (LM) iteration, will be reviewed in this section.

In many estimation problems, the error function is explicit in $\mathbf{x}$:

$$\arg\min_{\mathbf{p}} \|\mathbf{e}\|, \quad \mathbf{e} = \mathcal{F}(\mathbf{p}) - \mathbf{x}. \tag{2.82}$$

Newton iteration starts from an initial parameter estimate $\mathbf{p}_0$, close to the true value $\bar{\mathbf{p}}$, and iteratively refines the solution, under the assumption that $\mathcal{F}(\mathbf{p})$ is locally linear:

$$\mathcal{F}(\mathbf{p} + \triangle) = \mathcal{F}(\mathbf{p}) + \mathbf{J}\triangle. \tag{2.83}$$

The Jacobian matrix $\mathbf{J}$ represents the approximated linear mapping from $\mathbb{R}^n$ to $\mathbb{R}^m$ as the local gradient $\mathbf{J} = \partial \mathcal{F}/\partial \mathbf{p}$. One seeks to estimate an improved parameter vector $\mathbf{p}_1 = \mathbf{p}_0 + \triangle$ by minimizing $\|\mathcal{F}(\mathbf{p}_1) - \mathbf{x}\|$, which is a linear problem:

$$\mathcal{F}(\mathbf{p}_1) - \mathbf{x} = \mathcal{F}(\mathbf{p}_0) + \mathbf{J}\triangle - \mathbf{x} = \mathbf{e}_0 + \mathbf{J}\triangle. \tag{2.84}$$

Solving

$$\arg \min_{\triangle_{\mathbf{i+1}}} (\mathbf{e}_i + \mathbf{J}_i \triangle_{i+1}) \tag{2.85}$$

using normal equations gives an estimate for $\triangle_{i+1} = -\mathbf{J}_i^+ \mathbf{e}_i$. As such, $P_{i+1} = P_i + \triangle_{i+1}$, and one iteration step is finished. Convergence depends on the problem structure, its convexity, and the choice of the initial parameter estimate.

Applying different weights to measurements is achieved by introducing a covariance matrix $\mathbf{C}$ and solving the linear problem according to (2.80).

More general, Newton iteration seeks to minimize an arbitrary function $\mathcal{G}(\mathbf{p})$, which is scalar valued. An initial guess $\mathbf{p}_0$ of the solution is provided which is close to the actual minimum. The function is approximated locally around $\mathbf{p}_0$ by a Taylor series, which is terminated after the third term

$$\mathcal{G}(\mathbf{p}_0 + \triangle) = \mathcal{G}(\mathbf{p}_0) + \mathcal{G}(\mathbf{p}_0)' \triangle + \triangle^T \mathcal{G}(\mathbf{p}_0)'' \triangle / 2 + ... \tag{2.86}$$

The first and second derivatives of $\mathcal{G}$ with respect to $\mathbf{p}$ are given by gradient $\mathcal{G}'$ and Hessian $\mathcal{G}''$ respectively. A minimum with respect to $\triangle$ is found by setting the derivative to zero:

$$\frac{\partial \mathcal{G}(\mathbf{p}_0 + \triangle)}{\partial \triangle} = \mathcal{G}(\mathbf{p}_0)' + \mathcal{G}(\mathbf{p}_0)'' \triangle = 0. \tag{2.87}$$

Newton iteration consists of calculating $\mathcal{G}'$ and $\mathcal{G}''$ in each step, calculation of $\triangle$ and updating $\mathbf{p}$ as $\mathbf{p}_{i+1} = \mathbf{p}_i + \triangle_i$.

In the special case of least-squares optimization, $\mathcal{G}(\mathbf{p})$ has a special form:

$$\mathcal{G}(\mathbf{p}) = e(\mathbf{p})^T e(\mathbf{p}), \tag{2.88}$$

with $e(\mathbf{p}) = \mathcal{F}(\mathbf{p}) - \mathbf{x}$. The first derivative is given by $\mathcal{G}' = \mathbf{J}^T e$, where $\mathbf{J}$ is the Jacobian of $\mathcal{F}$. The second derivative would be $\mathcal{G}'' = \mathbf{J}^T \mathbf{J} + e''^T e$. If the Hessian of e is hard to compute, one may assume that $\mathcal{F}$ is linear, which leads to $\mathcal{G}'' = \mathbf{J}^T \mathbf{J}$. Substitution into (2.87) gives

$$\frac{\partial \mathcal{G}(\mathbf{p}_0 + \triangle)}{\partial \triangle} = \mathbf{J}^T e + \mathbf{J}^T \mathbf{J} \triangle = 0, \tag{2.89}$$

which is equivalent to the normal equations (2.77). This linear approximation is termed Gauss-Newton iteration. It avoids explicit computation of the Hessian and gives reasonable results, especially close to the actual minimum.

Gradient descent is the third iteration procedure. In each iteration it follows the direction of steepest negative gradient to reach the minimum:

$$\triangle = -\lambda \mathcal{G}' = -\lambda \mathbf{J}^T e, \tag{2.90}$$

where λ controls the step size. Gradient descent generally has slower convergence than Gauss-Newton iteration, due to zig-zagging around the desired optimum, but it may be preferred if far away from the optimum.

LM iteration is a hybrid method which seeks to combine the advantages of gradient descent and Gauss-Newton iteration. The normal equations from (2.89) are replaced by

$$\mathbf{J}^T e + (\mathbf{J}^T \mathbf{J} + \lambda \mathbf{I}) \triangle = 0, \tag{2.91}$$

where scalar λ is a weighting factor, and is updated in each step. Starting from a typical value of 10^{-3}, λ is decreased (e.g. $\lambda_{i+1} = \lambda_i/10$), if the error has decreased. If the error has increased, λ is repeatedly increased (e.g. $\lambda_{i+1} = 10\lambda_i$), until the error decreases again. As a consequence, small λ favors Gauss-Newton like iteration, and large λ favors gradient descent-like iteration.

2.6   Estimation Evaluation

In most parameter estimation problems it is not sufficient to retrieve a parameter estimate. Some information about precision and confidence of the result is also required, which leads to the problem of evaluation.

From intuition it is clear that the quality of an estimated parameter vector $\hat{\mathbf{p}}$, defined in parameter space $\mathbb{R}^m$, depends on the set of noisy measurements $\mathbf{x}$, defined in measurement space $\mathbb{R}^n$. The more measurements are available, and the smaller the measurement error, the better the result will be. Further, configuration of the measurement vector is of interest, i.e. if all model parameters are sufficiently constrained by measurements.

If only the estimated parameter vector $\hat{\mathbf{p}}$, measurement vector $\mathbf{x}$, and the model func-

tion $\mathbf{x} = \mathcal{F}(\mathbf{p}) : \mathbb{R}^m \to \mathbb{R}^n$ are available, one can evaluate the difference between noisy measurements and the estimated model $\mathbf{x} - \mathcal{F}(\hat{\mathbf{p}})$, also termed residual error.

The residual error does not allow to assess model accuracy, but it measures precision, i.e. how well the model fits the data. This type of measure should be used only if no external reference information is available, because a small residual error does not necessarily imply a good model estimate. For example, if a 2D line is estimated from a set of nearly identical points, or from exactly two points, the residual error might be low, but the line estimate is not very good. In the first case, the model will not be sufficiently constrained and has essentially one undetermined parameter. In the second case, the residual will always be exactly zero and model accuracy will depend on the accuracy of the two measured points.

If information on the measurement accuracy of $\mathbf{x}$ is available, the covariance of $\hat{\mathbf{p}}$ can be calculated by backpropagation. Under the assumption that $\mathbf{x} = [x_1, ..., x_n]^T$ is chosen according to a Gaussian distribution with mean at the true value $\bar{\mathbf{x}}$ and covariance Σ, the covariance Σ_p of $\hat{\mathbf{p}}$ can be calculated. This allows for further diagnostics. Considering the problem from above, where a line is estimated from nearly identical points, the resulting variance of one line parameter would be high, indicating that the parameter is not well constrained. Parameter covariance further contains information about parameter correlations. Especially in estimation problems, where the number of necessary model parameters may not be determined in advance, strong correlations are a hint for over-parametrization.

A more concise evaluation of accuracy is possible, if external reference data is available, which is the case in most calibration procedures, and in synthetic experiments. The difference between true values $\bar{\mathbf{x}}$ and model leads to the estimation error $\bar{\mathbf{x}} - \mathcal{F}(\hat{\mathbf{p}})$, which allows to assess estimation accuracy in absolute measures.

In the remainder of this section the above presented concepts are stated more formally. An estimation problem consists of estimating an m-dimensional parameter vector $\mathbf{p}_{m \times 1}$ from n measurements $\mathbf{x}_{m \times 1}$. The measurements follow independent Gaussian distributions with mean at the true values $\bar{\mathbf{x}}$ and identical variance σ^2. The model is given by a functional relation $\mathbf{x} = \mathcal{F}(\mathbf{p})$.

The root mean square (RMS) residual error for a maximum likelihood (ML) estimator is then given by

$$\epsilon_{res} = E[\|\hat{\mathbf{x}} - \mathbf{x}\|^2/n]^{1/2} = \sigma(1 - m/n)^{1/2}, \tag{2.92}$$

where $\hat{\mathbf{x}} = \mathcal{F}(\hat{\mathbf{p}})$.

The RMS estimation error for a ML estimator is

$$\epsilon_{est} = E[\|\hat{\mathbf{x}} - \bar{\mathbf{x}}\|^2/n]^{1/2} = \sigma(m/n)^{1/2}. \tag{2.93}$$

A geometric proof is sketched in [42]. If the measurement error is not independent, $\|A\|$ may be replaced by the Mahalanobis distance $\|A\|_\Sigma$.

With increasing number of measurements, the residual error also increases and asymptotically reaches the measurement σ. At the same time, the parameter estimate will become better, and the estimation error decreases. The expected average error of the MLE forms the gold standard for estimation algorithms, because it is the best that may be achieved without further prior information.

While the above error measures capture the dependency of a parameter estimate on the number of measurements, the covariance matrix captures dependencies of the measurements. In the following, rules for forward and backward propagation of covariances under affine and nonlinear mappings are given. These rules are taken without proof from [42].

Definition 1 (Affine forward propagation.) *Let* $\mathbf{v}$ *be a random vector in* $\mathbb{R}^M$ *with mean* $\bar{\mathbf{v}}$ *and covariance matrix* $\mathbf{\Sigma}$, *and suppose that* $\mathcal{F} : \mathbb{R}^M \to \mathbb{R}^N$ *is an affine mapping defined by* $\mathcal{F}(\mathbf{v}) = \mathcal{F}(\bar{\mathbf{v}}) + \mathbf{A}(\mathbf{v} - \bar{\mathbf{v}})$. *Then* $\mathcal{F}(\mathbf{v})$ *is a random variable with mean* $\mathcal{F}(\bar{\mathbf{v}})$ *and covariance matrix* $\mathbf{A}\mathbf{\Sigma}\mathbf{A}^T$.

Definition 2 (Non-linear forward propagation.) *Let* $\mathbf{v}$ *be a random vector in* $\mathbb{R}^M$ *with mean* $\bar{\mathbf{v}}$ *and covariance matrix* $\mathbf{\Sigma}$, *and let* $\mathcal{F} : \mathbb{R}^M \to \mathbb{R}^N$ *be differentiable in a neighborhood of* $\bar{\mathbf{v}}$. *Then up to a first order approximation,* $\mathcal{F}(\mathbf{v})$ *is a random variable with mean* $\mathcal{F}(\bar{\mathbf{v}})$ *and covariance* $\mathbf{J}\mathbf{\Sigma}\mathbf{J}^T$, *where* $\mathbf{J}$ *is the Jacobian matrix of* $\mathcal{F}$, *evaluated at* $\bar{\mathbf{v}}$.

Definition 3 (Affine backward propagation.) *Let* $\mathcal{F} : \mathbb{R}^M \to \mathbb{R}^N$ *be an affine mapping of the form* $\mathcal{F}(\mathbf{P}) = \mathcal{F}(\bar{\mathbf{P}}) + J(\mathbf{P} - \bar{\mathbf{P}})$, *where* $\mathbf{J}$ *has rank* M. *Let* $\mathbf{X}$ *be a random variable in* $\mathbb{R}^N$ *with mean* $\bar{\mathbf{X}} = \mathcal{F}(\bar{\mathbf{P}})$ *and covariance matrix* Σ. *Let* $\mathcal{F}^{-1} \circ \eta : \mathbb{R}^N \to \mathbb{R}^M$ *be a mapping that maps a measurement* $\mathbf{X}$ *to a set of parameters corresponding to the* MLE $\hat{\mathbf{X}}$. *Then* $\hat{\mathbf{P}} = \mathcal{F}^{-1} \circ \eta(\mathbf{X})$ *is a random variable with mean* $\bar{\mathbf{P}}$ *and covariance matrix*

$$\mathbf{\Sigma}_{\mathbf{P}} = (\mathbf{J}^T \mathbf{\Sigma}_{\mathbf{X}}^{-1} \mathbf{J})^{-1}. \tag{2.94}$$

Definition 4 (Non-linear backward propagation.) *Let $\mathcal{F} : \mathbb{R}^M \to \mathbb{R}^N$ be a differentiable mapping and let $\mathbf{J}$ be its Jacobian matrix evaluated at a point $\bar{\mathbf{P}}$. Suppose that $\mathbf{J}$ has rank M. Then $\mathcal{F}$ is one-to-one in a neighborhood of $\bar{\mathbf{P}}$. Let $\mathbf{X}$ be a random variable in $\mathbb{R}^N$ with mean $\bar{\mathbf{X}} = f(\bar{\mathbf{P}})$ and covariance matrix $\mathbf{\Sigma_X}$. Let $\mathcal{F}^{-1} \circ \eta : \mathbb{R}^N \to \mathbb{R}^M$ be the mapping that maps a measurement $\mathbf{X}$ to a set of parameters corresponding to the* MLE *$\hat{\mathbf{X}}$. Then to first order, $\hat{\mathbf{P}} = \mathcal{F}^{-1} \circ \eta(\mathbf{X})$ is a random variable with mean $\bar{\mathbf{P}}$ and covariance matrix $(\mathbf{J}^T \mathbf{\Sigma_X}^{-1} \mathbf{J})^{-1}$.*

In the non-linear case these rules allow to propagate covariance matrices forward and backward, under the assumption that $\mathbf{J}$ is invertible and that the MLE has been computed. Another method of evaluating a parameter estimate is Monte Carlo simulation, where several measurements are drawn from an assumed noise distribution, or are effectively measured. The parameter covariance is then calculated statistically from several trials. This method is computationally expensive and only feasible for small estimation problems, but it is independent of the underlying measurement noise distribution.

2.7 Conclusion

In this chapter, the theoretical foundation for this work has been laid out. The principles of projective geometry are used to develop the sensor system proposed in Chapter 3. The multi-view geometrical layout and structure estimation method as proposed in Chapter 5 also utilize projective geometry and especially the linear and nonlinear central perspective projection model. The theory of rigid motions plays an important role in kinematic robot modeling and calibration (Chapter 7).

To summarize, the different fields of application used in this work, ranging from computer vision over photogrammetry to robotics and calibration, are based on a common set of theoretical foundations. Differences in notation make it often difficult to use this common basis in a single framework.

Chapter 3

System Concept

Contents

3.1 Introduction

Industrial quality inspection and process control require the ability to conduct a large variety of measurement tasks on machined parts. A typical example is the inspection of a freshly produced motor block for compliance with given quality constraints. The motor block would be moved to an inspection station, where numerous object parameters like borehole diameters, planarity of flange surfaces and relative position of parts are measured. Furthermore, recognition tasks like reading data object codes, barcodes or engraved characters would be performed. The result would be a decision whether the object fulfills predefined quality requirements or not.

Currently, such inspection problems are solved by combining a number of specialized sensor systems in one measurement station. Each sensor solves one measurement problem at a single location and the results are combined by a central controller to form a decision on object quality.

Implementation of such a measurement cell is difficult. Automation companies employ experts in mechanical engineering, electrical engineering, robotics and measurement

instrumentation to build automated systems dedicated to a given task. As a result there is an economic need to simplify construction and task planning. It would be desirable to have a single measurement system, capable of performing a number of various measurement tasks at different positions within a given workspace, and relate the measurements to each other in one global `CF`. Type and position of the measurements should be easy to program, and the system should be applicable to many inspection tasks without changing its hardware configuration.

In this chapter, the concept of an automated, highly flexible sensor is introduced. It is movable within a medium scale workspace, capable of performing a large variety of measurements and combining them in a common `CF`. The system is composed of a vision sensor, which is rigid and lightweight enough to be mounted on a robot manipulator. The vision sensor can perform measurements ranging from `3D` inspection to 'traditional' image processing tasks. The included projection system allows to optimally illuminate the object for a given tasks, and also acts as the signal source for coded light `3D` reconstruction.

In the following section, the position of the proposed system within the range of existing measurement principles is discussed. The concept of the proposed sensor system is introduced next, followed by a short overview of the hardware architecture. Thereby, scalability of the system is demonstrated through the construction of two prototypes with different fields of view, size, and price.

3.2   Competing Technologies

The market of industrial `3D` surface metrology is traditionally shared among coordinate measuring machines (`CMMs`) and systems based on signaling and recording of light waves. In a market survey from Quality Digest Magazine in 2004 [25], 16% of dimensional metrology equipment in use by respondents was `3D` inspection equipment. A qualitative comparison of the capabilities of industrial metrology tools is given in Figure 3.1. `CMMs` typically cover the accuracy range from 1μm to 15μm over a range up to 1000mm. Laser trackers and light detection and ranging (`LIDAR`) devices offer a measurement range up to $100m$. Measurement accuracy is as low as 10μm for laser trackers and in the millimeter range for `LIDAR` devices. Sensors based on the triangulation principle, including photogrammetric systems, laser triangulation and coded light systems, typically offer relative measurement accuracies of up to 1 : 3000 in a range of several millimeters up to few meters.

In the following sections, an overview of the working principles of these measurement principles and their realizations is given.
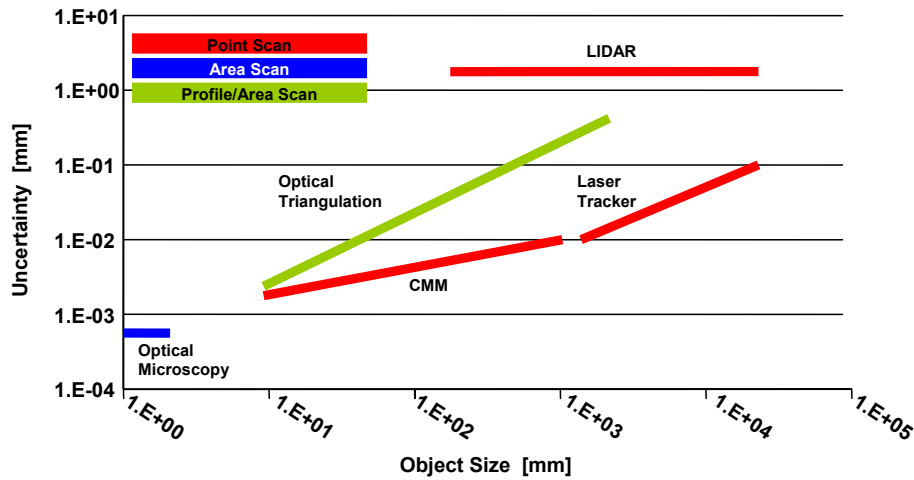
Figure 3.1: Qualitative overview of the accuracy range of industrial metrology tools in the above-millimeter range.

3.2.1 Tactile Coordinate Measuring Machines

Tactile coordinate measurement machines are arguably the most widespread means for acquiring dimensional information of 3D objects in industrial use. The measurement principle is simple: a dimensional measuring probe is mounted on a highly accurate electromechanical actuator and moved to measurement positions on the target object. The actuator records its current position, and the probe records its position relative to the object. The combination of these two measurements gives the 3D position of a measured point.
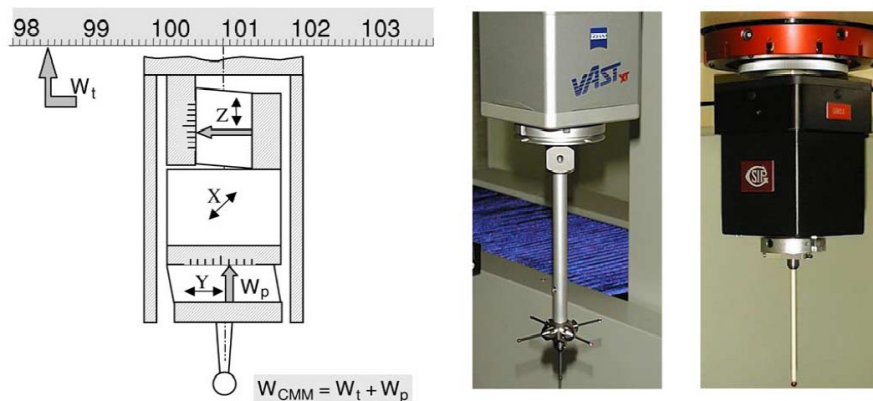
Figure 3.2: Schematic and Images of a CMM Measuring Probe. Figure by courtesy of [112].

While the type of measurement probe is not predetermined by the measurement principle, most CMMs use tactile sensors. The probe is moved until it touches the object

surface. To increase accuracy and enable dynamic measurements like surface profiling
and contour following, displacement sensors are included to record bending as the probe
needle touches the object (see Figure 3.2). Other possible types of probe would be laser
sensors for non-tactile measurements or fringe-projection devices for area measurements.
The electromechanical actuator is usually built in a bridge, cantilever, column or gantry
configuration.

Comparing the specifications of state-of-the-art `CMMs`, measurement accuracy lies be-
tween 1μm and 20μm within a measurement range of up to 1m. Accuracy evaluation of
tactile `CMMs` is evaluated subject to ISO10360-4 [51], or in older specifications VDI2617
[104].

3.2.2   Optical 3D Imaging Systems
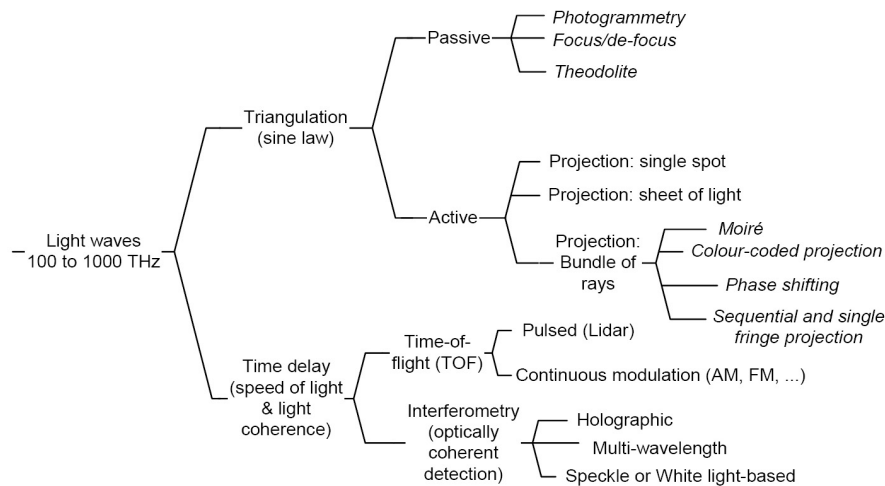
Figure 3.3: Summary of optical 3D imaging techniques. Diagram by courtesy of [6].

Optical 3D measurement systems are fundamentally based on signaling and recording
of light waves (wavelengths from $300nm$ to $3000nm$). Geometric information is derived
either by measuring time (travel time or wave coherence), or by measuring geometric
displacement (triangulation of rays). In [6] a survey of optical 3D measurement methods
is given, as depicted in Figure 3.3.

The travel time measurement principle is commonly realized in form of `LIDAR` systems.
A pulsed or modulated light beam, typically generated by a laser source, is pointed to-
wards a target. The receiver, being at the same location as the source, records the reflected
signal and measures travel time of light pulses, or the phase shift of modulated signals.

With known speed of light, the relative displacement between sender/receiver unit and target point can be calculated. The target point may be either an artificial retroreflective target or a natural surface point. To reference multiple point measurements in one CF, the sender/receiver is mounted on a pan/tilt unit which precisely records rotational sensor motion. In principle such a sensor gives point measurements in a common CF. To perform profile or area measurements, the beam may be automatically swept over the target using mirrors or mechanical motion devices. A measurement range of up to $100m$ and measurement accuracy in the order of millimeters allows applications in geodesy and in the industrial field for displacement measurements of large parts, and safety applications like laser curtains. Recently, area based time of flight (TOF) sensors became available, using infrared light emitting diode (IR-LED) arrays as light source and dedicated pixelated CMOS sensors as receivers (Figure 3.4). This way, 2.5D depth maps with approximate resolutions of 200×200pixels are generated, covering a depth range of approximately $20m$. Accuracy is comparatively poor and lies in the range of a few millimeters up to several centimeters.
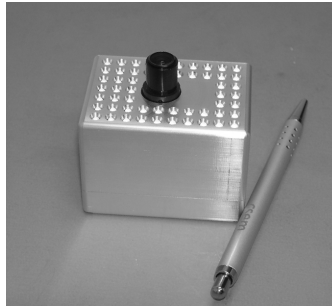
Figure 3.4: Realization of a PMD sensor, including emitter (IR-LED) and receiver in one package [85].

Laser trackers utilize the interferometric measurement principle to determine the relative displacement between sensor and an artificial retroreflective marker to a high level of accuracy. The tracker consists of an interferometric distance sensor (Schematic shown in Figure 3.5) mounted on a high resolution motorized pan/tilt unit. It is able to automatically follow a marker during movement or consecutively scan a high number of markers, providing 3D measurement accuracies as low as 10μm.

Laser triangulation sensors (Figure 3.6) offer a limited measuring range as compared to TOF sensors. Based on the principle of geometrical triangulation of rays emitted by the sender, and rays back projected to the receiver, both sender and receiver need to be placed at different positions in space. The emitter is realized as a point or line laser, whereas the receiver is an area sensor like a CCD or CMOS chip. Depending on the type of emitter,
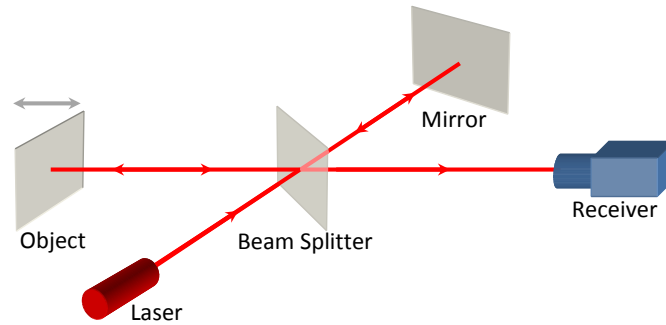
Figure 3.5: Laser interferometry. Distance to an object is determined to high accuracy by comparing the phase of an emitted light beam to the reflection [70].

the sensor gives point or profile measurements. Its range and theoretical accuracy are determined primarily by the baseline (i.e. offset between sender and receiver).

(a) Triangulation based point scanner.

(b) Triangulation based line scanner.
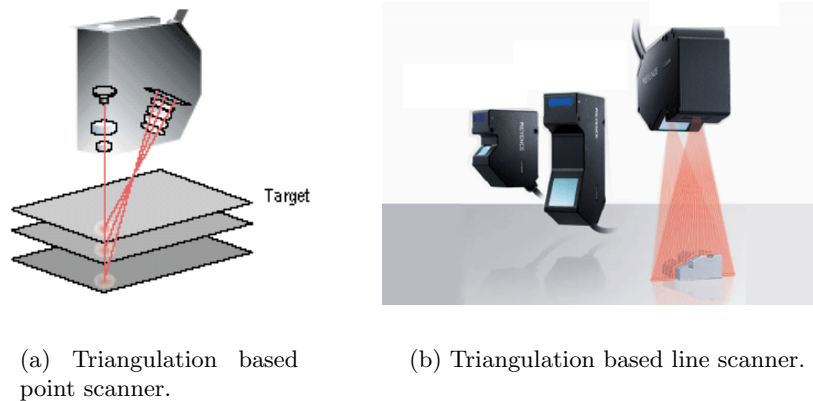
Figure 3.6: Industrial triangulation based scanners [61].

Area triangulation sensors are also based on the triangulation principle. There exist passive realizations, where at least two area based receivers observe a common volume from different positions, or active ones, where an area based emitter signals the object and at least one receiver observes part of the signaled volume. Accuracy and observed volume are functions of the relative emitter and receiver positions. This measurement principle is applied in the proposed measurement system. A detailed introduction is given in Chapter 6.

Traditionally, projection systems have limited applicability in industrial metrology. The reason for this is mainly the technological limitation imposed by projectors, which are usually large, heavy and vibration sensitive devices. As a consequence, projector-camera

metrology systems are applied to medium-scale measurements with a measurement range between 100mm and 1m in a well conditioned environment. Sensors typically consist of one projector and one or two cameras. In the last few years the applicability of projector-camera systems has improved due to the emerging market of light emitting diode (LED) projectors based on digital light processing (DLP) or liquid crystal on silicon (LCoS) technology. These projectors are smaller, lighter and can sustain considerable forces, which makes them suitable for mounting on a robot manipulator. The underlying measurement principle, however, is still mostly limited to medium-scale 3D measurement using coded light projection, following a conservative and proven measurement scheme based on signal processing theory.

Projector-camera systems for medium scale measurement have become popular in quality inspection. The KOLIBRI System by Fraunhofer-Institut für Angewandte Optik und Feinmechanik (Fraunhofer IOF) (Figure 3.7(a)) is a dedicated 3D geometry sensor, which may also be mounted on a robot manipulator and, according to the specification sheet, offers relative 3D measurement accuracies between 1 : 10000 and 1 : 20000, depending on measurement range and surface properties. Measurement time is between 30 seconds and several minutes. The robot is primarily used as a carrier device and not integrated in the measurement chain.

(a)   (b)   (c)
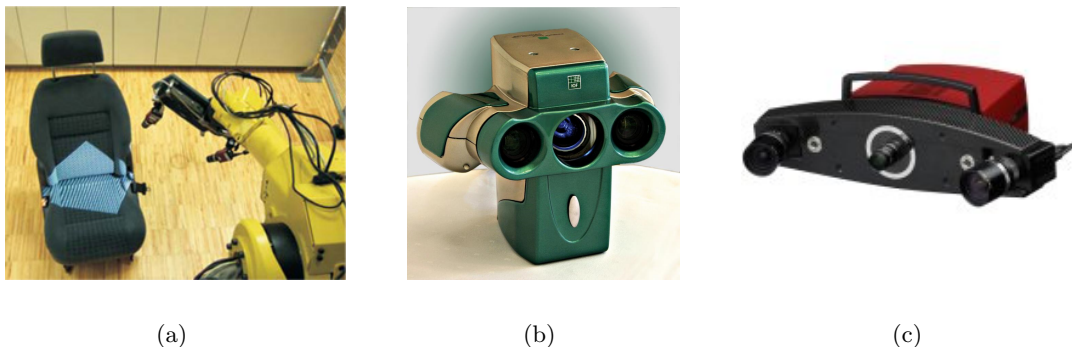
Figure 3.7: Various structured light systems. The Kolibri system (a), a miniature version (b), the ATOS system (c).

A miniaturized version of the sensor is the Kolibri CORDLESS (Figure 3.7(b)), a miniature structured light 3D sensor equipped with two cameras, a digital micromirror device (DMD) projector and a WLAN adapter. The system acts as a pure 3D sensor and offers relative measurement accuracies of 1 : 5000 over a measurement range of $220 \times 170 \text{mm}^2$, at an acquisition time of $150 ms$. Total measurement time is not specified. A

comparable system is also offered by Vialux.

The ATOS System by GOM (Figure 3.7(c)) is also a pure 3D Sensor based on the structured light principle. It offers a variable measurement range from 150×150mm^2 to 2000×2000mm^2, 4 million measurement points are generated in a measurement time of $2s$.

3.3   System Concept

Goal of the proposed measurement system is to combine the high accuracy and medium measurement range of a triangulation sensor with the lower accuracy and high working range of an industrial manipulator. In principle, a comparably cheap and more flexible CMM is constructed. Apart from pure 3D measurements, the included camera system is able to cover the whole range of image processing problems. The included projection system allows adaption of active illumination to a given problem.

In theory, if one would build a vision system to cover a measurement area of 1000×500mm^2, at a spatial resolution of 0.025×0.025mm^2, a camera resolution of 800megapixels would be necessary. If measurements at occluded parts are also required, several cameras at different locations would have to be applied, and if 3D measurements need to be performed, additional sensors need to be used as well. All these systems would have to be rigidly mounted in a common frame, and geometrically calibrated with respect to each other.

A robot mounted vision system on the other hand is able to solve two fundamental types of measurement problems: local measurements with high accuracy within a limited field of view, and global measurements with lower accuracy in the robot workspace.

In accordance with user requirements for industrial quality inspection tasks, a local measurement accuracy of $< 50\mu$m is desired, within a volume of $150 \times 100 \times 80$mm^3. Measurement problems to solve within this range include surface inspection, borehole detection and dimensional measurement, distance measurements and existence checks. Pointwise scanners would not be able to scan the desired volume in a reasonable time, and lack the flexibility for performing different measurements, so a vision based sensor with active illumination by projectors is applied.

Global measurements are performed by moving the sensor using a robot. If relative displacement over a larger distance has to be measured, the robot is moved from one measurement position to the next. Two local measurements, e.g. the position of boreholes, are combined with the relative robot motion to calculate relative displacement of the boreholes. Measurement accuracy and range are ultimately determined by the robot used.

The proposed local sensor must be sufficiently small and lightweight to be carried by an industrial manipulator. It should be scalable to different measurement volumes, and sufficiently rigid to be carried about at high speed by a robot. For this reason, a structured light sensor has been constructed. Consisting of at least one projector and two cameras, it is able to autonomously calibrate itself, perform simple 2D inspection tasks using the projector as a simple light source, as well as 3D measurements by means of structured light. The measurement volume can be enhanced by employing more projectors and cameras, or simply by changing the sensor baseline, at the cost of accuracy. With ongoing miniaturization of projection systems - the smallest sensor constructed in this work was only $150 \times 50 \times 50 \text{mm}^3$ - the measurement principle can be used also with small and fast manipulators.

3.4   Hardware Architecture

To demonstrate flexibility of the principle, two prototypes of the sensor have been constructed.

The first prototype is a miniature version, consisting of a micro-projector, two cameras and a small scale manipulator with a working range of roughly $1.5m$. The effective local measurement range is $50 \times 30 \times 10 \text{mm}^3$. The small scale sensor, shown in Figure 3.8, utilizes a micro-projector based on the LCoS principle for signalling and two cameras in a verged stereo configuration for measurement. Because part of the infrastructure for data transfer and control needs to be mounted on the robot hand, image data transfer is done via gigabit ethernet (GigE), over a single cable. A schematic is shown in Figure 3.10(a). GigE has been applied for the reason of cable length and scalability. Using a switch on the manipulator, all cameras are connected to the PC workstation using a single cable. The small-scale prototype can be moved by a comparably small manipulator with a maximum load as low as $1kg$.

The second prototype, consisting of four cameras, two projectors and a large scale manipulator with a working range of $4m$, features a local measurement range of $150 \times 100 \times 80 \text{mm}^3$. Because, at the time of construction, small scale projectors have a modest resolution of $800 \times 600 \text{pixels}$, two projectors have been combined over a mirror system as shown in Figure 3.9. The cameras are arranged circularly around the projection system, as shown in Figure 3.10(b).

Hardware architecture is comparable to the small prototype, with the difference that four cameras and two projectors are used (Figure 3.10(a)). Projector light sources have to
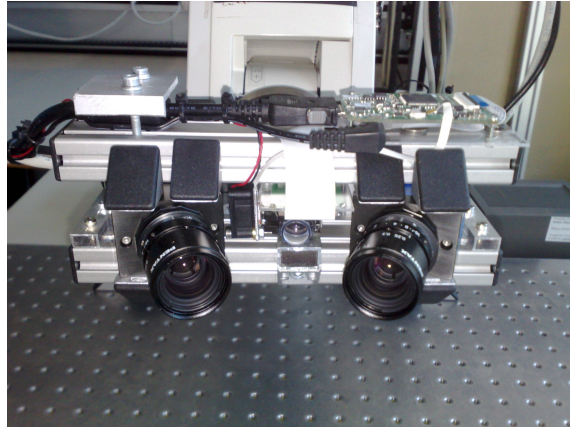
Figure 3.8: Small scale sensor prototype constructed with two industrial cameras and a micro-projector (center).
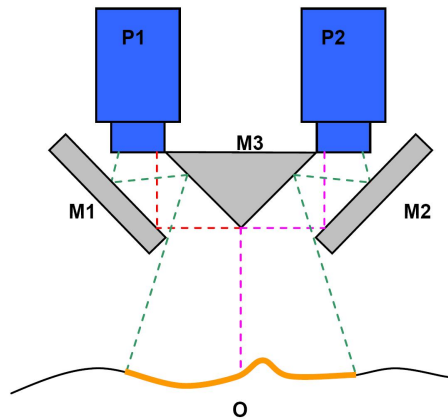
Figure 3.9: Combination of two projector images using a mirror system. Mirrors $M1$, $M2$ and $M3$) are used to align the image rays of projectors $P1$ and $P2$.

resist vibrations and high acceleration forces during robot movements. The standard projector light emitting diodes (LEDs) are replaced by blue high power LEDs, which enhances robustness under stray light and color aberration in combination with color filters on the camera. The weight of the second prototype is higher ($\sim$ 12kg), but still well within the load capacity of a medium scale manipulator.

3.5   Conclusion

Considering the flexibility of triangulation based vision sensors in terms of measurement range, accuracy and construction layout, they are among the best choices for a generic

measurement setup. The realization in form of a camera-projector setup adds the benefit that cameras can be used for more measurement applications than just 3D reconstruction.

The constructed prototypes are primarily meant as research prototypes to demonstrate the system concept. For a production prototype, issues like temperature stability and long-term robustness against high acceleration forces need to be considered as well.

The small-scale prototype was employed in a laboratory environment, while the medium-scale prototype was employed as a demonstrator in a manufacturing hall, and was used to solve measurement tasks as the one described in Chapter 8.
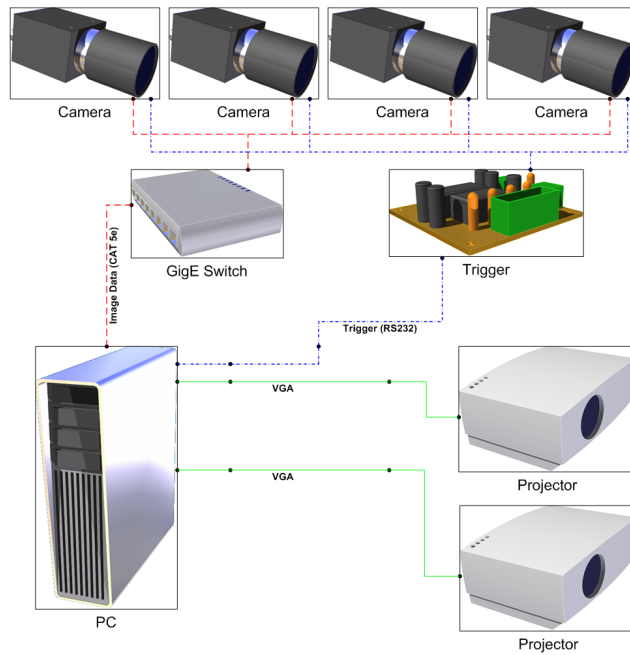
(a) Hardware architecture of the medium scale proto-
type. Four cameras, controlled by a common trigger,
are connected via a dedicated LAN to the processing unit.
The projectors receive their signals via a VGA connec-
tion.
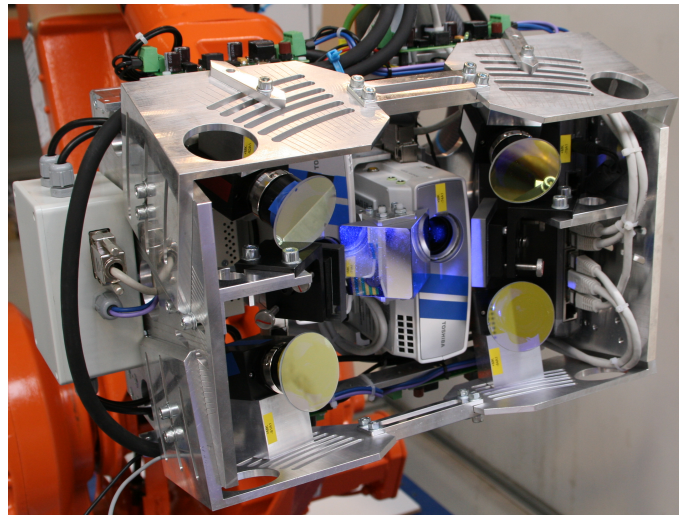
(b) Sensor prototype. The four cameras are equipped with
color filters and arranged circularly around the projection
unit, consisting of two projectors. The aluminium chassis
allows to adjust sensor baseline and mounting to the ma-
nipulator.

Figure 3.10: Hardware architecture of the proposed 3D vision sensor.

Chapter 4

Illumination and Radiometry

Contents

4.1 Introduction

Industrial computer vision applications obtain much of their robustness and accuracy from problem specific image acquisition. Camera and light source are selected and adjusted in a way that the resulting image is ideally suited for further measurement. While the engineer typically has tight control of the camera, control of the light source is limited. The type of illumination (diffuse, directional, ...) is controlled by selecting a specific source. Illumination direction, time and intensity are controlled globally by mechanical mounting and programming.

Replacing the classical light source by one or more projection systems adds a considerable amount of flexibility. Color, intensity and exposure time can be controlled on a per-pixel basis, so small patches of the scene may be illuminated independently. In the simplest case, projectors may be used as ideal point light sources. They may also be used to project specific patterns for the purpose of geometric measurements, or they may project scene specific dampening functions to suppress specular reflections and effectively enhance the dynamic range of an image acquisition system.

In this section, first the process of radiometric calibration of projector-camera systems is reviewed. Knowledge of the radiometric mapping from grey level intensities to physical light intensities is mandatory to perform low-level image enhancement, such as HDR imaging, or local illumination adaption. Methods for enhancing the camera dynamic range are discussed in the next section. A different method of dealing with high scene dynamics is the local adaption of illumination intensity, which is shown in the final section.

4.2   Radiometric Calibration

A digital camera system is a passive sensor which captures the amount of incoming light intensity over a specified amount of time and returns a proportional, dimension-less digital value. The camera response function (CRF) typically is approximately linear, but unknown. Projectors work in a similar way, but in opposite direction, mapping a digital value to a proportional radiant flux.
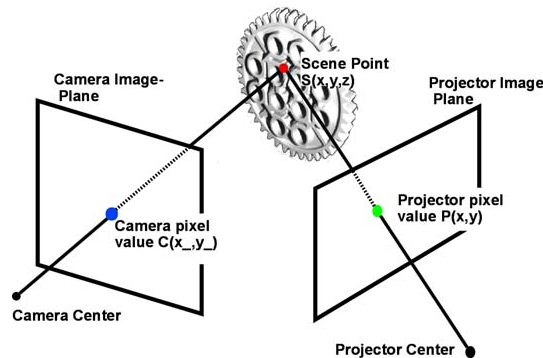
Figure 4.1: Geometry of the projector-camera chain. A point on the object surface is observed by a camera pixel and illuminated by a projector pixel [62].

For a camera, the linear mapping is a consequence of the sensor response function, which is characteristic for photodiodes or photogates. For projectors, nonlinearity is more on purpose, to compensate for the nonlinear response of the human eye and boost visual quality.

In machine vision, it is desirable to have a linear transfer function. A linear relation between physical illumination intensity, projector intensity, camera intensity and exposure time allows an easy fusion of several measurements with different exposure, or different projector intensity. If, for example, the entire scene cannot be captured at once, due to the limited dynamic range of the camera, several images may be acquired, with different

exposure times. Even if in each image a part of the scene is saturated, a weighted linear combination of all images would lead to a complete result.
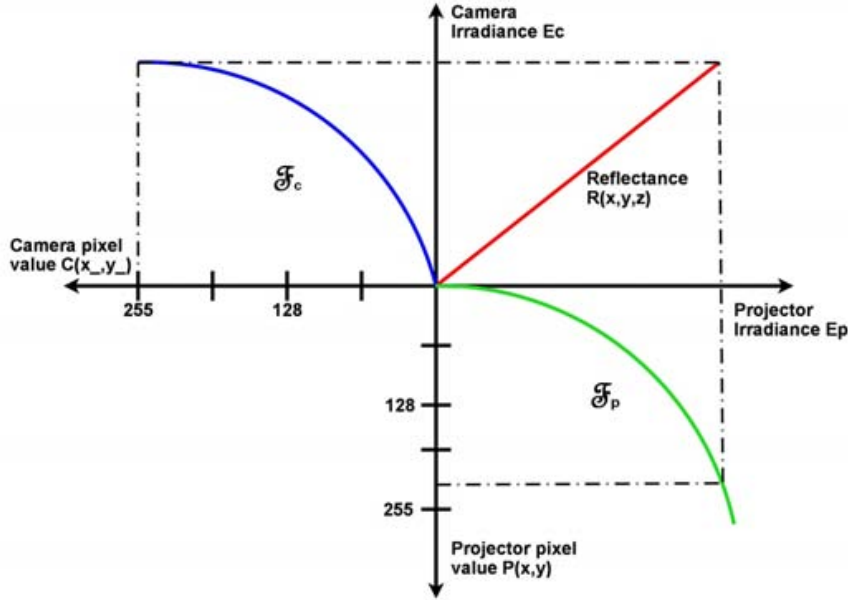
Figure 4.2: Radiometric relations of the projector-camera chain [62].

Geometric light transfer is sketched in Figure 4.1. Light emitted from a single pixel at projector coordinates $\mathbf{P}$ is approximated by a ray, which intersects the object at a scene point $\mathbf{S}$, where the incoming light energy is reflected according to a bidirectional reflectance distribution function (BRDF). The amount of light reflected towards the camera is captured on the image plane at point $\mathbf{C}$.

The radiometric transfer function is shown in Figure 4.2. Illumination intensity for a projector pixel $(x, y)^T$ is specified by a dimension-less value $\mathbf{P}(x, y)$, which is typically digital, and eight bits wide. Projector irradiance is then given by a nonlinear mapping $\mathcal{F}_P(\mathbf{P})$. The object surface, which reflects the incoming projector ray, acts as an amplitude modulator and generates reflected irradiance towards the camera according to a reflectance function R. Surface reflectance is typically isotropic and nonlinear. The distribution of reflected light intensity as a function of incoming intensity is modeled by the BRDF. A classical model including specular reflections is given by Phong [86] as

$$R_{Phong}(\mathbf{L}, \mathbf{V}) = k_d + k_s \frac{(\mathbf{R} \cdot \mathbf{V})^{n_s}}{\mathbf{N} \cdot \mathbf{L}} I_L, \tag{4.1}$$

where $\mathbf{L}$ is the unit vector pointing from $\mathbf{S}$ to the light source with incoming light
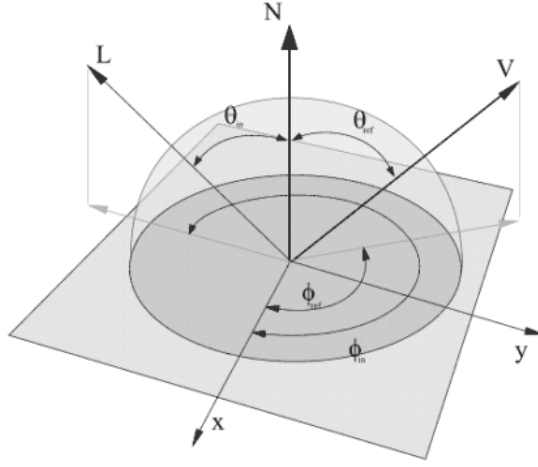
Figure 4.3: Geometric variables of the reflectance function. Surface normal $\mathbf{N}$, direction $\mathbf{L}$ towards the light source, and direction $\mathbf{V}$ towards the camera [76].

intensity I_L, $\mathbf{V}$ is the unit vector pointing towards the camera and $\mathbf{R}$ is given by $\mathbf{L}$, mirrored over the surface normal $\mathbf{N}$ (see Figure 4.3). The second term describes specular surface reflections, k_s and n_s are object surface properties specifying the intensity and crispness of the specular component. The first term models diffuse reflection properties, where incident light is reflected uniformly in all directions. Diffuse reflection is modeled by the Lambertian model

$$k_d = \mathbf{L} \cdot \mathbf{N} \rho I_L. \tag{4.2}$$

For radiometric calibration, a planar, homogeneously colored object with approximately Lambertian reflection properties is observed. Surface reflectance is nearly linear and constant over the surface area. Incoming irradiance on the camera side is mapped to a digital value $C(x_p, y_p) = \mathcal{F}_C(\mathbf{P})$. Thus, $C(x_c, y_c)$ can be expressed as a function of projector pixel intensity $P(x_p, y_p)$ as follows:

$$C(x_p, y_p) = \mathcal{F}_c^{-1}(\Delta t \ R \ \mathcal{F}_p(P(x_p, y_p))). \tag{4.3}$$

Estimation of $\mathcal{F}_c$ and $\mathcal{F}_p$ are related problems. Debevec and Malik [23], as well as Mann et al. [74] proposed methods for recovering $\mathcal{F}_c$. Following the method of [23], camera exposure $X[\sim Wsm^{-2}]$ defines the light energy captured by a camera. The value is proportional (hence the $\sim$ sign) to incoming light irradiance $E[\sim Wm^{-2}]$ and exposure time Δt: $X = E\Delta t$. Camera output is a digital value $Z(x_c, y_c)$ for each pixel, and the

exposure time. The desired mapping $\mathcal{F}_c$ is estimated from

$$Z = \mathcal{F}_c(X) = \mathcal{F}_c(E\Delta t). \tag{4.4}$$

$\mathcal{F}_c$ is a discrete function, defined for all possible values of Z, $S_z = \{z_j, j = 1...n\}$. If a single image is acquired by an uncalibrated camera, each pixel contributes a measurement $Z(x, y), x = 1...m, y = 1...n$, where m, n are the number of image columns and rows, respectively. Exposure time Δt is known and identical for all pixels, incoming irradiance is unknown and possibly different for each pixel. As such, (4.4) is ill-posed. Adding more images of the same (rigid) scene with different exposure times and constant irradiance adds redundancy and allows to compute a MLE of $\mathcal{F}_c$ and $E(x, y)$:

$$(\mathbf{p}_{min}, E_{min}) = \arg\min \sum_{x,y,\Delta t} (Z(x, y, \Delta t) - \mathcal{F}_c(\mathbf{p}, E(x, y)\Delta t))^2, \tag{4.5}$$

where $\mathbf{p} = [p_1, ..., p_n]$ is the parameter vector of model $\mathcal{F}_c$. Reformulating the error function and taking the logarithm of both terms gives

$$Z - \mathcal{F}_c(E\Delta t) \Rightarrow \mathcal{F}_c^{-1}(Z) - E\Delta t \Rightarrow \mathcal{G}(Z) - \ln E - \ln \Delta t, \tag{4.6}$$

where the new function $\mathcal{G} : Z \to (\ln E + \ln \Delta t)$, is parameterized by a new vector $\hat{\mathbf{p}}$. The minimization problem is finally given by

$$(\hat{\mathbf{p}}_{min}, E_{min}) = \arg\min \sum_{x,y,\Delta t} (\mathcal{G}(\hat{\mathbf{p}}, Z(x, y, \Delta t)) - \ln E(x, y) - ln\Delta t)^2 + \lambda \sum_j \mathcal{G}''(\hat{\mathbf{p}}, z_j)^2. \tag{4.7}$$

The last term, weighted by a scalar λ, punishes large second derivatives of $\mathcal{G}$ to reach a smooth curve, especially if some values z_j are not observed by measurements. A least squares method as described in Section 2.5 can be used to estimate a solution. In Figure 4.4 a typical correction curve $\mathcal{G}$ is shown. The correction mapping is a typical smooth gamma correction curve, therefore the use of a smoothing constraint is justified.

The capability of on-scene calibration is an advantage, especially in industrial manufacturing inspection, where the same type of object is observed in a repeated manner. Gray values and exposure times observed during calibration will most likely occur also during measurement. Further, with known $\mathcal{G}$ and Δt, it is possible to calculate the incoming irradiance per pixel, which further allows to create HDR images and to calibrate a
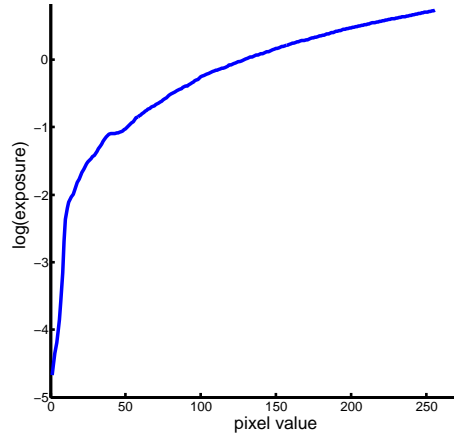
Figure 4.4: Calibrated `CRF`. Digital intensity values are mapped to camera exposure. If no ground truth information on light intensity is available, exposure is also dimension-less, but approximately linear in incoming light intensity and exposure time [62].

projection system.

Radiometric projector calibration is closely related to camera calibration in terms of methodology. A calibrated camera is required to close the radiometric chain as depicted in Figure 4.2. If the projector illuminates a uniform, Lambertian surface, which is observed by the camera, the relation between camera irradiance and projector irradiance is approximately linear. Further, intensity values z_j on the projector side can be set manually, allowing tighter control of the observations. Similar to (4.5) the goal is to obtain a solution for

$$(\mathbf{p}_{min})_p = \arg \min \sum_{x_p, y_p} \left(Z_p(x_p, y_p) - \mathcal{F}_p(\mathbf{p}_p, E_p(x_p, y_p)) \right)^2. \tag{4.8}$$

In contrast to camera calibration, projector gray values $Z_p(x, y)$ are not a function of the time integral over irradiance, but related directly. Emitted irradiance E_p is linearly related to camera irradiance $E_c = R(E_p) = \rho_a E_p + \rho_b$. In case of a homogeneous, Lambertian object surface, constant stray light and no light attenuation through optical components, reflectance is identical for all projector and camera pixels.

$Z_p(x, y)$ can be controlled by the user, so a simple calibration procedure would be to subsequently set all possible values $z_{j,p}$, acquire an image and estimate $\mathcal{G}_p(z_{j,p})$ by combining (4.3) and (4.7):

$$(\hat{\mathbf{p}}_{min})_p = \arg\min \sum_{x,y} \left(\mathcal{G}_p(\hat{\mathbf{p}}_p, Z_p(x,y)) - R^{-1}(\mathcal{G}(Z) - ln\Delta t) \right)^2 + \lambda \sum_j \mathcal{G}_p''(\hat{\mathbf{p}}_p, z_{j,p})^2.$$

(4.9)

Looking at the data term of (4.9), the linear components of $\mathcal{G}_p$ and R^{-1} are redundant and can not be calibrated simultaneously. If the reflectance function is unknown, it may be set to fixed values, e.g. $\rho_a = 1$ and $\rho_b = 0$. This allows to determine $\mathcal{G}_p$, but only up to scale. Setting $\rho_b = 0$ is reasonable, if no stray light is present. In practice, setting $\rho_a = 1$ does not make much of a difference, because the main interest lies in calibrating the nonlinearities of $\mathcal{G}_p$. If projector gray values can be mapped to a linear space of irradiance values, relative intensities can be set to high accuracy, which is important e.g. for projecting phase shift gray value patterns.
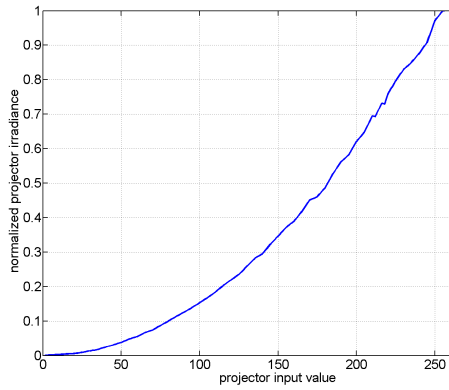
Figure 4.5: Calibrated PRF. Digital intensity values are mapped to projector irradiance [62].

An example of a PRF is shown in Figure 4.5. The complete radiometric calibration procedure is fully automatic. External constraints are the availability of a uniform, flat, Lambertian surface and the absence of stray light. Both constraints can be fulfilled fairly easy by providing a closed chamber and a calibration target.

4.3 High Dynamic Range Imaging

Measurement accuracy and robustness of image processing methods depend on image contrast (radiometric resolution) and camera dynamic range (radiometric range). Considering a scene with shadow regions and specular highlights, it is practically impossible to avoid under- or over-saturation with state-of-the-art camera equipment.
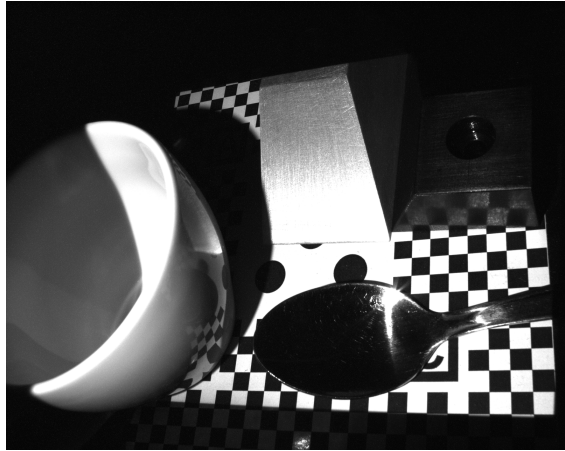
Figure 4.6: Scene with high dynamic range. Objects with different reflectance properties enlarge the dynamic range. Here, a ceramic cup, a teaspoon and two metal workpieces are shown on planar cardboard with a black/white checkerboard pattern printed on it [62].

One solution to the problem is to tune illumination and avoid shadows and speckles in the first place. This may not always be possible, because diffuse illumination is hard to implement, bulky and expensive. In some metrology applications the light source has to fulfill geometric constraints, too. For example in structured light metrology or photometric stereo.

Another possibility to cope with a high dynamic range of the scene would be to expand the dynamic range of the camera. According to a review tutorial of El Gamal [36], illumination intensities in a scene may vary over $100dB$ and more. Consumer `CMOS` sensors feature a dynamic range of $54dB$, whereas high-end `CCD` sensors give $78dB$ and more. The imaging problem is illustrated in Figure 4.6, where the shadow parts of the scene are under-saturated, while specular highlights exceed the sensor dynamic range.

4.3.1   Basic Definitions

To quantify sensor dynamics, basic definitions regarding the sensor noise model, signal to noise ratio (`SNR`) and dynamic range (DR) are required. The classical image sensor is a photodiode, which linearly converts incident light into photocurrent i_{ph} in range of few femtoamperes. Photocurrent is integrated over time, resulting in a charge, which is converted to voltage, digitized, and read out. The entire process is subject to noise, which ultimately limits contrast and dynamic range. Figure 4.7 gives a model of photocurrent to charge conversion [36].

Photocurrent i_{ph} is added to the dark current i_{dc}, which is a result of junction leak-
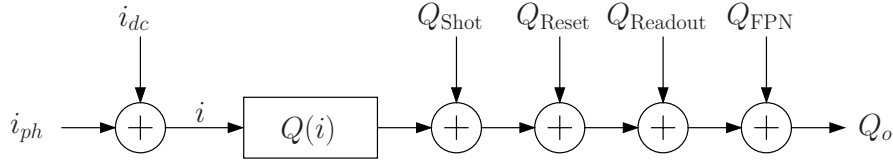
Figure 4.7: Photocurrent to charge conversion [36].

ages. The sensor transfer function $Q(i)$ linearly integrates current over time and gives a proportional charge, which is limited by sensor saturation:

$$Q(i) = \begin{cases} 1/q(it_{int}) & \text{for } 0 < i < qQ_{sat}/t_{int} \\ Q_{sat} & \text{for } i \geq qQ_{sat}/t_{int} \end{cases} \tag{4.10}$$

Well capacity Q_{sat} imposes an upper limit to the effective signal range. Further, several independent noise functions are added to the signal charge:

- Q_{shot}: zero mean integration noise (shot noise),

- Q_{reset}: reset noise,

- $Q_{readout}$: zero mean readout noise, including quantization noise,

- Q_{FPN}: systematic fixed pattern noise (offset and gain) due to device mismatches,

resulting in an equivalent additive input noise current I_n (see Figure 4.8) with power

$$\sigma_{I_n}^2 = (q/t_{int})^2(1/q(i_{ph} + i_{dc})t_{int} + \sigma_r^2), \tag{4.11}$$

where $\sigma_r^2 = \sigma_{reset}^2 + \sigma_{readout}^2 + \sigma_{FPN}^2$. Modern CMOS pixels allow for correlated double sampling (CDS), which is a relative measurement of dark pixel value and exposed pixel value. Systematic noise components, like σ_{reset} in CCD and σ_{FPN} in CMOS sensors, are hereby eliminated, whereas random noise powers are doubled.
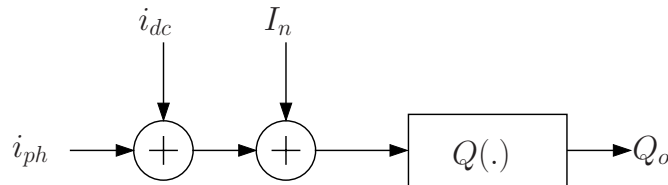
Figure 4.8: Equivalent additive noise [36].

Signal to noise ratio (SNR) is the ratio of input signal power to input noise power. It is given by

$$SNR(i_{ph}) = 10 \log_{10} \frac{i_{ph}^2}{(q/t_{int})^2(1/q(i_{ph} + i_{dc})t_{int} + \sigma_r^2)}. \tag{4.12}$$

In Figure 4.9 an example of SNR versus i_{ph} is given for three levels of dark current. Dark current i_{dc} effectively determines the lower bound for a usable signal.
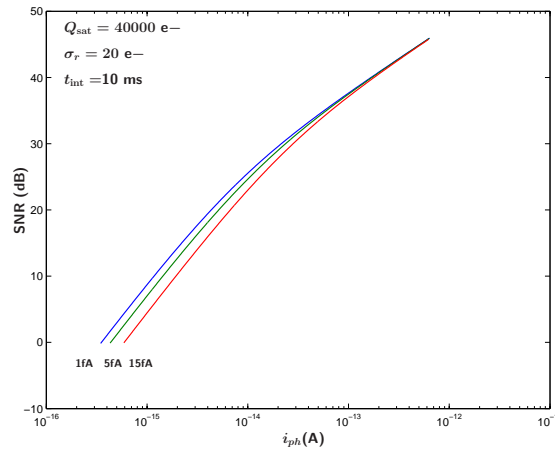
Figure 4.9: SNR over photocurrent for three levels of dark current [36].

Dynamic range quantifies the ability of a sensor to adequately image highlights and dark shadows in a scene, and is defined as the ratio of largest non-saturating input signal to lowest detectable input signal $DR = 20 \log_{10}(i_{max}/i_{min})$ [36]. From (4.10), the largest signal is given by

$$i_{max} = q\frac{Q_{sat}}{t_{int}} - i_{dc}. \tag{4.13}$$

The smallest detectable signal must exceed the noise level under dark conditions, and is hence given by σ_{I_n} with $i_{ph} = 0$:

$$i_{min} = \frac{q}{t_{int}}\sqrt{\frac{1}{q}i_{dc}t_{int} + \sigma_r^2}. \tag{4.14}$$

In Figure 4.10 a typical DR curve over integration time is shown. Because dark current is integrated over time, DR must decrease with growing t_{int}.
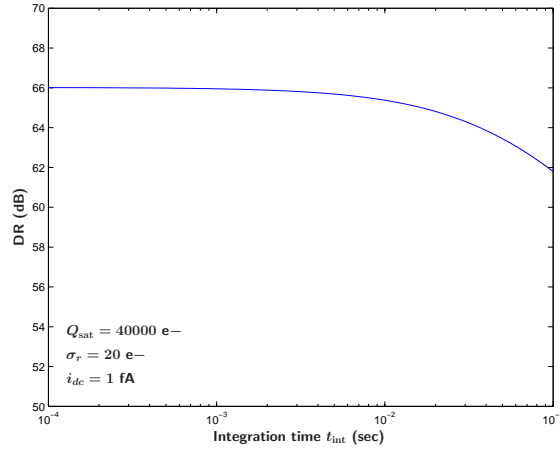
Figure 4.10: Dynamic range over integration time [36].

4.3.2 HDR Imaging Principles

An increase in sensor dynamic range can be achieved by increasing i_{max}, decreasing i_{min}, or changing the measurement principle. Looking at the physical parameters in (4.13) and (4.14), the tunable parameters are Q_{sat} and t_{int}.

Integration time is most easy to adapt, but a global change in t_{int} shifts i_{min}, as well as i_{max}, resulting in constant DR. The algorithm proposed in [23] produces a combination of temporal multiplexed images with different t_{int}. If the sensor transfer function is calibrated (see Section 4.2), the input signals can be linearized and transferred to a global radiance map by

$$\ln E_i = g(Z_{ij}) - ln\Delta t_j. \tag{4.15}$$

A weighted linear combination of the signals produces the resulting image:

$$\ln E_i = \frac{\sum_{j=1}^{P} w(Z_{ij})(g(Z_{ij}) - \ln \Delta t_j)}{\sum_{j=1}^{P} w(Z_{ij})}, \tag{4.16}$$

where w is a weighting function which decreases the influence of signals near the saturation levels. Hardware implementations use a simpler combination scheme, e.g. by taking the largest sample before saturation. Multiple image capture increases DR by raising i_{max}, the lower bound stays fixed. A possibility to lower i_{min} would be redundant image capture. Calculating the mean of several images with identical t_{int} decreases the influence of random noise, again at the cost of capture time.

Essentially the same principle may be applied in spatial multiplex, where neighboring pixels are exposed differently. Nayar et al. [84] proposed to apply light blocking filters in front of a sensor array, e.g. constant density filters, DMD or liquid crystal display (LCD) panels. Here, the gain in radiometric resolution of course comes at the cost of spatial resolution and sampling artifacts, if filters and pixels are not well aligned.
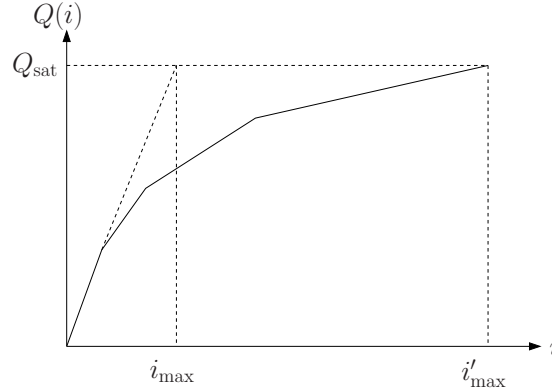
Figure 4.11: Well capacity adjustment progressively flattens the charge transfer function to increase i_{max} [36].

Well capacity Q_{sat} is predetermined by electrical characteristics of the photodetector. Yet, the charge transfer function (4.10) can be flattened to increase i_{max}, as shown in Figure 4.11. This flattening can be achieved by providing controlled drain from the well, e.g. through a half-open reset signal in a CMOS pixel. The increased input current range is still mapped to a constant output charge range, which leads to decreased SNR and a compression of the input signal. Dynamic compression of $Q(t)$ during integration gives a nonlinear sensor response curve with increased DR, at the cost of lower SNR for bright pixels. Fixed pattern noise also increases, as the drain output becomes an integral part of the measurement process.

In a different approach, pixel intensity is created by measuring the saturation time instead of i_{ph}. Saturation time is given from (4.10) as

$$t_{sat} = qQ_{sat}/i_{ph}, \tag{4.17}$$

and is $1/n$ proportional to i_{ph}. The minimum input current is hereby determined by the maximum allowed integration time. Maximum current is limited by the precision to which t_{sat} can be measured. One difficulty with this approach is precise detection of sensor saturation.

4.3.3 Experiments

In the following, the applicability of HDR imaging to industrial measurement applications will be evaluated. While the improvement of SNR and DR in general has been shown before, we focus on projector-camera systems in industrial settings, i.e. working volumes of approximately $200 \times 200 \times 80 \text{mm}^3$, containing objects with glossy surface (metal parts), as well as dark, diffuse reflecting surfaces. As a reference measurement task, the problem of 3D reconstruction using a structured light method with subsequent refinement using phase-shifting has been chosen. Focus of the experiments is not to evaluate the reconstruction procedure itself, therefore the reader is referred to Chapter 6, but to design experiments, where radiometric inaccuracies, in this case inaccuracies of the acquired sinusoidal pattern, can be related to ground truth information, in this case 3D surface planarity. As quality measures for a set of reconstructed points, the standard deviation σ of point normal distance to the best-fit plane (in a least-squares sense) has been chosen. If applicable, spatial resolution of the reconstruction is given in terms of point density ρ, the number of reconstructed points per surface area.

(a) **(b)** **(c)** **(d)**
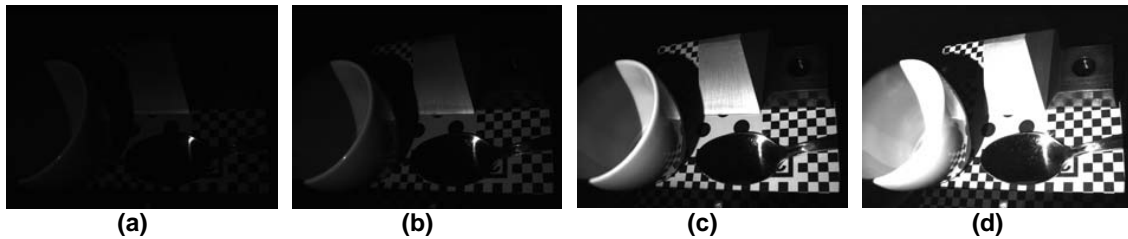
Figure 4.12: HDR image stack. Four images of a scene, with different exposure time settings: $5000\mu s$ (a), $20000\mu s$ (b), $100000\mu s$ (c), $350000\mu s$ (d) [62].

The result of a classical HDR experiment using multiple exposure times is shown in Figure 4.13. The test scene contains a ceramic coffee cup, a diffuse reflecting checkerboard pattern (printed paper), one teaspoon and two shiny metal parts. The scene is illuminated by a LED video beamer and observed by an industrial CCD camera.

It is shown in Figures 4.12 (a)-(d), that no exposure setting allows to capture the full dynamic range of the scene. Combination of the differently exposed images, according to the method described in Section 4.3.2, results in a HDR image as shown in Figure 4.13. The simultaneously obtained CRF is shown in Figure 4.14. One problem which is inherent to the HDR approach can be observed on the teaspoon. Exposure time can not be chosen deliberately small, so there may still remain saturated regions in the image. Further, to resolve dark regions, exposure time must be high, resulting in a large number of saturated
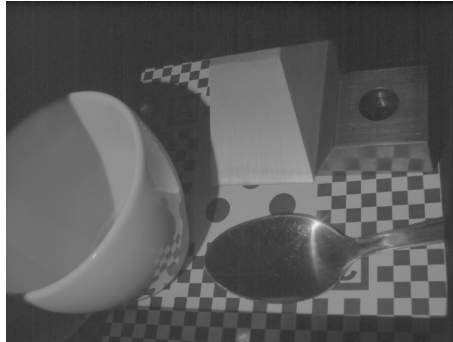
Figure 4.13: `HDR` image. An image stack of four differently exposed images allows to cover the dynamic range of the test scene in one image. Gray values are scaled for visualization [62].
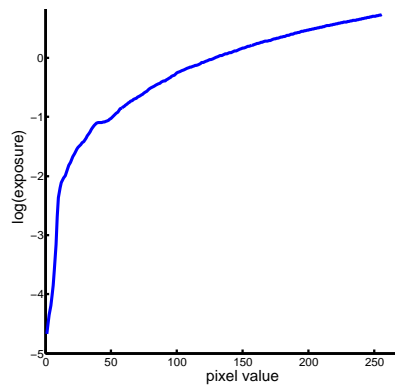
Figure 4.14: Mapping from camera gray values to exposure [62].

Image	Exposure time [ms]	Oversaturated [%]	Undersaturated [%]
Plain, Fig.4.15(a)	80	0.0	0.0
Over, Fig.4.15(c)	150	7.1	0.3
Under, Fig.4.15(e)	50	0.2	17.7
HDR, Fig.4.15(g)	$[1.5, 20, 80, 150, 250]$		

Table 4.1: Amount of saturation on checkerboard images.

pixels and corresponding artifacts, like blooming. Noise in the optical path, like dirt on the lens, become also more prominent when using long exposure times.

In general it is difficult to obtain ground truth information in the scene dynamic range, so evaluation is to some extent qualitative, with a focus on the benefit for subsequent image processing tasks. Two experiments are conducted using `HDR` imaging. First, a planar checkerboard pattern (printed paper) is illuminated by a projector with a sinusoidal
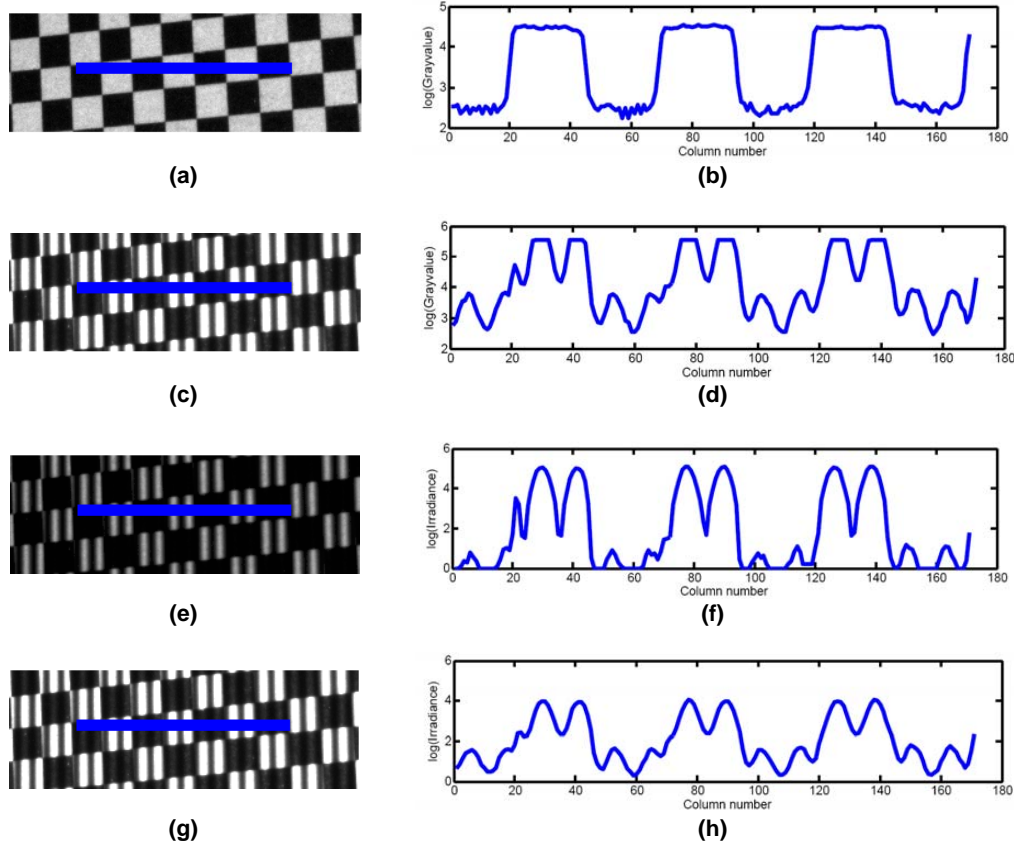
Figure 4.15: HDR image of a checkerboard pattern. A checkerboard pattern printed on paper is observed by a CCD camera (image in (a)), horizontal gray value profile (b)). A projector illuminates the object with a sinusoidal signal (c), (e). Regardless of the exposure time, the gray value profile runs into saturation (d), (f). Combination of five exposure times allows to capture the scene dynamic range (g), (h) [62].

signal. The resulting signal exceeds the camera dynamic range (see Figure 4.15). Using a combination of five exposures it is possible to cover the complete signal and recover the sinusoidal phase over the entire image, as given in Table 4.1. The same experiment is applied to a shiny metal plane (Figure 4.16). Due to the point light source, a direct specular reflection occurs in the image center, saturating the sinusoidal signal on roughly 30% of the surface area. A combination of four exposure times allows to cover the complete dynamic range of the scene, as given in Table 4.2.

The recovered sinusoidal signals are further used to perform 3D reconstruction on the two reference objects, based on structured light with phase-shift refinement. It is assumed that both objects are perfectly planar, so the difference between measured points and the

(a)

(b)

(c)

(d)
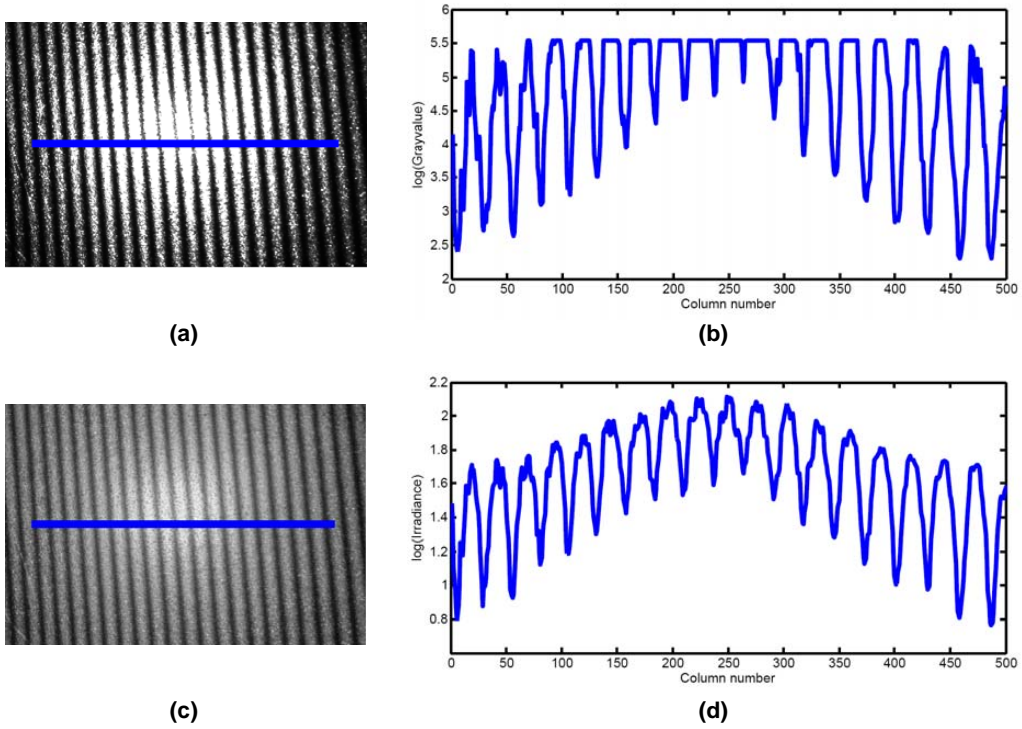
Figure 4.16: `HDR` image of a sinusoidal pattern. A shiny metal plate is observed by a `CCD` camera. A projector illuminates the object with a sinusoidal signal (image in (a), horizontal gray value profile (b)). Combination of five exposure times allows to capture the scene dynamic range and recover the sinusoidal pattern (c), (d) [62].

Image	Exposure time [ms]	Oversaturated [%]
Over, Fig.4.16(a)	5	29.8
HDR, Fig.4.16(c)	$[0.5, 1.5, 3, 9]$	N/A

Table 4.2: Amount of saturation on a planar metal surface.

best-fit plane (using a least-squares fit), is taken as a measure for reconstruction quality.

Results are given in Table 4.3. A significant increase in point density ρ on the checkerboard pattern, and a slight increase for the metal surface are visible. Standard deviation of the point normal distances from the best-fit plane is significantly improved in both cases.

Scene	σ [mm]	ρ [$numpoints$/mm^2]	σ_{HDR} [mm]	ρ_{HDR} [$numpoints$/mm^2]
Checkerboard	0.074	78	0.038	115
Shiny	0.085	113	0.045	124

Table 4.3: Structured light reconstruction results on HDR images.

4.4 Radiometric Illumination Adaption

A projection system offers the possibility to adapt light intensity locally on the emitter side. The goal is to illuminate a region on the object surface, which projects to one camera pixel, with sufficient intensity to stay within the sensor dynamic range. Bimber and Iwai [9] exploited this principle for contrast enhancement on planar scenes and reported an effective dynamic range of more than 150db (see Figure 4.17).
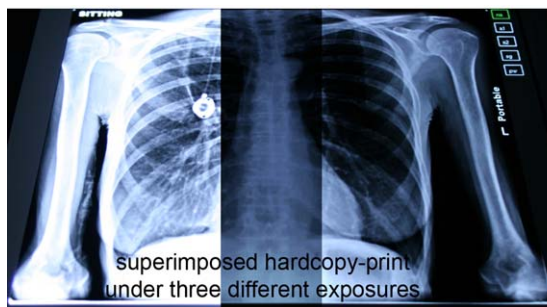
Figure 4.17: Contrast enhancement using projectors [9]. A projector pattern is iteratively adapted to surface texture (here the printout of an X-ray image). The initial contrast ratio of 100db is hereby increased to over 150db.

If a non-planar, textured scene is observed, the amount of light captured by a single pixel is a function of projector parameters, surface geometry, reflectance, and camera parameters. The goal is to adapt projector intensity, which is a local property of each projector pixel, and camera exposure time, which is a global property of all camera pixels, to maximize the dynamic range over the scene. Grossberg et. al. [39] performed pattern adaption under known object geometry to virtually change object appearance. Konickx and Van Gool [64] [63] recently proposed radiometric projector pattern adaption in the context of structured light reconstruction.

In measurement problems, surface geometry, texture and reflectance are usually unknown, which makes the problem of pattern adaption ill-posed. Known surface geometry provides correspondence information between projector and camera pixels. Known surface

orientation allows to calculate incidence angles to account for specularity. If rough geometry information can be obtained without radiometric adaption, an iterative procedure can be applied to alternatively refine geometry and the adaption pattern.

4.4.1   Iterative Pattern Adaption

To establish a unique correspondence between projector and camera pixels, scene geometry must be known. In case of limited saturated regions in the image, scene geometry can still be measured partially, and interpolated in saturated regions. As such, rough correspondence between projector and camera pixels is established which allows adaption of local illumination intensity. The adapted intensity pattern in turn allows to refine scene geometry, which closes the iteration loop.

In the following, we assume the projector and the camera to be radiometrically and geometrically calibrated. A rough estimate of scene geometry is also known. A uniform intensity pattern in the middle of the projector DR (e.g. gray value 128) is projected and the response is acquired by the camera. The surface reflectance map R can be calculated according to (4.3) as:

$$R = \frac{\mathcal{F}_p(P(x_p, y_p))\Delta t}{\mathcal{F}_c(C(x_c, y_c))}. \tag{4.18}$$

Based on (4.18) and (4.3), measured camera pixel intensities $z(x_c, y_c)$ are mapped to projector pixel intensities $z_p(x_p, y_p)$, as shown in Figure 4.18 (a). In regions where no 3D information is available, z_p is interpolated (Figure 4.18(b)). Because an estimate of R is also available, the desired camera gray value $z_{i,c}$ is also mapped to a desired projector intensity $z_{i,p}$. The compensation mask (Figure 4.18 (c)) in projector image space is given by $\Delta z(x_p, y_p) = z_{i,p} - z_p(x_p, y_p)$.

The resulting mask might be wrong, because the correspondence of projector and camera pixels is not known precisely and exact $\Delta z(x_c, y_c)$ could not be calculated due to saturation. However, if the scene geometry estimate does not contain gross outliers, the correspondence error is local. Saturation effects typically cover a significant image area, so it is likely that the dampening of a slightly wrong pixel still gives an improvement. In the next iteration, surface geometry is known with more detail, again improving the compensation mask, until it converges to a stable state. The iteration ends, if either the camera shows an ideal response to the projection, or the limits of projector dynamics are reached.

| (a) | (b) | (c) |
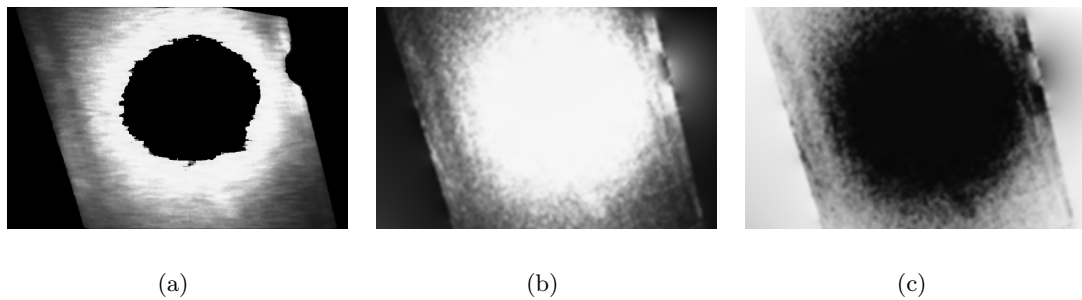
Figure 4.18: Pattern adaption iteration. With an initial 3D reconstruction, camera gray values are mapped back to projector gray values (a). Where no reconstruction is possible, values are interpolated (b). A compensation mask (c) is given by the difference of ideal projector intensity and measured intensity [62].

4.4.2 Light Source Separation

An active computer vision system consists of a carefully selected light source and one or more cameras. Arguably, most measurement errors in vision systems are caused by unexpected illumination conditions, due to stray light and reflections in the scene.

To avoid stray light, the best solution would be to carefully shield the measurement volume from external influences. If this is not possible, e.g. in robotic applications with a large working area, it is favorable to work with relative measurements, as shown in Figure 4.19. A pair of images is acquired, one with active illumination switched on, the second one with stray light only. In the difference image, stray light is effectively suppressed.

Figure 4.19: Relative intensity measurement. In the presence of stray light, active illumination provides the possibility to subtract the scene with stray light only from the actively illuminated scene [70].

The second source of unwanted illumination comes from interreflection. Depending on the illumination direction and scene geometry, one part of the object may reflect light to other parts, leading to false pattern decoding in structured light, or saturation effects in 2D image processing. Interreflections are harder to suppress, or even to identify. Especially

measurement principles like pattern projection, which are based on geometric constraints on the light source, are severely hampered by reflection. Nayar et al. [83] proposed a method, based on projection of high spatial frequency patterns, to identify interreflections. Xu et al [113] applied the method to 3D reconstruction. Although these methods do not allow to separate the original signal from an overlaid reflection, they allow to robustly detect and mask out such regions and prevent false measurements.

If no stray light is present in the observed scene, the total radiance L emitted from a surface patch has two components, depending on their origin:

$$L = L_d + L_g. \tag{4.19}$$
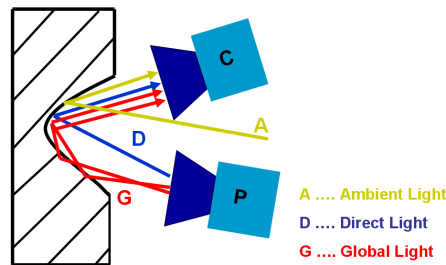
Figure 4.20: Direct and global reflection. Light intensity captured by a camera pixel may be caused by direct reflection of a projected ray at the observed surface patch, or by multiple interreflections [62].
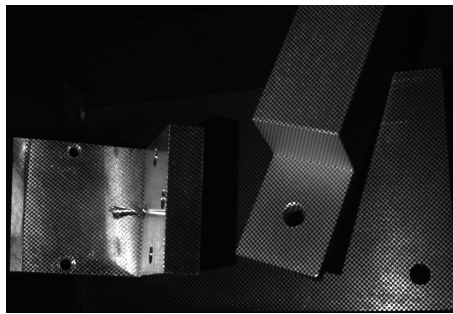
Figure 4.21: High frequency pattern for light source separation [62].

Radiance L_d is caused by direct illumination, whereas L_g originates from reflection. A projector allows to emit a binary pattern with high spatial frequency, e.g. by switching off every second pixel. An example is shown in Figure 4.21. Assuming that global radiance for a surface patch is caused by reflected light rays which are distributed over a larger surface area (large w.r.t. the applied spatial frequency, see Figure 4.20), the camera

receives radiance

$$L^+ = L_d + L_g/2 \tag{4.20}$$

from directly illuminated regions, and

$$L^- = L_g/2 \tag{4.21}$$

from regions corresponding to a switched off pixel. L_d and L_g may be identified by changing the high frequency pattern, such that every surface region is observed at least once with and without direct illumination. A minimum of two pattern projections are necessary to reach this condition, but in practice two problems arise: the object surface is sampled in finitely large intervals by the camera, and the projector may be slightly defocused, producing rays with medium intensity at the pixel transitions. As a consequence, on- and off-pixels may cancel out in some surface regions. Nayar tried to overcome the problem using a checkerboard pattern for projection. Shifting the pattern several times in small steps, L^+ and L^- are identified by taking the maximum and minimum intensity for every pixel over all images:

$$L_{m,n}^+ = \max\{x_{m,n}^i\} L_{m,n}^- = \min\{x_{m,n}^i\} \tag{4.22}$$

for all images $i = 1...n$. The size of the checkerboard patches is defined externally. One needs to find a compromise between violation of (4.20) and (4.21) for too large patches, and artifacts due to low projector and camera resolution.

4.4.3   Experiments

4.4.3.1   Pattern Adaption

A projector-camera system allows local adaption of illumination intensity to avoid saturation. For evaluation, several test objects are used, with surface properties which are common in industrial inspection. The feasibility of the general approach is evaluated on a planar checkerboard (printed paper). Because a non-saturated image can only be an intermediate result in the measurement chain, the applicability to structured light reconstruction is also evaluated.

Radiometric calibration of the projector-camera system is conducted as described in Section 4.2. The results are sensor response functions for camera (`CRF`, Figure 4.22(a)) and

projector (PRF, Figure 4.22(b)). The CRF shows approximately linear behaviour between ray levels 5 and 250, whereas the PRF is clearly nonlinear. All irradiance values are normalized to the maximum projector irradiance.
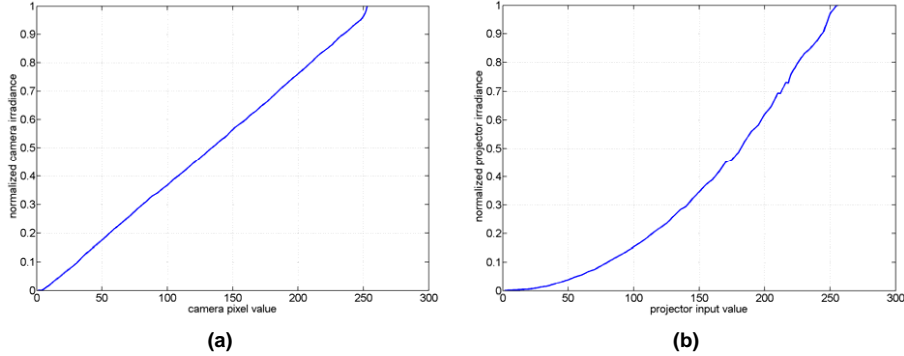
Figure 4.22: Camera and projector response functions. The camera response to incoming irradiance is approximately linear, while the projector clearly shows approximately quadratic behavior [62].

The feasibility of the proposed procedure is demonstrated on a planar checkerboard pattern, shown in Figure 4.23(a). A compensation pattern shall be obtained, which compensates the black and white pattern to obtain a uniform camera image with gray level 128. The result after one iteration is shown in Figure 4.23(b), the final result, after two iterations, in Figure 4.23(c). The original pattern, approximately $100 \times 100 \text{mm}^2$, has a bimodal grey level histogram under uniform illumination. The bright regions have a larger scatter, due to non-uniform illumination by the point light source, as shown in Figure 4.24(a). After compensation, the grey level histogram shows a significant peak at the desired gray level (Figure 4.24(b)). The compensated image shows only slight artifacts, especially at the pattern edges, which originate from the limited spatial projector resolution, it is not possible to reach an ideal compensation.

The pattern adaption process is further evaluated on a planar metal surface (Figure 4.25). Hereby, the goal is to compensate for specular reflections and modulate the resulting compensation pattern with a sinusoidal signal. As such, the signal is visible over the entire object surface. The sinusoidal signal is further used to obtain a sub-pixel accurate 3D reconstruction by means of a structured light procedure with phase-shifting refinement. Point density and scatter around the best-fit plane are given in Table 4.5 as quality measures for the adaption process. The procedure was repeated for different specimen (see Figure 4.26): the checkerboard pattern on printed paper (d), a chromed metal object (a), an anodized (b), and a black-finished object (c). In comparison to a standard structured
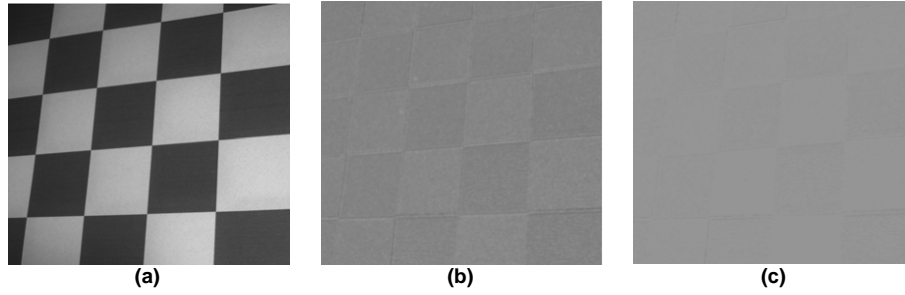
Figure 4.23: Pattern adaption on a checkerboard. The radiometric compensation method is demonstrated on a planar checkerboard (a). After one iteration, the printed pattern is effectively suppressed (b), and reaches a nearly uniform distribution after two iterations (c) [62].
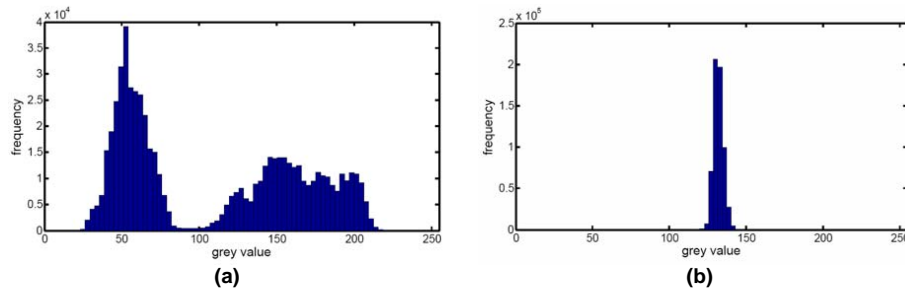
Figure 4.24: Histogram of adapted image. The original checkerboard image (Figure 4.23 (a)) shows a bimodal intensity histogram. After two iterations of pattern adaption, the checkerboard is effectively eliminated (b), see also Figure 4.23 (c) [62].

Iteration	σ_{PAT} [mm]	ρ_{PAT} [%]
1	0.319	0.74
2	0.129	0.93
3	0.064	0.99
4	0.063	1.0

Table 4.4: Pattern adaption for structured light reconstruction on a shiny metal plane (Figure 4.25), with an observed area of 2300mm^2. After four iterations, a total of 188602 points have been reconstructed.

light procedure, the adaption methods using HDR imaging and pattern adaption lead to improved accuracy in terms of point scatter. Extreme surface properties, like black finished steel, show the limits of both adaption procedures. Camera and projector dynamics fail to cover the scene dynamic range and result in little or no improvement compared to the standard structured light procedure.
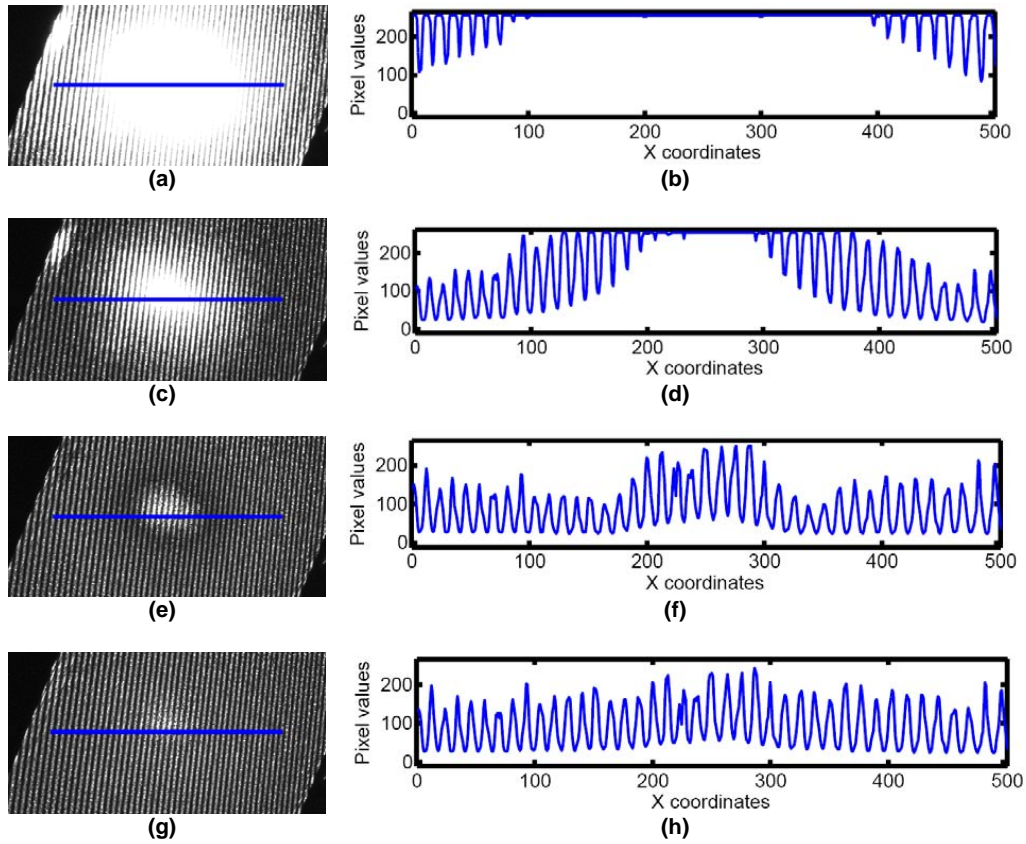
Figure 4.25: Pattern adaption on a metal surface. A planar object with a shiny metal surface is illuminated with a stripe pattern. Iterative adaption of the projector signal decreases the amount of reflection ((a), (c), (e), (g)). The decrease of saturated regions is clearly seen along the intensity profile ((b), (d), (f), (h)) [62].

Object	σ [mm]	σ_{HDR} [mm]	σ_{PAT} [mm]
Chromed	0.039	0.035	0.021
Anodized	0.012	0.010	0.012
Black	0.017	0.016	0.020
Checkerboard	0.022	0.014	0.012

Table 4.5: Structured light results using raw image acquisition, HDR imaging and pattern adaption. Planar objects with different surface properties were reconstructed. The numbers give the standard deviation in mm for all reconstructed points from the best-fit plane.
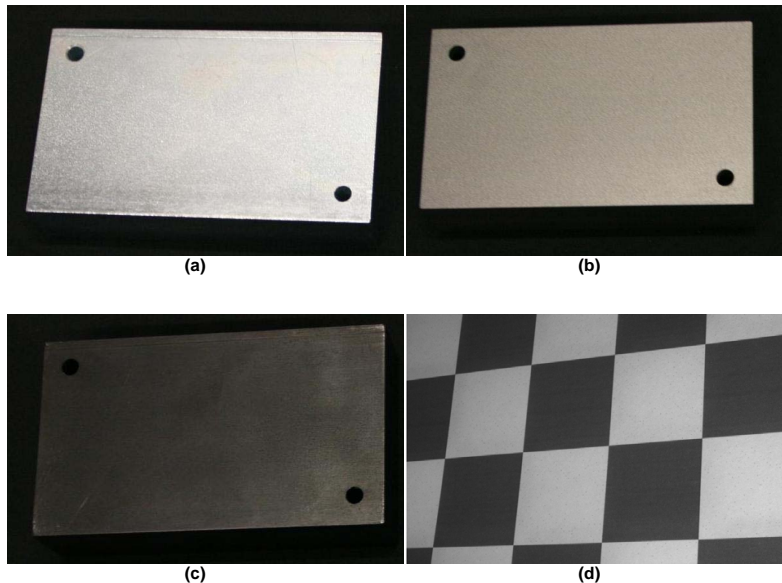
Figure 4.26: Reference objects for pattern adaption. A chromed metal object (a), an anodized object (b), a black-finished object (c) and a printed checkerboard (d) were used to evaluate pattern adaption [62].

Experiment	σ_1 [mm]	sum_1 [numpoints]	ϕ_1 [deg]
Standard	5.3	29821	99.81
Masked	0.023	1603	89.68

Table 4.6: Light source separation experiment. Without masking out invalid patterns, the top plane in Figure 4.27 shows considerably more point scatter than after masking.

4.4.3.2 Light Source Separation

The ability to mask out interreflections according to section 4.4.2 is evaluated on the test scene shown in Figure 4.27(a). A metal prism is positioned on a metal cube and illuminated by a projector. Directional illumination by a projector causes interreflections, especially from the prism onto the upper cube plane. The goal is to perform a structured light reconstruction of the cube planes in two regions, shown in Figure 4.27(b). Without identifying and masking out interreflection regions, the top plane shows a considerable amount of point scatter, as given in Table 4.6. After masking out invalid regions, the remaining valid measurements on the object surface (Figure 4.27(c)) give more precise results, of course with less point density.

(a)   (b)   (c)
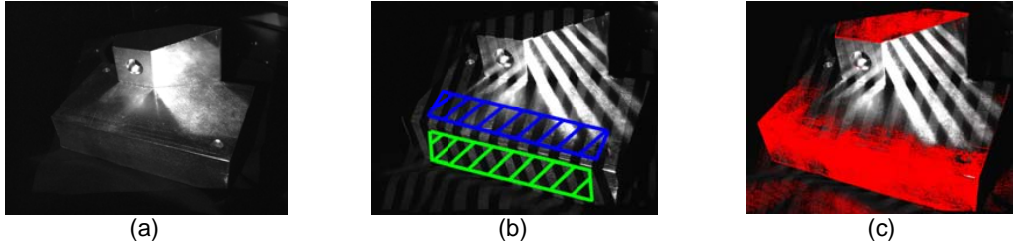
Figure 4.27: Reference scene for light source separation. A metal prism lies on a metal cube (a). Light source separation is performed on the scene. Valid regions are shown in red (c). The planar regions of interest for structured light reconstruction are shown in (b) [62].

4.5   Conclusion

In this chapter, the problem of limited dynamic range in image based measurement is discussed. Saturation effects and low radiometric resolution impose a physical limit to all subsequent image processing tasks. Consequently, artificially increasing the sensor dynamic range can be considered as lowest level image enhancement. On the other hand, methods like the one proposed in Section 4.4 utilize some very high-level concepts like structured light reconstruction to solve this problem.

The HDR imaging uses a temporary sequence of acquired images to cover the scene dynamics. While this method is theoretically appealing - no projector and no offline calibration is required - it has some shortcomings in practice, compared to pattern adaption. In pattern adaption, excessive illumination intensity is avoided, where it is generated, while in HDR imaging, it is always present in the scene and always reaches the camera and sensor chip. Unwanted artifacts are the consequence, e.g. interreflections and stray light in the scene, blooming in the camera and interreflections in the optical path, creating ghost images. Further, acquisition time is a problem. If the goal is a 3D reconstruction of the scene, HDR imaging requires temporal multiplexing to apply structured light, and also for HDR imaging. The number of patterns is multiplied, leading sometimes to a prohibitively large number of required images. Nonetheless, experiments showed that both methods are capable of increasing signal quality for structured light, and image quality for 2D measurements.

Light source separation is a way to effectively cope with interreflections. Image regions showing reflections from multiple sources are practically worthless for robust structured light metrology, so the choice is to identify them and mask them out. In industrial inspection, where objects of rather simple shape are considered, there is a high probability that

interreflections on an object point are caused by only one (planar) object region. It would be an interesting future problem to identify such reflectors and consequently optimize reconstruction quality in a limited region of interest.

Chapter 5

Sensor Geometry

Contents

5.1   Introduction

Measurement accuracy of triangulation based sensors depends on exact knowledge of the radiometric and geometric relations between sender and emitter. This knowledge can either be given by exact machining of the sensor components, by an offline calibration procedure, or by determining these relations online, from the measured data.

Exact machining of sensor components has the advantage that sensors are identical within the given requirements and hence can be freely interchanged during operation or maintenance. High production costs, sensitivity to environmental conditions like temperature, and the inability to recover from even small geometric defects make this approach less favorable in practice.

Offline calibration is by far the most widely applied approach to increase sensor accuracy at tractable costs. Each sensor undergoes a defined calibration procedure, either directly after production or during installation. The acquired calibration data becomes part of the sensor and is tightly coupled to it, usually by storing it internally. Hereby, a sensor is able to recover from minor defects by recalibration as part of a maintenance procedure.

Self calibration approaches have the advantage that calibration data is extracted from measurement data itself. Hence, there is no time difference between measurement and calibration and the sensor does not need to remain consistent with offline data. On the downside, self calibration results, and consequently measurement accuracy, depend by definition on the available measurement data, so no accuracy specifications of the sensor can be given without knowing the exact measurement object and measurement configuration.

All these approaches require exact modeling of the underlying sensor system. The parametric model of sensor components and relative geometry has to be general enough to describe the components with sufficient accuracy, but still sufficiently constrained to allow calibration from a small set of reference data.

This chapter is divided into two parts. First, the model applied in this work is motivated, and methods for calibration are reviewed and evaluated. The model is further used for computation of 3D structure from point correspondences. Hereby, two alternative methods are presented, one relying on explicit modeling of camera geometry, the other, being a model-free approach, relying on a dense grid of 3D reference data and corresponding image measurements. It is shown that the model-free approach produces significantly more accurate results, at the cost of higher calibration effort.

5.2   The Multi-View Sensor Model

The multi-view system for the purpose of 3D measurement, as it is presented here, consists of a number of cameras and projectors. The theory of modeling a central perspective camera has been covered in Chapter 2. In this chapter, methods for automatic calibration to high accuracy are reviewed. Further, the consistency of an off-the-shelf projector system with the camera model is experimentally evaluated. The complete calibration framework, consisting of monocular camera calibration, projector calibration and multi-view calibration, is presented and experimentally evaluated.

5.2.1   Monocular Camera Calibration

The central perspective camera model, including some type of nonlinear lens distortion, is the de-facto standard model for close-range cameras (see Chapter 2). It is applied in photogrammetry and computer vision to model film cameras, digital consumer cameras and industrial cameras. The necessity of calibrating the geometric configuration of a single

camera is about as old as aerial photogrammetry, where cameras were used as measuring instruments for the first time.

5.2.1.1 Related Work

Monocular camera calibration is an active research topic for over 100 years. A historical survey of methods and their development is given in [21]. The camera and lens distortion model by Brown [15] is, with minor variations, still the fundamental mathematical model.

State-of-the-art camera calibration relies on highly redundant reference information, which is acquired by observing three dimensional reference entities like points or lines. The projective mapping between reference points and their observation is found by optimization of an error function, usually in image space. Recently, little progress has been made in increasing calibration accuracy. Rather, much emphasis has been laid on ease of use, modeling of cheap consumer cameras and non-standard optics like fisheye lenses.

Roger Tsai [102] [101] initiated the shift in methodology from expensive photogrammetric laboratory calibration to more versatile methods. In two variants the camera either observes a planar or non-planar reference grid. In a two-stage procedure, the relevant calibration parameters are estimated, which are the effective focal length, one radial distortion coefficient and the camera extrinsics.

Heikkilä and Silveri presented a method for camera calibration from 3D reference points and a single observation [45] [46]. The projection model included focal length, principal point, radial and tangential lens distortion. They addressed the problem of feature location under perspective distortion, especially for circle centers and proposed an iterative procedure, where feature locations are re-estimated once an initial calibration and camera pose is known.

Sturm and Maybank proposed a calibration framework closely related to self-calibration techniques [98]. They rely on reference information from known points on a planar target, and are flexible in the projection model, where parameters can be known a priori and others may vary from one observation to another. Degenerate cases under certain configurations of the projection model are discussed.

The method of Zhengyou Zhang [116] arguably reached the highest popularity in the computer vision community. Using several (at least two non-degenerate) observations of a planar checkerboard pattern with known corner coordinates, the projection model is determined in a two stage approach. At first, the linear intrinsic parameters are estimated using constraints on the pattern to image homography, followed by an estimation of the

nonlinear distortion parameters and a complete `MLE`.

Kannala and Brandt give an overview of the state-of-the-art in fisheye lens calibration and propose a generic projection model for these types of lenses [59]. They address the problem that cameras with a 180 degree field of view can not be modeled simply by a distorted pinhole model. So, while in the pinhole model the radial distance of an image point from the image center is a function of $tan(\theta)$, where θ is the incoming ray angle, the authors model the radial distance as a Taylor series expansion of θ. Additionally a calibration procedure relying on a planar reference target is presented.

Wang et al. proposed a new projection model for pinhole cameras to handle the various distortion effects [108]. Instead of modeling decentering distortion, thin prism distortion and skewness, they introduce an out-of-plane rotation of the sensor plane with respect to the ideal image plane. The model has fewer parameters than the conventional approach and experiments showed comparable results in distortion correction.

A lot of research effort has been put into method evaluation and quality studies. Salvi et al. gave a non-complete survey of calibration methods [94]. They tried to unify the different notations used by the authors and compared the calibration methods on a common test set. The findings confirm the general assumption that nonlinear methods are more accurate than linear ones. Gonzalez et al. evaluated calibration methods in terms of stability and accuracy [37]. Their results mainly confirm the fact that highly correlated parameters, like focal length and translation in viewing direction, become unstable if their counterpart is unstable. Furthermore, according to their findings, calibration results highly depend on the choice of reference views.

Kopparapu and Corke examined the effect of measurement noise on the calibration parameters [65]. Their results show that Gaussian measurement error almost linearly propagates to Gaussian error in the calibration parameters, except for the principal point, which turned out to be more sensitive to noise levels above 0.5pixels. However, the authors did not do a `MLE` of camera parameters, which should cope well with Gaussian noise.

Mallon and Whelan studied biasing of feature extraction methods due to perspective, lens distortion and defocusing effects [73]. Four popular extraction methods, namely blob centroid, ellipse center, saddle point and edge fitting are evaluated. The authors show that distortion bias, followed by perspective bias, are the main sources of error in feature extraction and that edge fitting and saddle point methods are more robust than ellipse based methods.

Remondino and Fraser evaluated camera calibration methods against photogrammetric

bundle adjustment [90]. They showed that the linear direct linear transform (DLT) method performs worst, while the nonlinear methods give comparable results in terms of image space residual error. The authors' conclusion that bundle adjustment methods are superior due to the lower 3D RMS error is questionable, since no ground truth evaluation of the actual 3D points has been conducted.

Sun did an empirical comparison of camera calibration methods [99] [110]. The plane based calibration methods showed to be superior in versatility with a low calibration error. Yet, no evaluation on accurate ground truth data was given.

5.2.1.2   Methodology

Feature points extracted from images of a planar reference target form the input data for intrinsic camera calibration. A planar reference has been chosen, because it gives a viable compromise between accuracy and ease of implementation, although more care has to be taken in choosing camera poses, as will be discussed at the end of the section.

A projective model as described in Chapter 2.3 has been adopted to describe the world to image mapping. An empirical evaluation showed that two radial and two tangential lens distortion parameters, aspect ratio of one and zero skew are sufficient to model the cameras.

Because the projective mapping, including distortion, is nonlinear, parameter estimation leads to a nonlinear least squares problem. Such a method requires a good initial guess of the solution, for not to get stuck in a local minimum.

Initialization requires the determination of camera intrinsics and the relative pose of the cameras with respect to the reference target. Initialization of the intrinsic camera parameters can be done automatically using images of orthogonal vanishing points, which can be easily calculated from a square grid of planar reference points. Setting the principal point initially to the image center, assuming zero skew, square pixels, no distortion and a canonical camera pose (projection center at origin looking in positive z-direction), the projection model is reduced to

$$\mathbf{x} = \begin{bmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix}, \tag{5.1}$$

where $\mathbf{x}$ is the image point, f is the focal length and $[X_1 X_2 X_3]^T$ are the first three coordinates of the world point. Inversely, the direction vector d of a backprojected ray

$[x_1 x_2 x_3]^T$ becomes

$$
\mathbf{d} =
\begin{bmatrix}
f^{-1} & 0 & 0 \\
0 & f^{-1} & 0 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x_1 \\
x_2 \\
x_3
\end{bmatrix}
\tag{5.2}
$$

.

Now, if the images $\mathbf{v_1}$, $\mathbf{v_2}$ of at least two orthogonal vanishing points are available, their backprojected ray directions also have to be orthogonal, leading to the constraint

$$
\begin{bmatrix}
f^{-1} & 0 & 0 \\
0 & f^{-1} & 0 \\
0 & 0 & 1
\end{bmatrix}
\mathbf{v_1}
\begin{bmatrix}
f^{-1} & 0 & 0 \\
0 & f^{-1} & 0 \\
0 & 0 & 1
\end{bmatrix}
\mathbf{v_2} = 0
\tag{5.3}
$$

and simplified

$$
\begin{bmatrix} f^{-2} & f^{-2} & 1 \end{bmatrix} \mathbf{v_1} \mathbf{v_2} = 0,
\tag{5.4}
$$

which can be solved for f. This method is fairly robust and works for a large variety of cameras. However, when large distortions occur in the images, or the principal point is located far off the image center, the initialization may fail, sometimes leading to a complex solution for the focal length. In an industrial environment, where the sensor geometry is approximately known in advance, the parameters may be initialized straightforward from the sensor specification sheet.

For parameter refinement, assuming normal distribution of the residuals, a MLE problem can be formulated, according to Equation 2.1:

$$
\widehat{\mathbf{\Phi}} = \arg\min_{\Phi} \sum_{i=1}^{N} (\mathbf{x}_{img}(\mathbf{\Phi}) - \mathbf{x}_{meas})^2,
\tag{5.5}
$$

where $\mathbf{x}_{meas}$ denotes the measured feature coordinates, N denotes the number of measured reference points and $\mathbf{\Phi}$ denotes the unknown parameter vector. This is equivalent to the classical bundle adjustment problem, where $\mathbf{\Phi}$ holds the camera intrinsics and extrinsics, or, in the case of unknown reference geometry, also holds reference point locations.

Assuming that the reference target is planar and the in-plane reference point locations are known, a reference coordinate system O_{ref} is assigned to the target. As a consequence, the projection of X_{ref} to an image point x_{img} is formulated as follows:

$$\mathbf{x}_{img} = \mathbf{K}\mathcal{F}_{lens}([\mathbf{I}|\mathbf{0}]\mathbf{H}_{world\rightarrow cam}\mathbf{H}_{ref\rightarrow world}\mathbf{X}_{ref}), \tag{5.6}$$

where $\mathbf{H}_{ref\rightarrow world}$ is a rigid transformation from the reference CF to the WCF, and $\mathbf{H}_{world\rightarrow cam}$ is a mapping from world to CCF, resembling the camera extrinsics. Consequently, for each camera the set of intrinsic and extrinsic parameters has to be estimated, and for each observation the unknown rotation and translation in $\mathbf{H}_{ref\rightarrow world}$ has to be estimated. The number of unknowns n_Φ sums up to $n_\Phi = 14n + 6m$, where n is the number of cameras and m the number of observations, whereas the number of equations n_{equ} grows approximately with $n_{equ} = n_{pt}nm$, n_{pt} being the average number of observed reference points per camera.

The large degree of redundancy makes the calibration process more robust and computationally tractable than in the complete bundle adjustment case. A nonlinear least squares solver, such as the LM algorithm, can be used to compute a ML estimate of the solution, making use of the sparseness of the system Jacobian.

There are, however, some issues influencing calibration accuracy. By using a planar reference target, all camera poses relative to the reference plane and the camera intrinsics need to be estimated simultaneously. Therefore, one needs to take care, what sensor poses to choose in order to reduce correlations between extrinsic and intrinsics parameters. Usually a tradeoff has to be found between extremely slanted camera views, which considerably improve the geometric configuration, and frontoparallel views, which improve visibility of the feature pattern and give increased segmentation accuracy. In Section 5.3.5, an experimental comparison of plane-based calibration and calibration from a three-dimensional reference target is given.

5.2.2 Projector Calibration

The optical setup of a DLP projection system is comparable to that of a central perspective camera. A point light source illuminates a micro-mirror array situated in the focal plane of the optical system. The amplitude of reflection is modulated by the mirrors and passes through a lens system.

This leads to the assumption that the projector can be modeled as an inverse camera system, which backprojects image points to light rays according to the central perspective projection model. If the same projection model is applicable, also the same calibration methods as in Section 5.2.1 can be used. In contrast to camera calibration however, it is more difficult to generate the necessary reference information. Reference information is

based on some kind of known 3D structure, directly related to identifiable features in image space. In the case of forward projection, this correspondence is generated relatively easy, as salient 3D points directly project to image features. The inverse case is harder, because 3D reference information must be assigned to given image features. One option would be to align a projected pattern to some known reference target, for which an accurate reference target would be required, and an external measurement device, which measures the misalignment between projected pattern and reference. The measurement device only performs relative measurement, so it needs to be precise, but does not have to give absolutely accurate measurements. Typically, a camera system is used for this purpose, as it is anyway required for reconstruction purposes in a structured light measurement setup. It is clear that the residual alignment error directly affects the calibration result, whereas inaccurate camera calibration does not, due to the relative measurement. Another option is to impose a number of constraints on the projected features instead of generating 3D reference information.

The method applied in this work is related to the second option. Because a camera network is available anyway, it is possible to do a 3D reconstruction of projected features on reference objects, without knowing the projector calibration or the object structure. This way, 3D reference information is assigned to projected image features and the methodology of the previous chapter can be applied, resulting in a bundle adjustment problem where the reprojection error in the cameras and the projector are minimized in a least squares sense. The parameter set to be optimized contains the scene structure, as well as the projector intrinsics and extrinsics. The camera parameters are kept constant, because these are already optimized with respect to known 3D reference information. It is clear that camera calibration inaccuracies adversely affect the quality of the projector calibration. On the other hand, no external reference is required, which makes the proposed method flexible and contributes to the practical applicability.

5.2.3   Experiments

Two questions need to be answered when analyzing camera calibration results. First, the consistency of the applied projection model with the physical imaging process needs to be evaluated, and second, precision and accuracy of the calibrated model needs to be evaluated. Although related, these two questions cannot be answered in one go, because of differences in the degree of redundancy. A projection model with more DOF might give more precise calibration results, while describing the projection process less accurately,

due to overfitting to the available reference information.

The first question, regarding consistency of the projection model, arises primarily in monocular camera calibration, due to the freedom of choice in the lens distortion model. Here, a number of experiments have been conducted with varying number of distortion parameters, as presented in the first section. Second, after motivating the choice of a suitable distortion model, the geometric uncertainty under measurement noise in image space is evaluated.

The experimental setup is sketched in Figure 5.1. Four cameras observe a common volume of $150 \times 150 \times 40\text{mm}^3$. The cameras are positioned at a distance of roughly $\Delta z = 180\text{mm}$ to the measurement volume. The camera viewing angles relative to the volume z-direction range between $\vartheta = 15°$ and $\vartheta = 30°$. Exact values are given in Table 5.1.
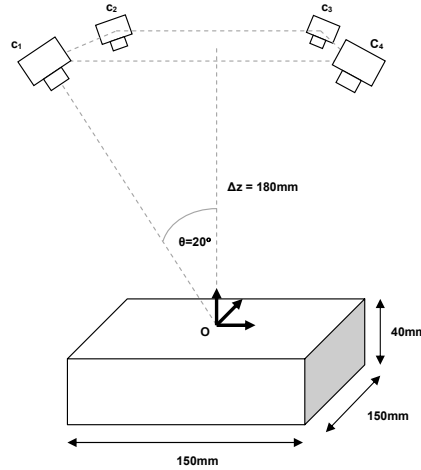
Figure 5.1: Experimental setup. Four cameras observe a common volume of $150 \times 150 \times 40\text{mm}^3$ from a distance of roughly 180mm. The cameras are manually mounted and aligned so exact geometric relation may vary.

Ground truth information was generated, using a planar glass target with circular reference markers, aligned in a rectangular grid (grid-spacing is 2mm). The markers have a nominal positioning accuracy of 900nm. To generate 3D reference, the target was moved vertically in 2mm steps, using a mechanical stage with nominal accuracy of 1μm (see Figure 5.2).

As such, a nearly rectangular reference grid $\mathbf{R}$ of 59250 points has been generated, covering a volume of $144 \times 132 \times 40\text{mm}^3$. The fact that the reference grid is not exactly rectangular, results from a slight misalignment of the stage motion

Camera	Δz [mm]	ϑ [deg]
1	185.27	16.95
2	183.56	29.26
3	174.96	24.07
4	176.28	20.16

Table 5.1: Vertical distance and angle for the four camera views. According to Figure 5.1, the vertical offset Δz and angle ϑ between viewing direction and vertical axis are given.
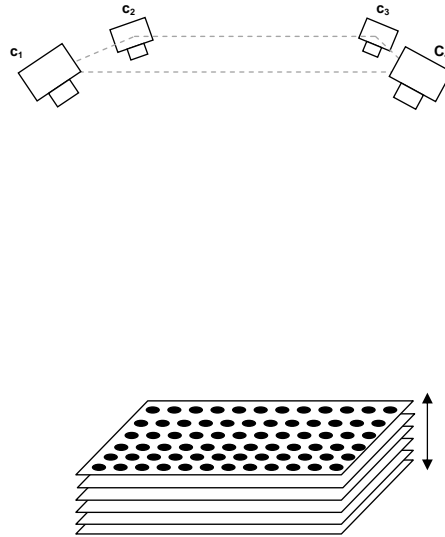
Figure 5.2: Acquisition of ground truth data. A planar reference target is moved vertically to cover the measurement volume. For evaluation, the target is also shifted horizontally, to generate different points for testing.

direction to the glass-target normal direction. The y-motion direction in the target reference frame was $\mathbf{d}_y = (-0.27e^{-3}, 1, 0.16e^{-3})^T$, and the z-motion direction $\mathbf{d}_z = (-4.49e^{-3}, -0.57e^{-3}, 999.99e^{-3})^T$.

To provide reference information for projector calibration, another set of images was acquired, exchanging the reference pattern by a planar, white surface and moving it in 2mm intervals in z-direction. A DLP projector (800×600pixels) was used to project the reference pattern onto the surface. A total of 16 calibration images were acquired.

Lens Model	Residual Err. [pixels]	f error [%]	T error [mm]
1rad	$(0.13, 0, 12)$	0.09	$(0.03, 0.08, 0.2)$
2rad	$(0.05, 0.06)$	0.04	$(0.01, 0.03, 0.09)$
2rad2tan	$(0.04, 0.05)$	0.04	$(0.01, 0.03, 0.08)$

Table 5.2: Residual error after monocular calibration. Different lens distortion models have been used. One radial coefficient (1rad), two radial coefficients (2rad) and two radial with two tangential coefficients (2rad2tan). The residual reprojection error is given, as well as the propagated relative 3σ error of focal length, and the mean translational error relative to the reference target.

5.2.3.1   Monocular Camera Calibration

Monocular camera calibration has been performed on four cameras, using the 21 distinct planes of the reference data (see previous section) as calibration normals. The relative plane poses are assumed to be unknown, as well as the relative camera poses and camera intrinsics. Residual error in image space, as well as propagated 3σ bounds on the intrinsic parameters were used as precision measures. Because no ground truth information on camera and lens parameters is available, a ground truth evaluation has been done on the 3D reference points, by comparing points reconstructed from the calibration data to the ground truth points.

Monocular calibration results for a single camera (1680×1220pixels, CCD, $f = 8$mm lens), are given in Figure 5.3 and Table 5.2. Results are given for different lens distortion models, one radial component only (Figures 5.3(a), (b)), two radial components (Figures 5.3(c), (d)), and two radial plus two tangential components (Figures 5.3(e), (f)).

Figures 5.3(a),(c),(e) show the lens distortion curve along the image radius, measured from the principal point as the image center. The curve is normalized to a maximum of one, and reprojection errors of all image points are given as a function of radius. For a single radial distortion component, a systematic characteristic of reprojection errors is visible in the images. By adding a second radial distortion component, the residual error is distributed more or less randomly, with small residual systematic component, which is obviously not radially symmetric. Two additional tangential distortion parameters give a minor improvement in precision, as shown in Table 5.2.

(a)

(b)

(c)

(d)

(e)

(f)
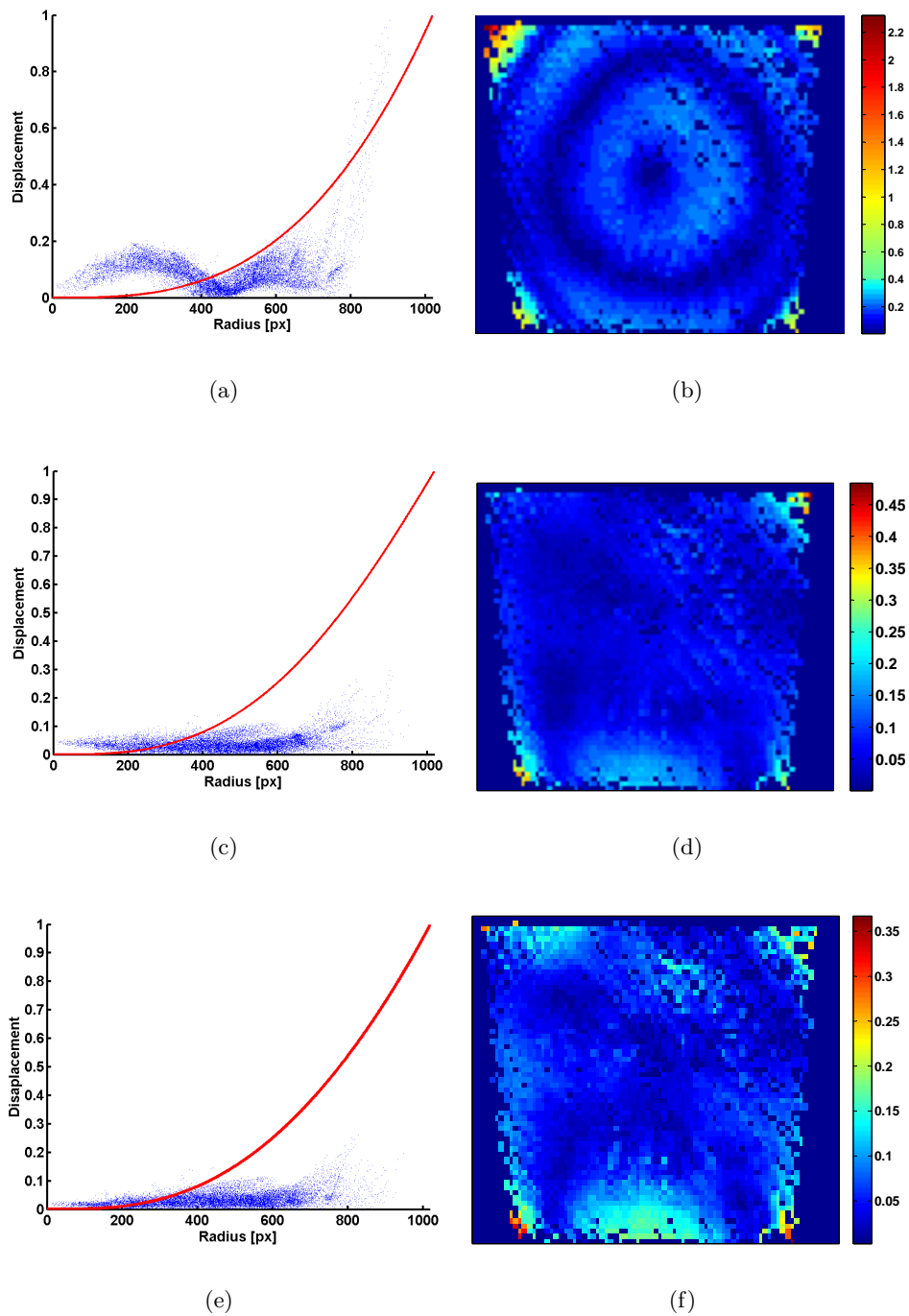
Figure 5.3: Residual error after camera calibration. One radial component (a)(b), two radial components (c)(d), two radial and two tangential components (e)(f). The left figures show the distortion curve, normalized to a maximum of one, and the residual reprojection error (blue dots) for all reference points over the image radius. In the right figures, the distribution of residual errors over the image area is shown.

Camera	Residual Err. [pixels]	f error [%]	T error [mm]
1	$(0.038, 0.041)$	0.005	$(0, 0, 0)$
2	$(0.038, 0.048)$	0.005	$(0.005, 0.007, 0.017)$
3	$(0.039, 0.049)$	0.0047	$(0.004, 0.007, 0.015)$
4	$(0.038, 0.037)$	0.0048	$(0.004, 0.003, 0.014)$

Table 5.3: Calibration result after bundle adjustment. The error figures include mean reprojection error, relative focal length error, and translational errors of the projection centers w. r. t. the first camera.

5.2.3.2 Multi View Calibration

Monocular camera calibration was done for our cameras on the same reference information, from different view angles. All precalibrated cameras have been refined together in a bundle adjustment procedure. Multi-View calibration results are given in Table 5.3, in terms of accuracy of reference points and reprojection error. Results for intrinsic parameters are comparable to the monocular results.

5.2.3.3 Projector Calibration

(a)   (b)
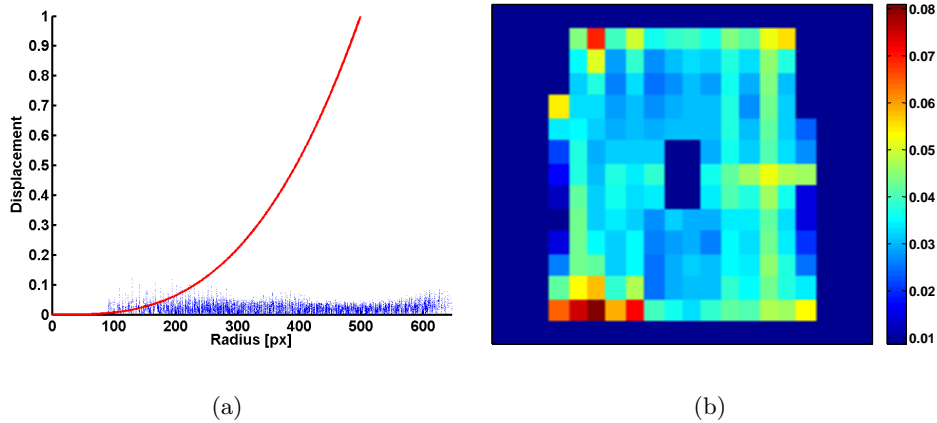
Figure 5.4: Residual error of projector calibration. The DLP projector has been calibrated under the assumption, that its optics comply with the projection model introduced in Section 2.3, using one radial component and two tangential distortion components (c)(d). The left figure shows the distortion curve, normalized to a maximum of one, and the residual reprojection error (blue dots) for all reference points over the image radius. In the right figure, the distribution of residual errors over the image area is shown.

Features of the projected calibration pattern were reconstructed using the previously

Lens Model	Residual Err. [pixels]	f error [%]	T error [mm]
2rad2tang	$(0.04, 0.02)$	0.013	$(0.01, 0.02, 0.04)$

Table 5.4: Residual error after projector calibration. Two radial with two tangential coefficients were used to model lens optics. The residual reprojection error is given, as well as the propagated relative 3σ error of focal length, and the mean translational error relative to the reference target.

calibrated multi-view setup. Calibration results and precision in terms of reprojection error are given in Table 5.4. The projector's viewing frustum is essentially a half cone, so the principal point lies in the lower image area, at $y = 647$pixels on an 800×600pixels projector image. At 0.04pixels, the reprojection error is moderately low, so the camera model is consistent with the projector hardware. In Figure 5.4, the reprojection error is shown over the radius. Errors are evenly distributed over the radius, which also suggests consistency. Due to the high pixel size of the projector (13μm compared to 4.8μm of the camera), the metric error of the projection unit is considerably higher.

It is important to note that a slight aspect ratio of the projector pixels could be noticed. At 1.04pixels, this aspect can not be neglected and must be included in the parameter set.

5.2.4   Conclusion

Calibration of multiview systems and projector-camera systems have been studied extensively before. In this work it is shown, how such a sensor can be calibrated to high accuracy with a minimum of external devices. Using a planar reference pattern, and an approximately planar, white board for projector calibration, the entire system can be calibrated bottom up, a property which is important especially in industrial, day to day use. A robot mounted sensor may recalibrate itself as part of a daily maintenance process and continuously check for calibration consistency on the reference target.

The often tricky part of projector calibration is solved by providing geometric references by the camera-system. The calibration process hereby becomes simple, because the projector is treated like any other camera, but inaccuracies in determining the reference points influence projector calibration accuracy. Nonetheless it was shown that an off-the-shelf DLP projector is consistent with the central perspective camera model and applicable also to high accuracy measurement tasks.

5.3 Structure Estimation

The problem of structure estimation is to identify the set of world points, which projects to sets of measured and corresponding points in two or more views. Due to measurement noise and other factors, the projection will not be exact (Figure 5.5), leading to an estimation problem.

Traditionally, the error measure is defined in image space, either as an algebraic error or as Euclidean reprojection error. The projection mapping from $\mathbb{P}^3$ to $\mathbb{P}^2$ is defined by the central perspective camera model as described in Chapter 2.3. Least squares estimation leads to a solution for the three DOF per world point.
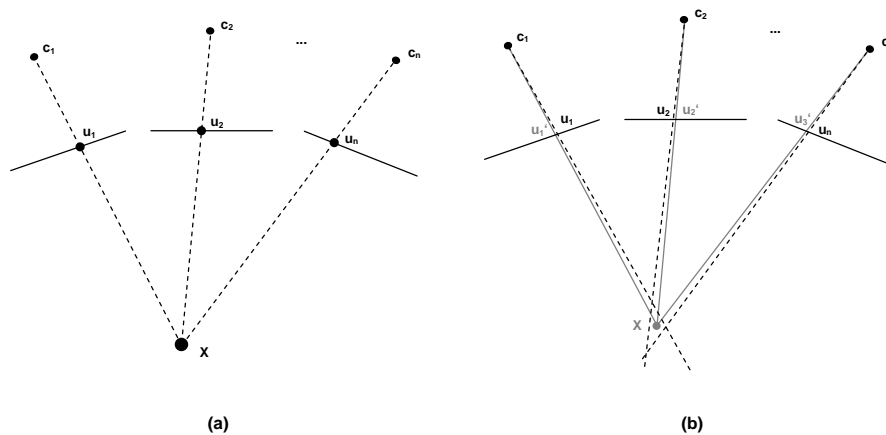
Figure 5.5: Triangulation in a multi-view setup. In the noise-free case (a), all image rays intersect exactly in the respective world point. Noisy image features, or systematic errors in image space (b), make an exact intersection infeasible.

This approach has two shortcomings in terms of accuracy. First, the central perspective camera model will never be complete. Effects like depth dependent lens distortion, image plane unflatness, and optical effects other than decentering or radial distortion are hard to model and calibrate. As a consequence, back-projected world rays will be affected by segmentation errors, as well as systematic errors originating from overly restricted camera models.

The second problem arises in multi-view configurations with more than two cameras. Redundancy in the number of views clearly increases reconstruction accuracy, yet, in practice, a world point will not necessarily be visible in all views. A partial occlusion causes a considerable degradation of reconstruction accuracy, especially in least squares estimation. The problem is sketched in Figure 5.6. If, due to occlusion, image ray n is
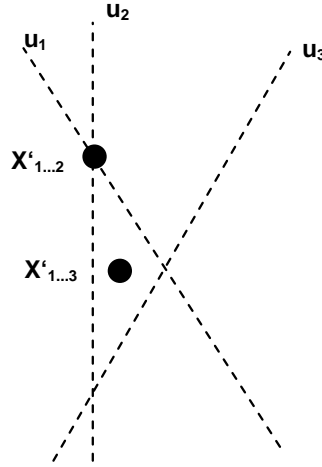
Figure 5.6: Effect of occlusion. If $\mathbf{u}_3$ is omitted from the triangulation, the estimated point $\mathbf{X}'$ might experience considerable shift.

omitted, the reconstructed world point X will move towards the intersection of rays 1 and 2.

In this section, a different approach to structure estimation is introduced, which is especially applicable for rigid multi-view setups which observe a limited volume. The mapping from the scene space to image space is no longer described by an algebraic model, but by a dense grid of reference world points and their projections in each camera view. Intuitively it is obvious that this information implicitly encodes all systematic errors of the projective mapping. A world point is expressed through barycentric coordinates with respect to its closest reference points. It is shown in Section 5.3.1 that central perspective projection of barycentric coordinates does not involve any intrinsic camera parameters. As a consequence, the structure estimate is directly related to the reference point and not influenced by model parameters. An obvious drawback of the approach is the necessary effort for calibration.

After introducing the classical triangulation approach and barycentric reconstruction, it is shown that the barycentric approach gives considerably higher measurement accuracy, and that accuracy remains high under partial occlusions. Further, it is shown that the proposed method is robust, even if the reconstructed point lies at the border or slightly outside its enclosing reference points. As an alternative to least-squares estimation, barycentric structure estimation may be formulated as a quasi-convex problem, where a global solution can be found with respect to the L_∞ norm. Finally, a concluding experimental comparison between the different reconstruction approaches is given.

5.3.1 Barycentric Coordinates

Expressing a point inside a polyhedron as a weighted combination of the vertices of the polyhedron is a fundamental problem in computational geometry. Applications include surface parametrization, free-form deformation and interpolation.

In particular, given a convex polytope P with vertices $\mathbf{v}_i$, an interior point $\mathbf{x}$ may be written as a convex combination [55]

$$\sum_i b_i \mathbf{v}_i = \mathbf{x} \qquad \text{where} \qquad b_i \geq 0 \qquad \text{and} \qquad \sum_i b_i = 1. \qquad (5.7)$$

The weights b_i are called barycentric coordinates of $\mathbf{x}$ on polytope P. Barycentric coordinates are easily constructed on a simplex (triangle in $\mathbb{R}^2$ or tetrahedron in $\mathbb{R}^3$). For each $\mathbf{v}_i$, one calculates the volume of the simplex formed by $\mathbf{x}$ and the opposite face of $\mathbf{v}_i$, divided by the total volume of P. An example in 2D is shown in Figure 5.7, where P is a triangle, defined by points $v_1 \ldots v_3$.
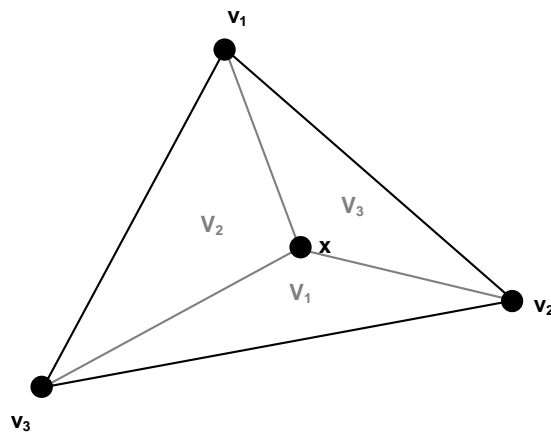
Figure 5.7: Classical barycentric coordinates in a 2D triangle.

This construction has some drawbacks. It is not defined outside P, and it can not be generalized to more complex polytopes, because the opposite face is no longer defined.

Wachspress [106] introduced a new barycentric coordinate representation, which is valid also for convex polygons in 2D. Warren [109] later generalized the concept to convex polytopes of any dimension. The resulting functions are smooth, positive inside P and conicide with multilinear Cartesian interpolation. Yet, no practical formulation for the construction of Wachspress coordinates in dimensions higher than $\mathbb{R}^2$ have been given.

Floater [27] introduced mean value coordinates in 2D to express points in the kernel of a

star-shaped polygon. He later generalized the concept to convex polyhedra and the kernels of star shaped polyhedra [28]. Ju et al. [54] generalized mean value coordinates from closed 2D polygons to closed triangular meshes. They showed that mean value coordinates are smooth inside P and continuous everywhere. In 2007, Ju et al. [53] provided a unified geometric framework to construct Wachspress coordinates, mean value coordinates and discrete harmonic coordinates in a convex 2D polygon and 3D triangular polyhedra.
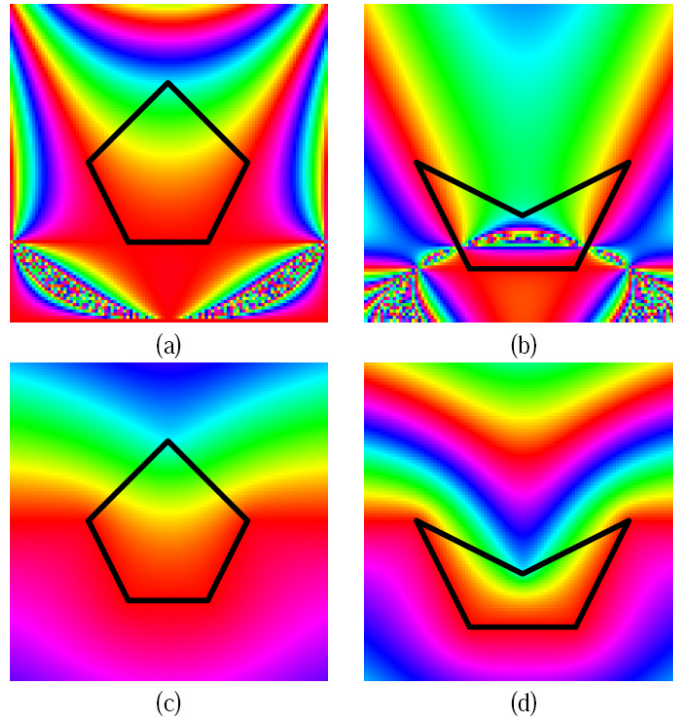
Figure 5.8: Planar interpolation using Wachspress coordinates (a), (b), and mean value coordinates (c), (d) [54].

In Figure 5.8, the interpolation of hue values around a convex and concave polygon are shown.

In order to construct mean value coordinates for a point $\mathbf{X}$ in a closed triangular mesh, one first computes the mean vector for each spherical triangle (i.e. the projection a mesh triangle onto the unit sphere centered at $\mathbf{X}$):

$$\mathbf{m} = \sum_i \frac{1}{2} \theta_i \mathbf{n}_i. \tag{5.8}$$

The spherical triangle is hereby sketched in Figure 5.9.

Weights per triangle vertex are further calculated as
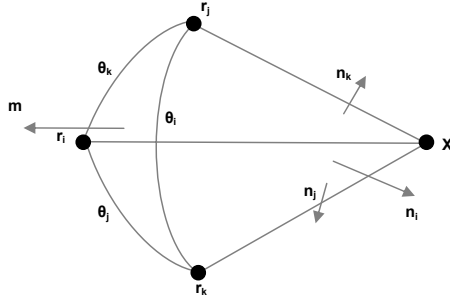
Figure 5.9: Spherical triangle. Normals $\mathbf{n}_i$ and angles θ_i used for mean vector computation.

$$\mathbf{w_i} = \frac{\mathbf{n}_i^T \mathbf{m}}{\mathbf{n}_i^T (\mathbf{r}_i - \mathbf{X})}. \tag{5.9}$$

This construction method may become numerically unstable if the area of the spherical triangle becomes small, i.e. the mesh triangle plane is nearly parallel to the mean vector. For a more robust construction scheme, the reader is referred to [54].

5.3.2   Triangulation

A multi-view camera system, as depicted in Figure 5.5(a), is modeled by a number of central perspective cameras $\mathbf{c}_i$, $i = 1 \ldots n$. A single camera $\mathbf{c}$ is described by intrinsic parameters, lens distortion parameters and extrinsics, which relate the camera poses to a common WCF C_w. During triangulation, the goal is to estimate the location of a world point $\mathbf{X}$ from a set of known projections $\mathbf{u}_1, \ldots, \mathbf{u}_n$. If lens distortion has been removed, the problem may be solved using DLT, resulting in a linear equation system of the form:

$$\begin{bmatrix} [\mathbf{u}_1]_\times \mathbf{P}_1 \\ \vdots \\ [\mathbf{u}_n]_\times \mathbf{P}_n \end{bmatrix} \mathbf{X} = \mathbf{0}, \tag{5.10}$$

where $[a]_\times$ designates the cross-matrix of a, and $\mathbf{P}_i$ is the i-th projection matrix. Due to measurement noise and systematic distortion effects, it will not be feasible to find an exact solution (see Figure 5.5(b)). Rather, a linear least-squares approximation will be sought.

A geometrically more intuitive problem is given by least-squares minimization of the reprojection error in image space, which leads to a bundle adjustment problem of the form

$$\hat{\mathbf{X}} = \arg\min \sum_{i=1}^{n} d(\mathbf{P}_i \mathbf{X}, \mathbf{u}_i)^2, \tag{5.11}$$

where $d(\mathbf{a}, \mathbf{b})$ designates Euclidean distance between vectors $\mathbf{a}$, $\mathbf{b}$. Other possible error measures include angular error between projected rays and measured rays, or orthogonal distance between $\mathbf{X}$ and projected rays.

5.3.3   A Least-Squares Model-Free Approach

A rigidly connected multi-view system, consisting of at least two cameras, can only observe a limited volume in world space. We propose to observe a dense three-dimensional grid of reference points, equally distributed in this volume. The reference point locations $\mathbf{r}_i$, $i = 1 \ldots n$, as well as their projections in each camera view $\mathbf{v}_{i,j}$, $i = 1 \ldots n$, $j = 1 \ldots m$ are saved.
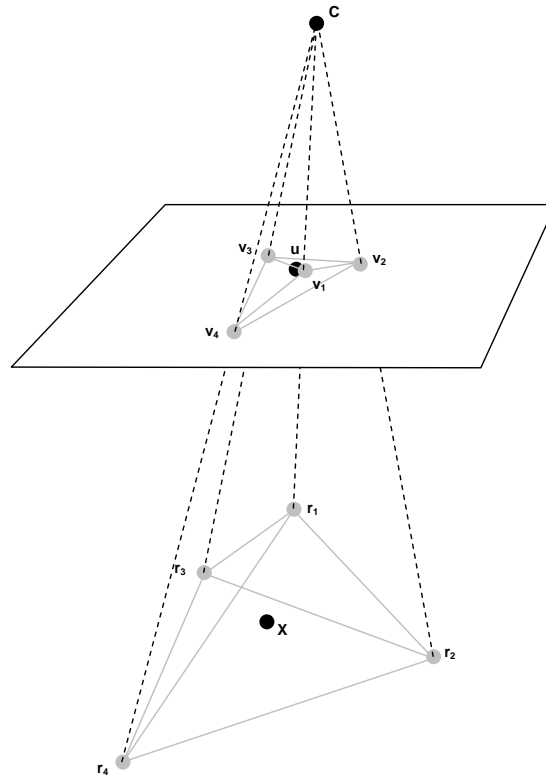

Figure 5.10: Projection of 3D barycentric coordinates. A world point X can be expressed as a linear combination of a set of enclosing points r_i. The corresponding image point u can again be expressed as a linear combination of v_i, the projections or r_i.

To project a world point $\mathbf{X}$ to image space, first the closest set of enclosing reference points $\mathbf{r}_i$, $i = 1 \ldots k$ is searched. The location of $\mathbf{X}$ is then expressed by barycentric coordinates $\mathbf{W} = (W_1, \ldots, W_k)^T$, such that $\mathbf{X} = [\mathbf{r}_1, ..., \mathbf{r}_k] \, \mathbf{W}$. The 3D barycentric coordinate vector $\mathbf{W}$ is then projected to a camera view as $\mathbf{w}_j = (w_{1,j}, \ldots, w_{k,j})^T$, where the Euclidean image point is reconstructed as $\mathbf{u}_j = [\mathbf{v}_{1,j}, ..., \mathbf{v}_{k,j}] \, \mathbf{w}_j$. The geometric relations are sketched in Figure 5.10. Hereby, the important observation is made, that $\mathbf{W}$ projects to $\mathbf{w}_j$ without knowledge of the intrinsic camera parameters, as will be shown in Section 5.3.3.1. Structure estimation is further done in analogy to traditional triangulation, by minimizing the reprojection error in image space.

To make this procedure feasible, the following assumptions are made:

- Nonlinear image distortions can be approximated linearly within a small image patch. This is necessary, because nonlinearities are essentially interpolated between reference points.

- Barycentric coordinates need to be well defined inside the set of basis points, at the border, and outside border within a small neighborhood. This is necessary, because an initial estimate for $\mathbf{X}$ may lie close to the border and move outside during optimization.

Both prerequisites can be fulfilled in practice. Smooth behaviour of image distortions is assumed also in the traditional lens distortion models, and generalized 3D barycentric coordinates, which have smooth behavior in border regions have been introduced recently. Further, model accuracy can be controlled by changing the density of the available image points.

5.3.3.1   Barycentric Projection

In the Section 5.3.1 it has been shown that a world point $\mathbf{X}$ can be expressed as the weighted sum of a set of enclosing vertices $\mathbf{r}_i$, as long as these are connected by a closed triangle mesh with no self-intersections. The weight vector $\mathbf{W} = (w_1, \ldots, w_k)^T$ is addressed as barycentric coordinate vector. Among the possible barycentric representations, mean values have been chosen in this work, mainly for ease of construction and smoothness at the border of a polytope P. The general concept in this work however applies to any kind of barycentric representation.

Now, the problem of central perspective projection is addressed. Without loss of generality, a one-dimensional camera with focal length f, as shown in 5.11, is given. The

camera is canonical, with projection center at the world coordinate origin, looking in positive z-direction.
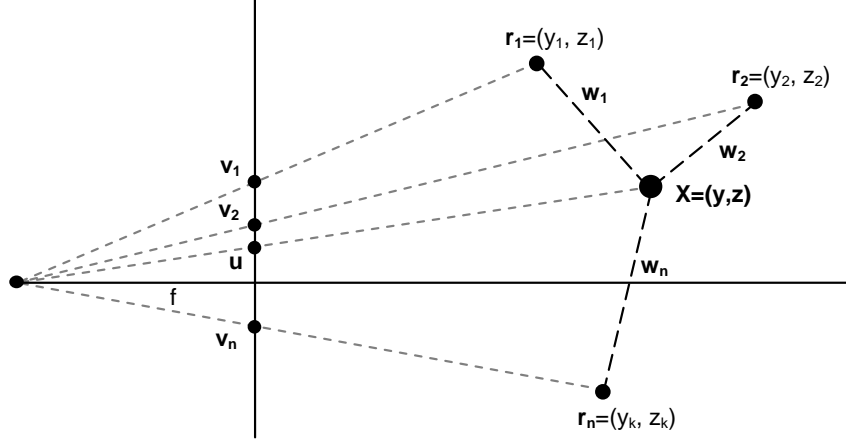
Figure 5.11: Projection of barycentric reference points to image space.

A world point $\mathbf{X}$ in $\mathbb{P}^2$ is expressed in barycentric coordinates $\mathbf{W}$ with respect to a set of enclosing vertices $\mathbf{r}_1, \ldots, \mathbf{r}_k$. The projections of $\mathbf{X}$ and $\mathbf{r}_i$ are given by

$$u = \frac{yf}{z} = \frac{f \sum_i y_i w_i}{\sum_i z_i w_i} \qquad \text{and} \qquad v_i = \frac{y_i f}{z_i}, \tag{5.12}$$

respectively. Now, a barycentric coordinate vector $\mathbf{w} = (w'_1, \ldots, w'_k)^T$ in image space is sought, such that

$$u = (v_1, \ldots, v_k)\mathbf{w}, \tag{5.13}$$

which is equivalent to

$$\frac{f \sum_i y_i w_i}{\sum_i z_i w_i} = f \frac{y_1}{z_1}, \ldots, f \frac{y_k}{z_k}^T \mathbf{w}. \tag{5.14}$$

As a consequence, $\mathbf{w}$ is given by

$$\mathbf{w} = \frac{1}{\sum_i z_i w_i} (z_1 w_1, \ldots, z_k w_k)^T. \tag{5.15}$$

An important observation is that $\mathbf{w}$ does not depend on f, so camera intrinsics are practically eliminated, as long as image skew is zero.

5.3.3.2 Reconstruction

In analogy to model based reconstruction, structure computation based on barycentric coordinates can be formulated as a nonlinear least-squares estimation problem, which minimizes the reprojection error in image space. Given a function $\mathbf{W} = \mathcal{F}_b(\mathbf{X}, \{\mathbf{r}_j, \dots, \mathbf{r}_k\})$, which converts 3D Cartesian coordinates to barycentric coordinates, and a function $\mathbf{w} = \mathcal{F}_p(\mathbf{W}, \mathbf{X})$, which projects 3D barycentric coordinates to their 2D equivalents, a least-squares optimal estimate for $\mathbf{X}$ can be found by

$$\hat{\mathbf{X}} = \arg\min \sum_{i=1}^{n} d(\left[\mathbf{v}_j^i, \dots, \mathbf{v}_k^i\right] \mathcal{F}_p(\mathcal{F}_b(\mathbf{X}, \{r_j, \dots r_k\}), \mathbf{X}), \mathbf{u}^i)^2, \tag{5.16}$$

where n is the number of views and $\mathbf{a}_j^i$ designates the jth vector w.r.t. the ith view.

An initial solution $\mathbf{X}_0$ can be easily found by linear triangulation. An additional problem is the efficient identification of $\{\mathbf{r}_j, \dots, \mathbf{r}_k\}$. In an industrial setting, the reference points may be aligned in a rectangular grid, allowing identification in constant time, once $\mathbf{X}_0$ is known. In an irregular grid, techniques like space partitioning may be applied to find the enclosing references in logarithmic time.

5.3.4 Barycentric Reconstruction Under the L_∞ Norm

The structure estimation problem may be formulated differently, estimating $\hat{\mathbf{W}}$ instead of $\hat{\mathbf{X}}$. This leads to a considerably simpler optimization task. The number of DOF increases with the number of reference points, so at least four DOF need to be estimated in case of an enclosing tetrahedron. Under perspective projection, one DOF is lost due to scale invariance, so w_4 can be fixed to one and $\mathbf{W} = (w_1, w_2, w_3, 1)^T$.

The least-squares optimization problem 5.16 then becomes

$$\hat{\mathbf{X}} = \arg\min \sum_{j=1}^{n} d\left(\frac{\sum_{i=1}^{4} \mathbf{v}_i^j z_i^j w_i}{\sum_{i=1}^{4} z_i^j w_i}, u_i^j\right)^2. \tag{5.17}$$

Note that z_i^j, is also specified w.r.t. a camera view, because it designates projective depth, which is equivalent to the z-coordinate only for canonical cameras.

The error function for a single camera view j can be rewritten to

$$\mathcal{F}_j(\mathbf{W}) = \left\| \frac{(v_{11}z_1, v_{21}z_2, v_{31}z_3, v_{41}z_4)\mathbf{W}}{(z_1, z_2, z_3, z_4)\mathbf{W}} - u_1, \frac{(v_{12}z_1, v_{22}z_2, v_{32}z_3, v_{42}z_4)\mathbf{W}}{(z_1, z_2, z_3, z_4)\mathbf{W}} - u_2 \right\|_2, \tag{5.18}$$

where v_{ik} designates the kth coordinate of the ith reference image point. View index j is omitted for ease of reading. Setting $\mathbf{V}_1 = (v_{11}z_1, v_{21}z_2, v_{31}z_3)^T$, $\mathbf{V}_2 = (v_{12}z_1, v_{22}z_2, v_{32}z_3)^T$, $\mathbf{Z} = (z_1, z_2, z_3)^T$ and rewriting (5.18), gives

$$\mathcal{F}_j(\mathbf{W}) = \frac{\left\| \begin{pmatrix} \mathbf{V}_1^T - \mathbf{Z}^T u_1 \\ \mathbf{V}_2^T - \mathbf{Z}^T u_2 \end{pmatrix} \mathbf{W}_{1\ldots3} + \begin{pmatrix} v_{41}z_4 - u_1 z_4 \\ v_{42}z_4 - u_2 z_4 \end{pmatrix} \right\|_2}{\mathbf{Z}\mathbf{W}_{1\ldots3} + z_4}. \tag{5.19}$$

Equation 5.19 has the form of a second order cone function

$$\mathcal{F}_j(\mathbf{x}) = \frac{\|\mathbf{A}_j \mathbf{x} + \mathbf{b}_j\|_2}{\mathbf{c}_j^T \mathbf{x} + \mathbf{d}_j}, \tag{5.20}$$

which is a quasi-convex function [57]. A global solution to the minimax (L_∞) optimization problem

$$\min_{\mathbf{x}} \max_j \mathcal{F}_j(\mathbf{x}) \tag{5.21}$$

can be found by second order cone programming (SOCP). The equivalent SOCP minimax problem is hereby

$$\min_{\mathbf{x}} \max_j \frac{\|\mathbf{A}_j \mathbf{x} + \mathbf{b}_j\|_2}{\mathbf{c}_j^T \mathbf{x} + \mathbf{d}_j} \quad , \quad \mathbf{c}_j^T \mathbf{x} + \mathbf{d}_j \geq 0, \tag{5.22}$$

which can be solved by bisectioning and solving the SOCP feasibility problem

$$\|\mathbf{A}_j \mathbf{x} + \mathbf{b}_j\|_2 \leq \mathbf{c}_j^T \mathbf{x} + \mathbf{d}_j. \tag{5.23}$$

For a more in-depth treatment of the problem, the reader is referred to [57].

5.3.5   Experiments

To experimentally prove robustness and increased accuracy of the proposed method, a multi-view setup consisting of four cameras has been constructed. Absolute accuracy (i.e. deviation of of reconstructed points to ground truth points) has been evaluated for the triangulation method and the barycentric approach. To evaluate robustness against partial occlusions, the experiment has been repeated by omitting measurements from two views. Further, the barycentric approach allows to incorporate more than four enclosing reference points, which potentially reduces the influence of random noise in the reference image measurements. Reconstruction accuracy was therefore evaluated using an enclosing

Camera	MRE [pixels]
1	0.064
2	0.057
3	0.064
4	0.051

Table 5.5: Calibration result. The four-view camera system was calibrated using the reference grid as ground truth data. Bundle adjustment on camera motion was performed, including lens distortion, to retrieve intrinsic and extrinsic parameters. The MRE for the four views is shown here.

tetrahedron (four points), and an enclosing triangle mesh which approximately forms a cube (eight points).

5.3.5.1 Experimental Setup

The experimental setup is identical to the one used in Section 5.2.3.

For evaluation, another set of 21592 ground truth points $\mathbf{E}$, covering a volume of $144 \times 118 \times 1.2\text{mm}^3$ inside the reference volume was generated. The evaluation points were laid out within two z-layers of the reference grid. Because the behavior should be identical for different references, evaluation of different barycentric coordinates w.r.t. identical reference points was more of interest, than similar barycentric coordinates w.r.t. different reference points.

5.3.5.2 Results

The multi-view system was calibrated by estimating camera motion and camera intrinsics using a bundle adjustment procedure on the reference grid R. Lens distortion was modeled by two radial and two tangential distortion parameters, according to chapter 2.3. Residual error in terms of mean reprojection error (MRE) was roughly 0.06pixels, as given in Table 5.5.

Triangulation of the evaluation set $\mathbf{E}$ was done in four ways: linear triangulation using four views (*Lin4*), linear triangulation from views 1 and 2 (*Lin2*), least squares bundle adjustment using four views (*Bun4*), and least squares bundle adjustment using views 1 and 2 (*Bun2*). Results are shown in Figure 5.12 and Table 5.6.

The average distance to ground truth points for the four-view triangulation is in the range of 9μm, the difference between linear triangulation and bundle adjustment is

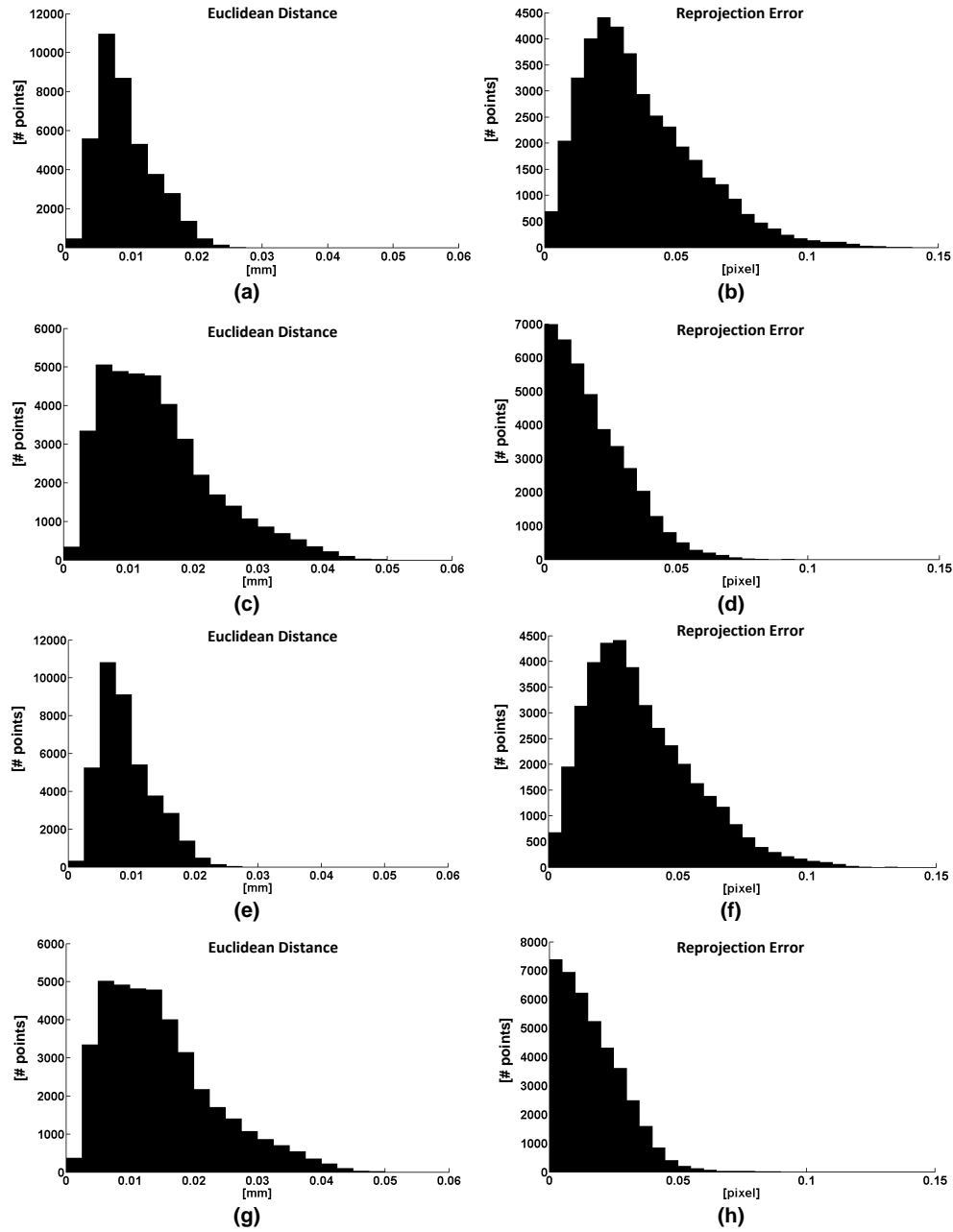Figure 5.12: Histograms of model based reconstruction errors. Results are given in Euclidean distance to ground truth point (DG) and reprojection error in view 1 (RE). Linear triangulation in 4 views (Lin4) DG (a), RE (b). Linear triangulation in 2 views (Lin2) DG (c), RE (d). Structure bundle adjustment in 4 views (Bun4) DG (e), RE (f). Structure bundle adjustment in 2 views (Bun2) DG (g), RE (h).

marginal. One important property is also visualized: a small reprojection error does not necessarily imply high reconstruction accuracy, especially when redundancy is missing, like in the two-view case.

Barycentric reconstruction was also performed in two ways. First, standard barycentric coordinates based on volume ratios in a tetrahedron were used in four views (BarT4) and two views (BarT2). Second, mean value coordinates were used on an enclosing cube in four views (Bar4) and two views (Bar2). Results are shown in 5.13 and Table 5.6.
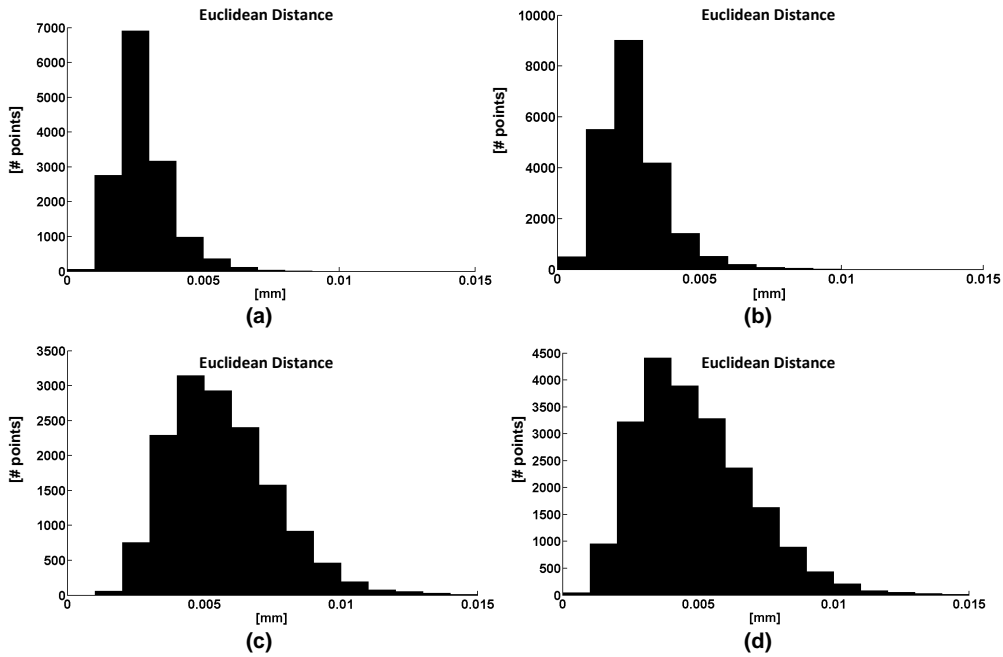
Figure 5.13: Histograms of barycentric reconstruction errors. Results are given in Euclidean distance to ground truth point (DG). Refinement in 4 views, tetrahedron (BarT4) (a). Refinement in 4 views, cube (Bar4) (b). Refinement in 2 views, tetrahedron (BarT2) (c). Refinement in 2 views, cube (Bar2) (d).

The residual error in terms of Euclidean distance to ground truth points is about 50% lower compared to model based reconstruction. While standard barycentric coordinates give roughly the same accuracy, these are considerably less robust. Roughly 30% of points had to be rejected during reconstruction, as these were moved to the tetrahedron boundary during optimization.

Reconstruction based on the L_∞ norm was evaluated, using an enclosing tetrahedron (4D weight vector) and four views (BarSOCP). Results are shown in 5.14 and Table 5.6. The global solution is comparable to the least-squares solution in terms of accuracy and robustness, and is also considerably better than the model-based approach.
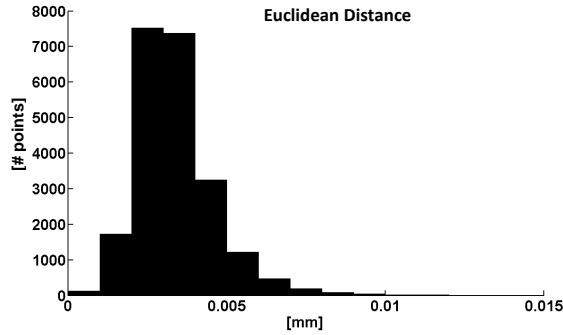
Figure 5.14: Histograms of barycentric reconstruction error based on the L_∞ norm. Results are given in Euclidean distance to ground truth point (DG).

Method	Mean error [μm]	# points	Mean RE [pixels]
Lin4	8.85	21592/21592	0.032
Lin2	11.24	21592/21592	0.02
Bun4	9.02	21592/21592	0.032
Bun2	11.22	21592/21592	0.018
BarT4	2.80	14467/21592	N/A
Bar4	4.16	21592/21592	N/A
BarT2	5.66	14950/21592	N/A
Bar2	4.86	21592/21592	N/A
BarSOCP	3.4	21472/21592	N/A

Table 5.6: Average reconstruction errors. Mean Euclidean error (DG), number of reconstructed points relative to total number of points, and `MRE` in view1.

5.3.6   Conclusion

Structure estimation based on barycentric coordinates requires considerably more hardware effort for calibration. A dense `3D` grid of reference points has to be created, and the corresponding image projections measured to high accuracy. Yet, the procedure has a number of advantages which should not be neglected. First, any type of projection and any type of distortion effect can be implicitly modeled. Even cameras which do not follow the central perspective model can be used without changing the method, a nice property, if projector devices are used. Second, reconstruction results are directly related to ground truth data. Systematic errors due to incomplete modeling of the camera system are eliminated. Accuracy is also directly related to ground truth accuracy. The denser the reference grid, and the more accurate the reference measurement, the more accurate the result will be.

During experiments, the usage of mean value coordinates proved to be useful, as these allow also extrapolation of data in the close vicinity of reference points.

The globally optimal solution based on `SOCP` showed similar results in terms of accuracy as the least-squares solution. Because the parameter vector linearly grows with the number of reference points, the problem was limited to an enclosing tetrahedron.

Chapter 6

Structured Light Metrology

Contents

6.1   Introduction

Among the measurement principles presented in Chapter 3, the coded-light measurement principle is industry-proven and one of the most versatile principles. It has been chosen in this work for the following reasons: active illumination makes it robust against object surface properties like color and texture. Area based reconstruction saves time during measurement, which reduces cycle times in industrial inspection and reduces the risk of false measurements due to sensor motion. Camera and projector optics and relative position are adjustable, which allows to change field of view and accuracy easily. A structured light system requires one or more cameras, which can be reused also for other inspection tasks, like 2D image processing, passive stereo or object recognition. With the advent of rugged micro projectors based on DLP or LCoS technology, structured light systems can be miniaturized such that they easily fit on an industrial manipulator, and are sufficiently robust to sustain the acceleration forces during motion.

    In this chapter, a review on the state-of-the-art in structured light metrology is given. While no direct contribution to the measurement principle itself is proposed, several en-

hancement methods of the input signal are discussed in Chapter 4, which allow to cope with surface specularity and high scene dynamics. Further, the volumetric triangulation method proposed in Chapter 5 may also be directly applied.

The chapter is structured as follows: a review of the state-of-the-art in structured light metrology is given in the next section, followed by a detailed introduction to the principle of time multiplexed structured light. Experiments validate the choice of structured light as a robust and accurate measurement technique. The chapter is closed by a short discussion.

6.2   Related Work

In structured light metrology, the correspondence problem is solved by actively signalling each point on the object surface with a unique code, usually by using a projector. A passive sensor, traditionally a camera, records the signal, and decodes it to obtain unique correspondence between projector and camera pixels. If camera and projector are intrinsically and extrinsically calibrated, it is sufficient to encode each column instead of each pixel, and compute the 3D reconstruction from epipolar geometry or directly from ray-plane intersection.

Over the past 30 years, a considerable amount of scientific work has been published in this field. According to the survey paper of Salvi [94], reconstruction methods are best categorized by the applied multiplexing technique.

6.2.1   Direct Coding

If each projector pixel column carries a unique code at every time instance, e.g. represented by color or intensity, one speaks of direct coding. Hereby, high temporal and spatial resolution may be obtained, at the cost of robustness, because the distance between codewords is small and easily distorted by surface texture or stray light. An early depth sensor based on intensity ratios has been proposed by Carrihill and Hummel [17], with poor results. Color patterns have been applied by Tajima and Iwakawa [100].

6.2.2   Spatial Neighborhood

The second approach relies on pattern coding in a spatial neighborhood. Hereby one loses spatial resolution, because for each measured point a local neighborhood is required to carry its codeword. High speed reconstructions are possible, as this is a one-shot technique. Maruyama et al. [75] proposed a non-formal coding technique using a pattern with vertical

slits of random length. The correspondence of one slit is decoded by the length ratios of of at least six neighboring slits. Mathematical coding techniques apply spatial codes like De Bruijn sequences and M-arrays.

6.2.3 Time Multiplexing

By far the most popular codification strategy is temporal multiplexing. Here, a unique code for each pixel column is generated by a temporal sequence of binary images. Spatial resolution is high, at the cost of temporal resolution. In the simplest case, the image sequence resembles a binary counter, which assigns a unique number to each column (Posdammer et al. [88]). A more robust codification technique through gray code patterns has been introduced later by Inokuchi et al. [49]. Using phase-shifting techniques, spatial accuracy can be boosted to sub-pixel level, at the cost of additional analogous patterns to be projected. As an example, see the work of Gühring [40].

6.2.4 Recent Developments

The work of Caspi et al. in 1998 [18] was one of the first who considered the radiometric camera-projector chain to get an improved structured light system. The main focus of this work was robust color coding in single shot patterns. In the paper of Je et al. [52], also pattern adaption is applied, but with the focus of optimizing the number of patterns for rapid range sensing of moving objects. In the recent work of Koninckx [64][63], scene adapted structured light is introduced. There, not only intensities, but also the shape of the pattern is optimized. Through a crude estimation of the geometry, the shape of the pattern is adapted to scene geometry to avoid aliasing caused by sampling foreshortened patterns. In [64] pattern adaption is used for single frame coded light techniques as well as for time based coding. Also the work of Grossberg et al. [39], where a projector is used to control an objects appearance in a camera, can be related to the problem of pattern adaption in structured light.

Novel ideas for pattern creation and analysis are mainly triggered by technological advances. In [82], the authors present a framework for fast active vision using DLP projectors. They take advantage of the temporal dithering of illumination, caused by the mirror flicker, which can be observed with high speed cameras. Each brightness value can be encoded uniquely due to dithering. When using this kind of technique, only one image is projected by the DLP projector, while an image sequence is acquired with a high speed camera.
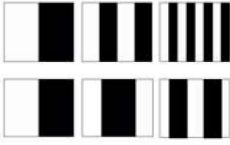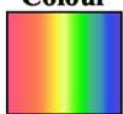
Approach	Method	Technique	Static	Moving	Binary	Grey levels	Colour	Periodical	Absolute
Time-multiplexing	Binary codes	Posdamer et al.	√		√				√
		Inokuchi et al.	√		√				√
		Minou et al.	√		√				√
		Trobina	√		√				√
		Valkenburg and McIvor	√		√				√
		Skocaj and Leonardis	√		√				√
		Rocchini et al.	√				√		√
	n-ary codes	Caspi et al.	√				√		√
		Horn and Kiryati	√			√			√
	Gray code + Phase shifting	Bergmann	√			√			√
		Sansoni et al.	√		√				√
		Wiora	√			√			√
		Gühring	√		√				√
	Hybrid methods	Kosuke Sato	√		√	√			√
		Hall-Holt and Rusinkiewicz		√	√				√
Spatial Neighborhood	Non-formal codification	Maruyama and Abe		√	√			√	
		Durdle et al.		√		√		√	
		Ito and Ishii		√		√			√
		Boyer and Kak		√			√		√
		Chen et al.		√			√		√
	De Bruijn sequences	Hügli and Maître		√			√		√
		Monks et al.		√			√		√
		Vuylsteke and Oosterlinck		√		√			√
		Salvi et al.		√			√		√
		Lavoie et al.		√			√		√
		Zhang et al.	√	√			√		√
	M-arrays	Morita et al.	√		√				√
		Petriu et al.		√	√				√
		Kiyasu et al.		√	√				√
		Spoelder et al.		√	√				√
		Griffin and Yee		√	√		√		√
		Davies and Nixon		√			√		√
		Morano et al.	√		√		√		√
Direct coding	Grey levels	Carrihill and Hummel	√			√			√
		Chazan and Kiryati	√			√			√
		Hung		√		√		√	
	Colour	Tajima and Iwakawa	√				√		√
		Smutny and Pajdla	√				√		√
		Geng		√			√		√
		Wust and Capson		√			√	√	
		Tatsuo Sato	√				√	√	

Scene applicability	*Static*	
	Moving	
Pixel depth	*Binary*	
	Grey levels	
	Colour	
Coding strategy	*Periodical*	
	Absolute	

Figure 6.1: Structured light codification strategies. The three main approaches are time multiplexed coding, spatial neighborhood coding, and direct coding [94].

Work on fast 3D scanning and automatic motion compensation is presented in [111]. A phase shift pattern sequence is rapidly projected and analyzed on the graphics processing unit (GPU) to obtain real-time performance, at 17 measurements per second. Due to the tracking capabilities of a fast system, the authors do not need to continuously project long code sequences. With the help of an additional color camera, textured depth maps are created. A method for dense shape reconstruction using a single structured light pattern is presented in [60]. The pattern consists of horizontal and vertical lines forming a grid. A number of coplanarity constraints are imposed on the grid connections and solved. Results have shown that reconstructions with a RMS error of half a millimeter can be obtained when using a coded structured light reconstruction as ground truth. In [12], a new method for the indexing problem of uncoded stripe patterns is presented using a maximum spanning tree of a graph to obtain possible connectivity and adjacencies of stripes. When using a 0.5megapixel camera, a reconstruction of the full image can be done within 0.17 seconds.

6.3 Methodology

In this section the reconstruction method based on time-multiplexed structured light is explained in detail. The coding strategy relies on a gray-code pattern. Accuracy is finally enhanced by employing a phase-shift technique for sub-pixel refinement.

6.3.1 Structured Light Using Gray Code Patterns

In time multiplexed structured light, each projector pixel column carries sufficient information to uniquely identify it in the camera image. The information is hereby contained in a temporal sequence of bits, which are mapped to intensity values, e.g. bright and dark (see Figure 6.3). The projector-camera setup is geometrically calibrated, so it is sufficient to decode the corresponding column for each camera pixel. Intersection of the epipolar line with the column finally gives a unique point correspondence (see Figure 6.4).

The temporal bit sequence resembles a gray-code binary counter, counting from 0 for the first column, to $n - 1$ for the final, n-th column. Equal to a standard binary counter, a total of n bits are required to encode 2^n columns, but it has the advantage that the Hamming distance is minimal for neighboring columns, which increases robustness. Wrong decoding of a bit will result in an error of only one resolution unit. Further, the width of the smallest pattern is always two resolution units, which facilitates detection and correct decoding, especially on slanted surfaces. An example of 16 gray code patterns
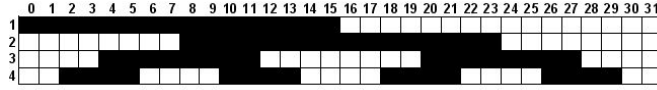
is shown in Figure 6.2.

Figure 6.2: Structured light code sequence. Binary encoding of 16 columns using a Gray-code sequence. Each column consists of a pair of pixels [62].
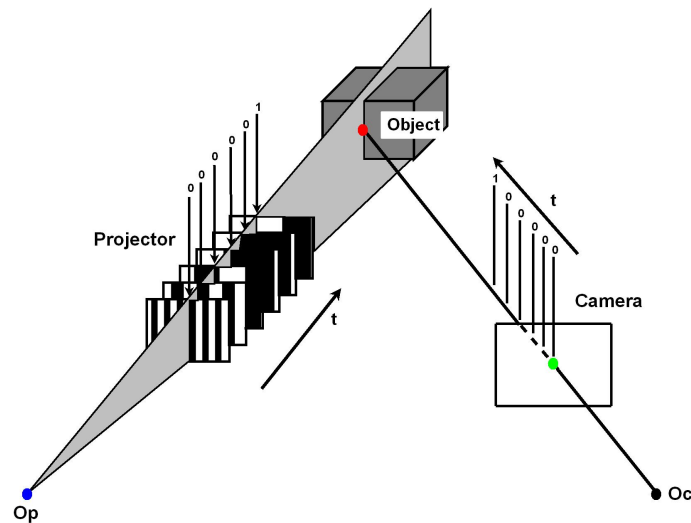

Figure 6.3: Binary pattern projection. Each column of the projector image is identified by a temporal sequence of binary stripe patterns to resolve the correspondence problem [62].

6.3.2 Phase Shifting

Discrete encoding of pixel positions limits the theoretically reachable measurement accuracy and introduces quantization noise. Projection of a continuous function helps to overcome this limitation and allows to determine a sub-pixel accurate correspondence in projector space for each camera pixel. In practice, several issues need to be considered when selecting a suitable continuous function. The function value for a single camera pixel should be determined from the intensity values at this pixel. No spatial neighborhood should be taken into account, as this would destroy the advantages of a time multiplexed method. Further, the function should be robust against stray light, slight defocusing of the projector, and properties of the object surface.

The phase shifting method can be considered as the gold standard in the field of optical,
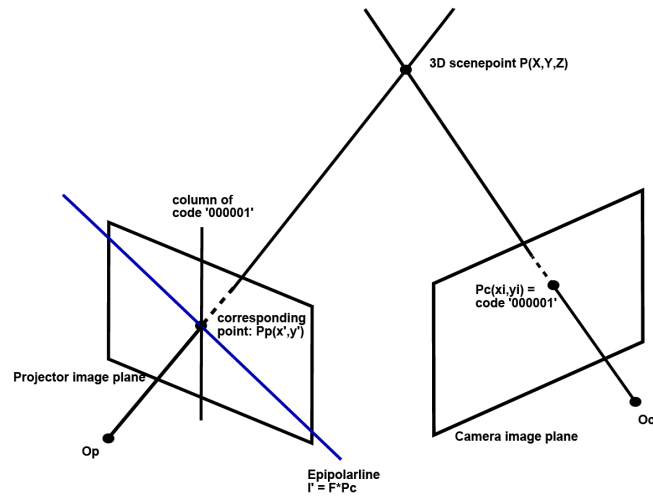
Figure 6.4: Point correspondence from column correspondence. Intersection of epipolar line and pixel column. Codification only establishes a correspondence of camera pixel to projector column. A unique point to point correspondence is established by intersecting the column with the respective epipolar line [62].

area based 3D reconstruction. A temporal sequence of a sinusoidal signal is projected, and sampled at each camera pixel. From the relative intensities of at least four samples, it is possible to reconstruct the signal phase. Assuming that projector defocus is equivalent to a low-pass filter, it will not distort the sinusoidal signal phase, and, because only intensity ratios are considered, the method is robust against surface properties like texture and color [94][11].

Consider four sinusoidal stripe patterns which are shifted in steps of $\pi/2$, as shown in Figure 6.5. In the simplest case, the pattern is generated by a binary sequence and projector defocus.
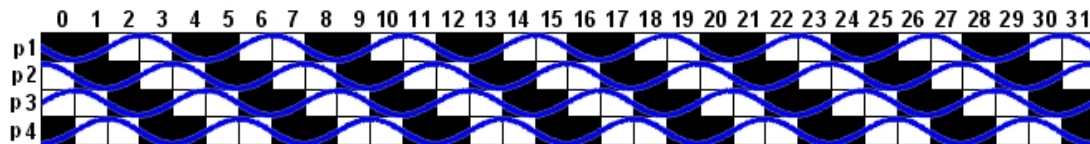

Figure 6.5: Phase shift sequence over 32 columns. A sinusoidal signal with a wavelength of four columns is desired. If the projector image is defocused, the signal may be approximated by a rectangular signal [62].

Using the following equation, the signal phase Φ_0 is obtained per pixel from the four

samples $p1 \ldots p4$:

$$\Phi_0 = arctan \left(\frac{p_2 - p_4}{p_1 - p_3} \right) \tag{6.1}$$

Figure 6.6: Reconstructed pixel phase. Pixel phase as calculated with (6.1), normalized to a range of $0 \ldots 1$ [62].

The resulting phase is depicted in Figure 6.6. In combination with the binary code (Figure 6.2), which determined the absolute phase as a multiple of 2π, a sub-pixel accurate column position can be calculated.

6.4   Experiments

Accuracy and robustness of a structured light technique is best evaluated by means of 3D reconstruction of known reference objects. To facilitate the generation of reliable ground truth information, planar reconstruction evaluation is used, where a planar surface is reconstructed as a point-cloud, and the deviation from the best-fit plane is calculated.

6.4.1   Theoretical Considerations

To provide a foundation for the practical evaluation, a few theoretical considerations on measurement accuracy in case of error-free geometrical calibration are presented. The workflow of a structured light algorithm starts with image acquisition, followed by a search for correspondences. In this course, a pixel in camera image space is selected and its correspondence in projector image space is sought. Location in camera image space is consequently considered noise-free, while the location in projector space is affected by radiometric effects like projection noise, image acquisition noise and external effects like changing stray light. Noise finally results in a location error. A schematic example is depicted in Figure 6.7. It is important to note that accuracy is ultimately determined by the recovered phase, not by the binary code, and phase accuracy is determined by the SNR. While the noise power depends on ambient light, object surface properties and

exposure time, the signal power is determined by the projector/camera radiometric range, radiometric resolution and the object surface. If one considers projector quantization noise as the principal source of error, the expected geometrical error is given by

$$\mathcal{E} = \frac{2 \times N}{S},\tag{6.2}$$

where N is the projector radiometric range and S the signal wavelength in pixels. A range of $S/2$pixels is subdivided into a maximum of N sub-pixel positions. As an example, one may set $S = 8$ and $N = 256$, and gets a maximal sub-pixel resolution of 0.0156pixels. With a typical geometric configuration, as sketched in Figure 6.7, the location error σ_p in the projector image plane is related to a geometric error of $\sigma_r = 9.7\mu m$ along the noise-free camera ray. Any other geometric error due to projector/camera noise can be derived in the same manner.
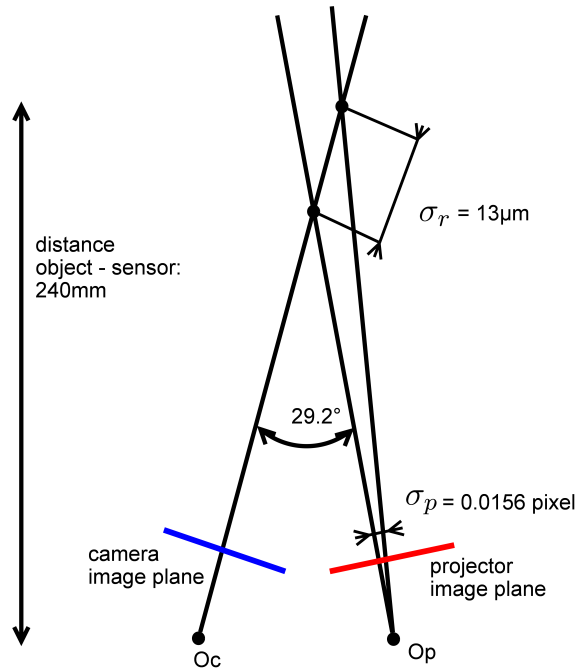
Figure 6.7: Relation of projector noise and location error. A location error in projector image space σ_p corresponds to a geometric reconstruction error σ_r in object space [62].

Apart from quantization noise, the recovered phase is subject to a variety of random distortions. According to experiments conducted in this work, phase information could be obtained with an average deviation of $\sigma_g = 0.5$ in units of grey values, leading to an error of $\sigma_p = 0.0078$pixels in the projector image plane and an error of $\sigma_r = 4.8\mu m$ in the

reconstruction.

Summing up the quantization error and the error due to noise, one gets

$$\sigma_r = \sqrt{\sigma_{r1}^2 + \sigma_{r2}^2} = 10.8 \mu m \tag{6.3}$$

In comparison, the measurements in Figure 6.10 show an error of $13.5\mu m$ at the angle of $90°$ (the plane was fronto-parallel to the projector image plane, which is equal to the measurement at $90°$ in Figure 6.10).

6.4.2   Planar Reconstruction

Reconstruction accuracy was primarily evaluated with plane-fitting experiments on a number of reference objects with different surface properties. The shape of the reference is shown in Figure 6.8, and the dimensional drawing is shown in Figure 6.9. For the plane-fitting experiments, the backside of the object is used. The standard deviation of reconstructed points from the best fitting plane in a least-squares sense has been calculated as a measure for planarity.

In a first step, accuracy under different orientations has been evaluated. An area of $10cm^2$ on the reference object was reconstructed from different orientations, by rotating the object around its axis in steps of $10°$ between $-80°$ and $+40°$. Results are shown in Figure 6.10.
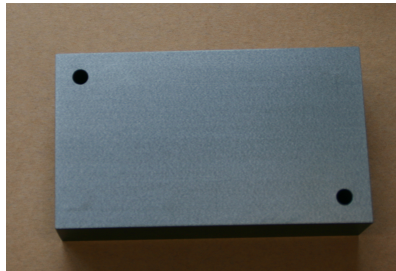
Figure 6.8: Anodized aluminum reference object. Dimensions are $30 \times 70 \times 120$ mm^3 [62].

As shown in Figure 6.10, an accuracy between $7\mu m$ and $35\mu m$ is reached within an angular range of $120°$ on the anodized aluminum workpiece. Reconstruction accuracy changes with orientation, which is due to perspective foreshortening and the resulting change of spatial resolution in object space. The problem is sketched in Figure 6.11. Projector stripes are narrow if the plane is nearly fronto-parallel to the projector. With

Figure 6.9: Dimensional drawing of the reference object [62].

Figure 6.10: Reconstruction accuracy under changing orientation. The best-fit plane was calculated from all reconstructed points, without outlier detection [62].

increasing slant angle, a single stripe has to cover more object surface area, and noise in phase reconstruction increases.

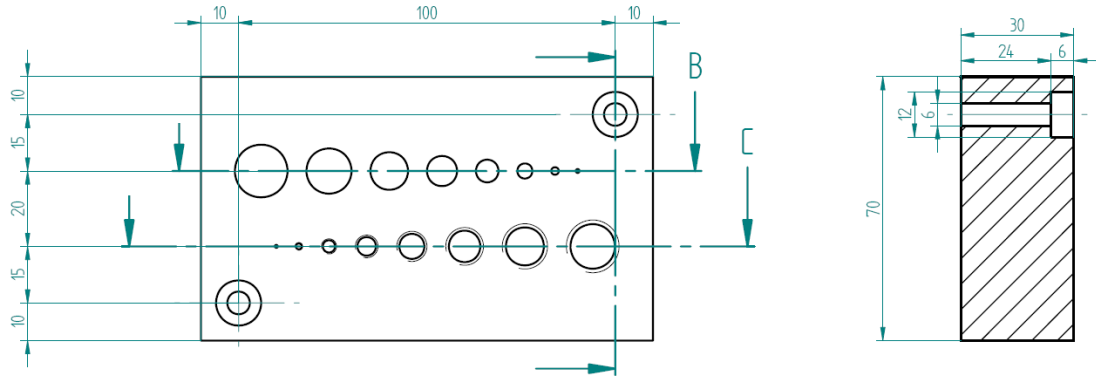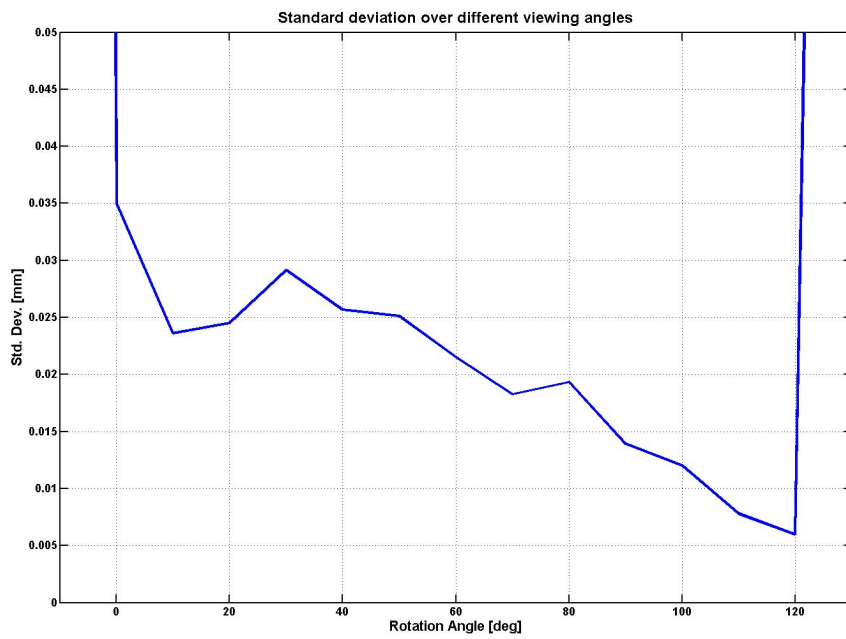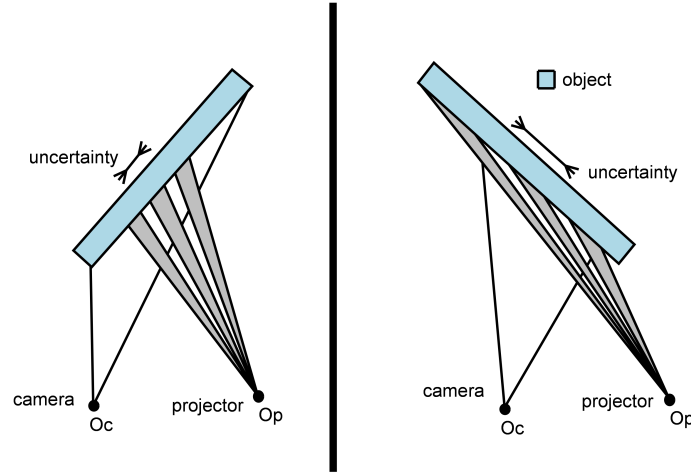Figure 6.11: Effect of perspective foreshortening on reconstruction accuracy. On a nearly fronto-parallel surface (left), spatial resolution of the projector stripes is higher than on slanted planes (right) [62].

Second, reconstruction accuracy on different surface types was evaluated. Six reference objects with identical shape but different surface finish, as shown in Table 6.1, have been reconstructed. The experimental setup is identical to the previous one. Each object is rotated around its axis in $10°$ intervals, and a planar area of $10cm^2$ is reconstructed in each step. The results, shown in Table 6.2, give the average error over all orientations, as well as the error of the best reconstruction between $-80°$ and $+40°$ of rotation angle.

6.4.3   Angle Measurement

The plane-fitting experiments presented in the previous section evaluate accuracy for a single plane at a given time. In a second experiment, several planes on a known reference object have been reconstructed simultaneously. The angular error between the planes serves as a quality measure.

The reference object shape resembles a prism with dimensions as shown in Figure 6.12. Three surface patches have been reconstructed. The same reference object has been constructed using three different materials, as shown in Figure 6.3. The ideal values for the angles are: $\alpha = 155°$ (angle between surface 1 and2), $\beta = 120°$ (angle between surface 2&3), $\gamma = 116.9°$(angle between surface 1 and 3). Due to an unknown production tolerance, the objects have been checked by a `CMM`. `CMM` measurement results are shown in
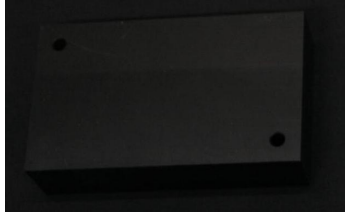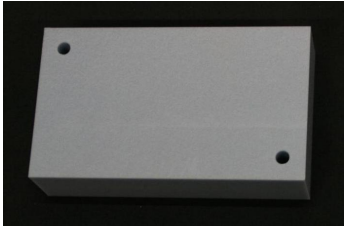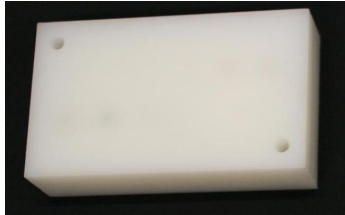
Workpiece 1		Workpiece 2	
anodized aluminum		chromed steel	
Workpiece 3		Workpiece 4	
black steel		black PVC	
Workpiece 5		Workpiece 6	
blue PVC		white PVC	

Table 6.1: All reference objects. Reference objects with different surface finish for evaluation of system robustness [62].

Workpiece	Average reconstruction results σ [mm]	Best reconstruction results σ [mm]
1	0.0135	0.0061
2	0.0724	0.0319
3	0.0235	0.0119
4	0.0359	0.0177
5	0.0467	0.0194
6	0.0962	0.0462

Table 6.2: Planar reconstruction under different surface properties. Using the reference objects from Table 6.1, every object was reconstructed under 12 different slant angles over a range of $120°$ in steps of $10°$. An area of $10cm^2$ was reconstructed. All reconstructed points (including outliers) were used for the computation of the standard deviation σ from the least-squares plane [62].

Table 6.4, and the structured light reconstruction results are shown in Table 6.5.
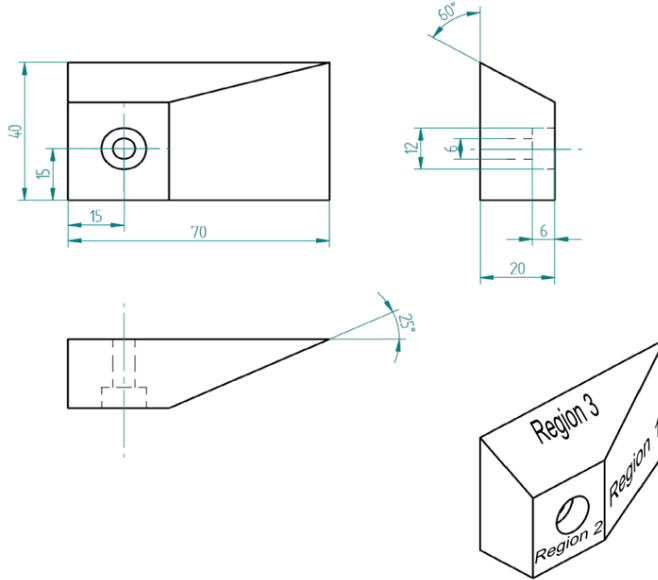
Figure 6.12: Prismatic reference object dimensions [62].

6.4.4   Effective Measurement Volume

The effective measurement volume of a structured light sensor is limited by the intersection of camera and projector fields of view, as well as depth of focus of both devices. In this experiment, the effective measurement volume for the medium scale sensor prototype, proposed in Chapter 3, is evaluated.

A planar surface was reconstructed over the possible reconstruction area at different depths. The hardware has been adjusted to get a field of view of approximately $120 \times 180 \text{mm}^2$ when projector and camera are in focus. From there, the plane has been moved along the sensor viewing direction, until the nearest and farthest position has been reached. The measurements are shown in Figures 6.13, 6.14 and 6.15. Between 170mm and 280mm distance to the sensor, the planar reconstruction gives reasonable results. Due to the central perspective nature of the sensor, spatial resolution is higher close to the sensor, and the field of view is higher at larger distances. In Figure 6.16, a sketch of the working volume is shown. Closest to the sensor, one has a reconstruction area of $150 cm^2$ and a point density of $220 points/\text{mm}^2$. Figure 6.13 shows that a reconstruction accuracy of 40μm can be obtained within the whole depth range as sketched in Figure 6.16.
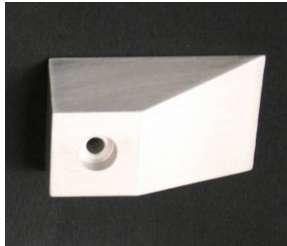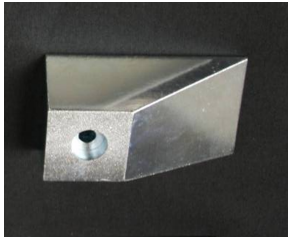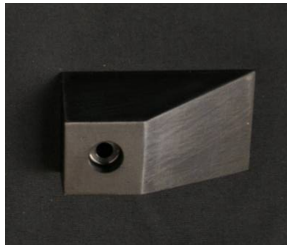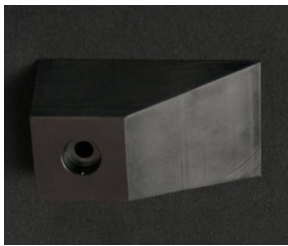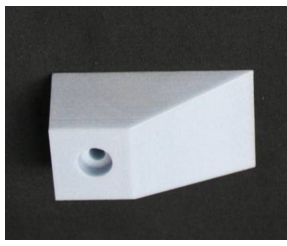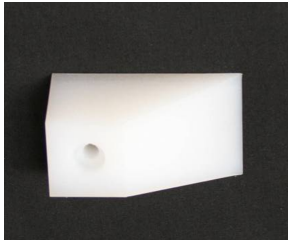
Prism 1		Prism 2	
brushed aluminum		chromed steel	
Prism 3		Prism 4	
black steel		black PVC	
Prism 5		Prism 6	
blue PVC		white PVC	

Table 6.3: Prismatic reference objects with identical shape but different surface properties [62].

	α [deg]	β [deg]	γ [deg]
Prism 1	155.094	119.939	116.866
Prism 2	155.106	120.335	116.943
Prism 3	155.042	120.421	117.293
Prism 4	155.231	120.257	116.526
Prism 5	154.985	119.914	116.952
Prism 6	155.135	119.915	116.837

Table 6.4: Ground truth measurement results of surface angles on the reference objects shown in Table 6.3. A CMM was used for this purpose. Nominal accuracy of the measurement system is 2.7μm over a range of 300mm.

	α [deg]	E_α [%]	β [deg]	E_β [%]	γ [deg]	E_γ [%]
Prism 1	154.942	0.098	120.516	0.481	117.318	0.386
Prism 2	155.134	0.018	120.407	0.059	116.560	0.327
Prism 3	155.055	0.008	120.664	0.208	117.297	0.003
Prism 4	154.976	0.164	119.808	0.373	116.723	0.169
Prism 5	155.102	0.075	120.412	0.415	117.165	0.182
Prism 6	154.821	0.202	119.298	0.514	116.971	0.114

Table 6.5: Angular measurement results on the prisms shown in Table 6.3. The angles between the three surface patches, labeled 1,2 and 3 in Figure 6.12, are computed. The error is computed as the relative deviation between this measurement and the measurements performed with the CMM, as shown in Table 6.4.
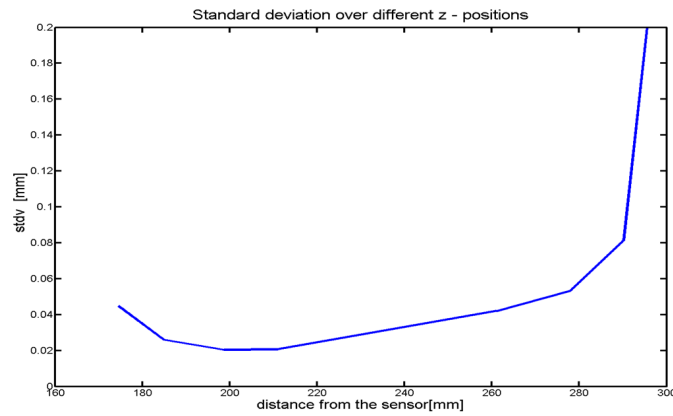
Figure 6.13: Planar reconstruction accuracy under varying depth. Due to limited depth of field of the projector and the camera, the working range is limited as well. A reasonable reconstruction accuracy is achievable within a depth range of about 100mm [62].
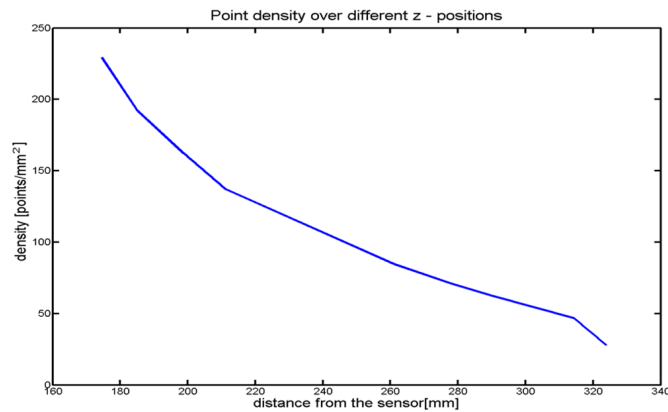
Figure 6.14: Planar reconstruction accuracy under varying depth [62].
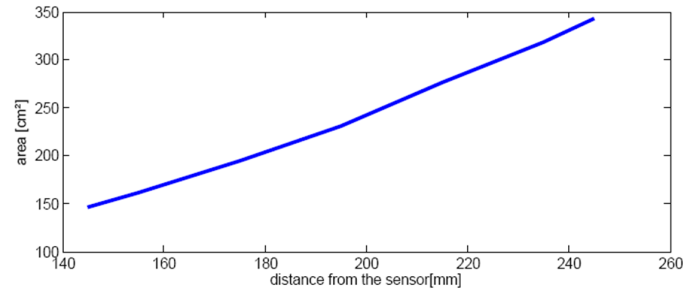
Figure 6.15: Maximal reconstructed area under varying depth [62].
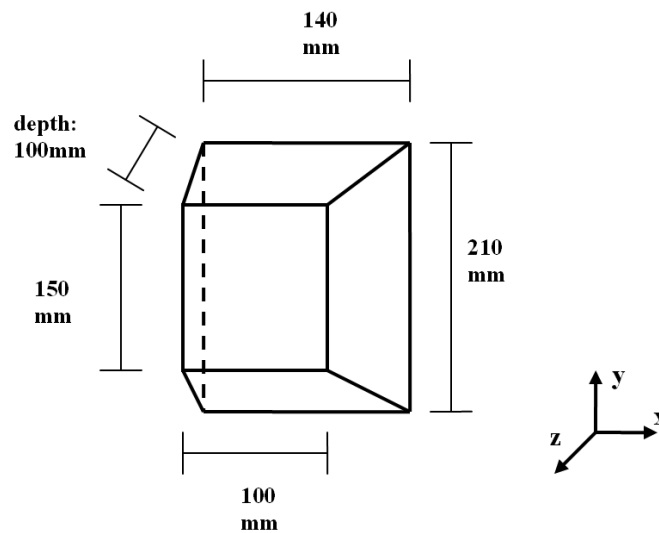
Figure 6.16: Sensor depth range. Depth range, limited by depth of focus and intersection of the fields of view, the sensor prototype is limited to a depth range of roughly 100mm. The maximal reconstruction area is $150cm^2$ at minimum depth and $300cm^2$ at maximum depth [62].

6.5 Conclusion

The structured light 3D reconstruction approach has been discussed in this chapter. The method, consisting of time-multiplexed binary coding, and sub-pixel refinement by phase-shifting, can be considered the gold standard in this field. The realization of the method on the measurement system proposed in Chapter 3 has been evaluated in terms of accuracy, robustness and measurement range. Experiments showed that under good environmental conditions, the method reaches an accuracy level close to the theoretical limit.

One issue not addressed in this chapter is geometry. The way from known correspondence to the 3D point includes intersection of the respective projector column with the

epipolar line and subsequent triangulation. Many different methods exist in literature, including ray-plane intersection and pure table lookup. With the method proposed in Chapter 5, effects like nonlinear lens distortion may also be handled, which is not possible if the projector column is modeled as a `3D` plane.

Robustness of the codification method suffers from external effects like stray light, specular reflections on shiny objects, interreflections, and high scene dynamics. Some of these problems are addressed in Chapter 4, where solutions are presented based on `HDR` imaging and pattern adaption.

Chapter 7

Kinematic Geometry

Contents

The measurement concept introduced in Chapter 3 incorporates an industrial manipulator in the measurement chain. Robotic manipulators, typically articulated arms with six DOF, show high repeatability in 3D position and orientation (pose), but may be inaccurate in absolute positioning. As stated in [56] and [8], errors in the kinematic parameters are among the main causes of positioning inaccuracy. If a manipulator is hereby used as a measurement device, its kinematic parameters should be accurately determined, which leads to a parameter estimation problem, commonly addressed as kinematic calibration or kinematic identification.

Part of this estimation problem is to determine the robot tool-hand transformation, which is also called eye-hand transformation for robot mounted cameras. The unknown Euclidean transformation from the robot hand CF to the tool CF depends on sensor geometry and mounting accuracy. So, if it has to be known accurately, it should be determined after mounting.

In this chapter, a short review of existing approaches to kinematic calibration are given, as well as a novel method for calibration of a robotic manipulator. The robot kinematic

chain and its tool are thereby observed by a hand mounted camera through a mirror. This enables hand-eye, hand-tool, and kinematic robot calibration without incorporating accurate external references, except a mirror. Using this particularly simple setup, hand-eye calibration becomes independent of the kinematic chain and parameter observability constraints in kinematic calibration become more relaxed, which makes pose planning for robot calibration more convenient.

7.1   Introduction

In this work, two novel contributions to the problem of robot calibration are presented. Both are based on the same principle: Observing the mechanical robot structure by a hand mounted vision system through a mirror (see Fig. 7.1).
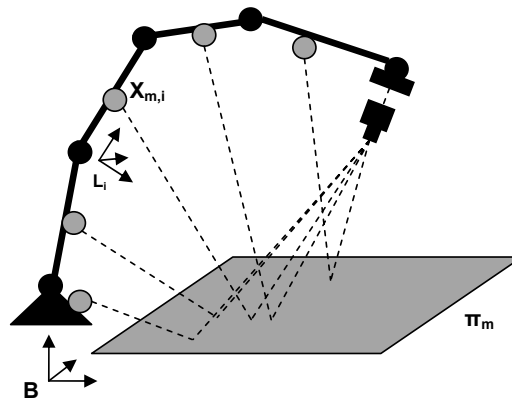
Figure 7.1: Mirror based kinematic calibration. A hand mounted camera observes reference markers, mounted on the kinematic chain, through a mirror.

The advantages of this configuration are as follows: for hand-eye calibration, the robot kinematic chain is eliminated from the estimation procedure and hence does not contribute to the final error. Additionally, besides determining the hand-eye transformation, any hand mounted tool, that is visible in the mirror, may be calibrated for relative pose as well. Only one mirror is required as an external device, it does not even need to be fixed during calibration.

In kinematic calibration, the motion of each robot link is observed by attaching reference markers to it and tracking their motion. This results in a considerably better conditioned estimation problem, compared to the state-of-the-art. The mirror can be arbitrarily placed to provide a good overview of the robot hand and body. In contrast to

hand-eye calibration, it must remain fixed during calibration.

The chapter is structured as follows: in Section 7.2, related work in the field of robot calibration is discussed. An automated method for mirror-based hand-eye calibration is presented in Section 7.3. Kinematic calibration is discussed in Section 7.4. In Section 7.5, experimental results on synthetic data are presented, and practical feasibility of the proposed method is validated. The chapter is concluded by a discussion in Section 7.6.

7.2   Related Work

7.2.1   Hand-eye Calibration

Most hand-eye calibration methods share one common principle: estimation of hand-eye pose, based on kinematic and optical measurements. On one side, hand motion $\mathbf{A}_i$ is calculated using the manipulator joint readings. On the other side, camera motion $\mathbf{B}_i$ is estimated relative to a reference target, using a pose estimation algorithm. The principle is depicted in Fig. 7.2. The problem may either be formulated in relative motions, as $\mathbf{A}_i\mathbf{X} = \mathbf{X}\mathbf{B}_i$, where $\mathbf{X}$ is the unknown hand-eye transformation, or in terms of absolute poses, incorporating also the position of the reference target $\mathbf{Z}$ as $\mathbf{A}_i\mathbf{X} = \mathbf{Z}\mathbf{B}_i$.

Shiu and Ahmad [96] derived a system of linear equations from which the angles of rotation, and subsequently the rotation matrices are determined. Tsai and Lenz [103] derived a closed-form solution from two relative pose measurements based on constraints on the eigenvectors of $\mathbf{R}_X$, $\mathbf{R}_A$, $\mathbf{R}_B$. Chou and Kamel [20] used unit quaternions to represent rotations and solved for $\mathbf{q}_X$ using algebraic, linear constraints and singular value decomposition (SVD). Wang [107] presented a survey of these methods and favored Tsai's method in terms of accuracy.

One argument against separately solving for $\mathbf{R}_X$ and $\mathbf{t}_X$ is error propagation. An uncertain estimate of the rotation negatively influences translation accuracy. To overcome this limitation, nonlinear methods to solve for rotation and translation simultaneously were introduced by Horaud and Dornaika [48], who also introduced an additional formulation of the calibration problem in terms of projection matrices, and Zhuang and Shiu [120], who put more emphasis on parameter observability through measurements. Daniilidis [22] proposed to model the rigid transformation as a dual quaternion, which represents the screw of motion in an elegant way, leading to a linear, simultaneous solution for $\mathbf{R}_X$ and $\mathbf{t}_X$.

More recently, Strobl and Hirzinger [97] addressed the problem of formulating a physi-
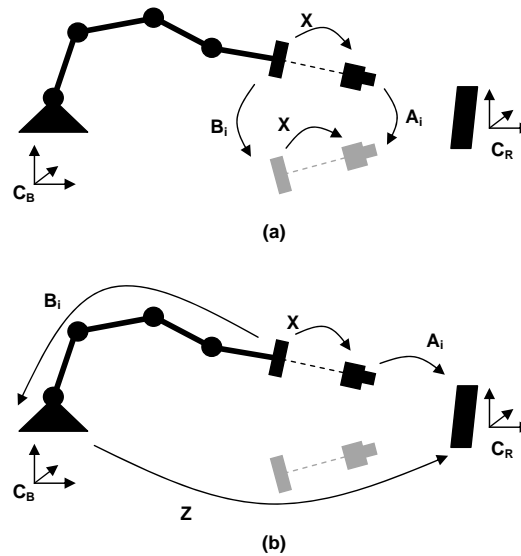
Figure 7.2: Hand-eye parameter estimation. The Euclidean transformation $\mathbf{X}$ is estimated by closing a loop of relative pose measurements (a), or absolute pose measurements (b).

cally relevant error model for nonlinear optimization in $SE3$, including automatic weighting of rotation and translation error during optimization.

7.2.2   Kinematic Calibration

In essence, kinematic calibration leads to a parameter estimation procedure, where the robot end-effector pose is measured by an external sensor. A parametric model, describing the forward kinematic mapping from joint coordinates $\mathbf{q}$ to end-effector poses $\mathbf{H}_{h \to b}$ is optimized until the deviation of calculated robot poses from the measured poses becomes minimal. A considerable amount of research effort has been put into kinematic calibration, addressing the following three key problems:

- Pose measurement

- Kinematic modeling

- Observability

Hollerbach et al. [47] presented a unified taxonomy of kinematic calibration methods. Their main categorization is *open-loop methods*, which is the principle described in the previous paragraph, *closed-loop methods*, where the end-effector is rigidly attached to the

ground (a varying number of DOF may be fixed), and *screw-axis measurement methods*, where joint axes are explicitly measured and kinematic parameters are derived analytically.

Open-loop methods require an external reference system, which measures one or more DOF of the end-effector pose. The measurement system may be a laser tracker [89] [4], a laser interferometer [67] [2], hand mounted laser beams and stationary cameras [69], vision systems [44] [91] [118] [119] [5] [33] [72] [79] and structured light systems [58]. Veitschegger [105] used a precision machined mechanical reference and aligned the end-effector manually, whereas in [78] a CMM was used for pose measurement. Goswami et al. [38] used an linear variable displacement transducer (LVDT) sensor to measure one DOF of the end-effector pose.

Closed-loop methods do not require external measurement devices. The end-effector is rigidly connected to the base, creating a closed kinematic chain. Because the range of possible arm motions is severely restricted, careful planning of the fixture type is necessary, such that all kinematic parameters can be identified [5] [81].

Among the vision-based calibration methods, open-loop methods are the most popular, because the camera pose relative to a known reference target can be measured through structure and motion techniques [42]. Either the camera, or the target is mounted to the end-effector and provides a measurement of all six pose DOF [118]. Thereby, work on camera calibration can be fused with kinematic calibration to formulate a large parameter estimation problem, where camera and robot are calibrated simultaneously [119]. Even self-calibration of the camera and kinematic system may be considered [77].

7.3   Hand-Eye Calibration

The calibration setup is as follows. A hand mounted camera observes the robot flange through a mirror. Hand pose and mirror poses relative to the camera are estimated directly.

The hand mounted camera observes the flange through a planar mirror, as sketched in Fig. 7.3. By moving the end-effector relative to the camera, different views of the flange are generated, which allow to formulate redundant constraints between CCF $\mathbf{C}_c$ and hand CF $\mathbf{C}_h$.

A very similar problem is described in Kumar et al. [66], where a set of stationary, diverging cameras is calibrated by using a stationary reference target and a moving planar mirror. In contrast to Kumar et al., in this work the mirror plane normal is estimated directly.
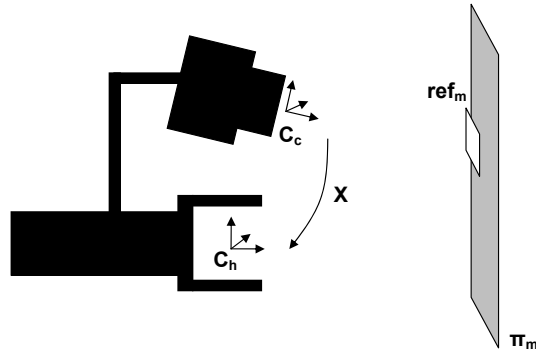
Figure 7.3: A hand mounted camera observes the hand through a mirror. For hand-eye calibration, the hand is moved, changing the position of π_m in the hand `CF`. A rectangular reference marker is placed on the mirror to determine the plane normal.

7.3.1   Theory

In the proposed method, the camera projection center $\mathbf{C}$ and rotation matrix $\mathbf{R}$ relative to the hand are estimated from measured mirror plane normals $\mathbf{n}_i$ and measured reflected camera poses $\mathbf{C}'_i$, $\mathbf{R}'_i$ in at least three views.

The Mirror plane π_m induces a reflection on the camera. The image of a real world point $\mathbf{X}$ in the mirrored camera $\mathbf{C}'$ is identical to the image of the mirrored point in the real camera, as shown in Fig. 7.4.
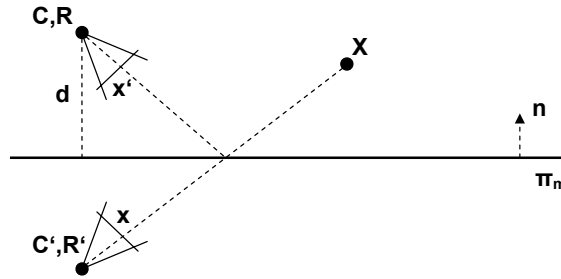
Figure 7.4: Mirrored camera pose. A camera (projection center $\mathbf{C}$, rotation $\mathbf{R}$) observes world point $\mathbf{X}$ through a mirror. The image of reflected point $\mathbf{X}'$ in the real camera is identical to the image of the real point in the reflected camera ($\mathbf{C}'$, $\mathbf{R}'$).

If the real camera pose is known, the mirrored camera is uniquely determined by the position of the mirror plane $\pi_m = (\mathbf{n}^T, -d)^T$, which has three `DOF`. Note that d specifies the distance to $\mathbf{C}$, not to the origin. The mirrored projection center $\mathbf{C}'$ must have the same distance to π_m as $\mathbf{C}$, and the line connecting both centers must be perpendicular to π_m. It follows that

$$\mathbf{C}' = \mathbf{C} - 2d\mathbf{n}. \tag{7.1}$$

For the mirrored rotation, the following proposition holds:

Proposition 1 *The vector connecting* $\mathbf{C}$ *and* $\mathbf{C}'$ *is orthogonal to* $(\mathbf{r}_k + \mathbf{r}'_k)$, *iff* $\mathbf{r}_k$ *is not orthogonal to the mirror plane.*

A proof is given in [66]. Vectors $\mathbf{r}_k$ and $\mathbf{r}'_k$, $k = 1 \ldots 3$ denote the three columns of rotation matrices $\mathbf{R}$ and $\mathbf{R}'$. The following nonlinear constraint in $\mathbf{C}$ and $\mathbf{r}_k$ holds:

$$(\mathbf{C}'^T - \mathbf{C}^T)(\mathbf{r}'_k + \mathbf{r}_k) = 0. \tag{7.2}$$

The constraint can be linearized by introducing additional coefficients of the form

$$\mathbf{C}^T \mathbf{r}_k = s_k. \tag{7.3}$$

Kumar used (7.2) and (7.3) to formulate a linear, homogeneous equation system in $\mathbf{C}$ and $\mathbf{r}_k$. The linear solution was projected to the closest orthogonal solution. It was further refined by a bundle adjustment procedure, enforcing orthogonality and unit norm of $\mathbf{r}_k$. His approach requires at least five images to find a solution.

However, if the mirror plane normal is known, (7.2) can be simplified to

$$(\mathbf{C}^T - 2d\mathbf{n}^T - \mathbf{C}^T)(\mathbf{r}'_k + \mathbf{r}_k) = 0, \tag{7.4}$$

which is equal to

$$\mathbf{n}^T(\mathbf{r}'_k + \mathbf{r}_k) = 0. \tag{7.5}$$

Each image contributes one linear constraint for each column of $\mathbf{R}$, so at least three images are necessary to solve for rotation. The center of projection is recovered from (7.1), by solving the system of equations

$$(\mathbf{C}' - \mathbf{C}) \times \mathbf{n} = \mathbf{0}. \tag{7.6}$$

As a consequence it was shown that knowing the mirror plane normal simplifies the estimation problem and allows calibration from fewer views.

7.3.2   Method

From a methodological point of view, the calibration procedure consists of the following steps: (1) Estimate $\mathbf{n}$, (2) use 7.6 to solve for $\mathbf{C}$, and (3) use a nonlinear bundle adjustment to increase accuracy. In this section, these steps are reviewed more closely.

It is easy to get an estimate for $\mathbf{n}$ from a single image, by using orthogonal vanishing points from a planar, right-angled object attached to the mirror. For a vanishing point $\mathbf{V}$ in $\mathbb{P}^3$, the image projection is given by

$$\mathbf{v} = \mathbf{M}\mathbf{V}_{1\ldots3}, \tag{7.7}$$

where $\mathbf{M}$ is the first 3×3 submatrix of projection matrix $\mathbf{P}$. Now, if $\mathbf{V_A}$ and $\mathbf{V_B}$ are two vanishing points orthogonal to the plane normal $\mathbf{V}_n$, the following equation holds for the first three coefficients of $\mathbf{V}_n$:

$$\mathbf{V}_{n,1\ldots3} = \mathbf{M}^{-1}\mathbf{v}_A \times \mathbf{M}^{-1}\mathbf{v}_B. \tag{7.8}$$

As an example, $\mathbf{V}_n$ may be determined from the mirror edges, or from any planar, rectangular object which lies flat on the mirror. If the reference marker dimensions are known, it would theoretically be possible to estimate the mirror plane directly. This step is omitted for practical reasons, as the reference marker is likely to be blurred in the image, resulting in a potentially erroneous depth estimate, and the mirror plane may not be accessible due to covering layers, or in case of a second surface mirror.

With known mirror plane normals, and known mirrored pose estimates, the linear equation systems described in 7.3.1 are solved using Singular Value Decomposition.

A nonlinear optimization procedure is applied to refine results. The problem is to estimate the set of unknown parameters, namely the mirror plane in each view and the hand-eye transformation. The mirror plane normals have been estimated from orthogonal vanishing points. The position of these may be uncertain, if the measured vanishing lines are close to parallel. Therefore, it is more natural to optimize the vanishing point positions, and derive the mirror plane normal from these. The hand pose has been measured from image features, it is also natural to minimize an image based error:

$$\arg\min \sum d(\mathbf{x}_i, \mathcal{P}(\mathbf{X}_i))^2 + \sum (\mathbf{l}_i^T \mathbf{v}_i)^2 + \sum (\|v_i\| - 1)^2, \tag{7.9}$$

where $\mathbf{x}_i$ are image measurements of the robot hand, and $\mathbf{X}_i$ are the corresponding 3D features. Function $\mathcal{P}(\mathbf{X})$ is the image projection function, which includes reflection by the

mirror, transformation to the CCF and central perspective projection. Also, reconstruction of the mirror plane from vanishing points is included in this function. The second summation term in (7.9) penalizes deviation of the vanishing points $\mathbf{v}_i$ from the measured vanishing lines $\mathbf{l}_i$, and the third term pulls the norm of the vanishing point vectors towards unit norm to constrain scale.

7.4   Kinematic Calibration

The constraints derived for mirror-based hand-eye calibration are not valid for kinematic calibration, as the eye-base transformation is not constant. Nonetheless, a mirror provides the opportunity to observe eye motion relative to the base, as well as link motions relative to camera and base.

A setup for kinematic calibration is sketched in Fig. 7.1. Each link, which is sufficiently large and observable by the camera, has at least one artificial point marker attached to it. Two markers on the robot base have a known distance, to uniquely constrain scale. The number of mirrors is not restricted, so one is free to add one for observing certain poses. The number of reference markers is also not restricted. If a link is mechanically too small to attach a marker, it may be omitted.

The resulting optimization problem has more free parameters than the original open-loop problem, but these are better constrained. Each point marker contributes three unknown DOF, namely the translation relative to its link CF.

7.4.1   Theory

The arm is modeled according to Robinson's notation [93], where each link is described by a rigid transformation relative to the previous link frame $\mathbf{H}_i$, consisting of a series of particular transformations of the form

$$\mathbf{H}_i = \mathrm{R}(p_1, z)\mathrm{T}(p_2, z)\mathrm{T}(p_3, x)\mathrm{R}(p_4, x)\mathrm{R}(p_5, y)\mathrm{R}(p_6, z), \tag{7.10}$$

where $\mathrm{R}(\varphi, a)$ denotes a rotation of angle φ around axis a, and $\mathrm{T}(d, a)$ denotes a translation of d along axis a.

Depending on the link structure, two or more parameters per link are held fixed during calibration. A hand mounted camera observes the robot arm through a mirror, described by plane π_m, which is given in the base coordinate system $\mathbf{C}_B$. Each link $\mathbf{L}_i$, which is at least partially observable through the mirror, has reference markers $\mathbf{X}_{m,i}$ rigidly attached

to it, where m specifies the point index, and i specifies the CF it is given in. For a single point measurement, the following equation holds:

$$\mathbf{x}_{m,q} = \mathbf{K}\,[\mathbf{I}|\mathbf{0}]\,\mathbf{H}_{e\to h}^{-1}\mathbf{H}_{h\to b}^{-1}\mathbf{D}_\pi\mathbf{H}_{i\to b}\mathbf{X}_{m,i}, \tag{7.11}$$

where $\mathbf{x}_{m,q}$ is the image point corresponding to $\mathbf{X}_{m,i}$ at robot pose q, and $\mathbf{D}_\pi$ is the reflection matrix of the mirror plane π_m.

7.4.2 Method

To estimate the set of unknown parameters, a LM procedure is applied. Assuming that intrinsic camera calibration and hand-eye calibration are known, a good initial solution for the camera pose is found by computing the forward kinematic solution from blueprint parameters. The mirror plane is initialized from a single square marker attached to the surface.

7.5 Experiments

7.5.1 Simulation

Both algorithms are evaluated in terms of sensitivity to image noise and initialization error. The simulated robotic model resembles a Mitsubishi RV-1A 6 DOF manipulator. As for the camera, a 2megapixels sensor with a pixel width of 4.4μm and 4.8mm focal length was simulated.
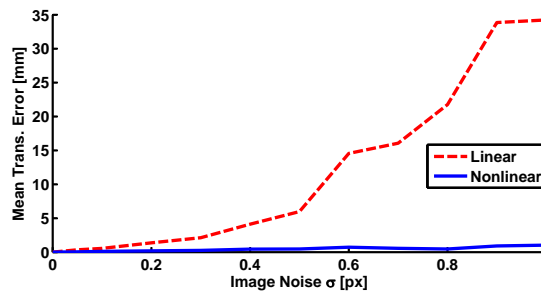
Figure 7.5: Mean translational error over image noise for the nonlinear hand-eye calibration result (solid line) and the linear solution (dashed line).

As for hand-eye calibration, several small square markers are distributed over the mirror surface. Each square marker is represented by 40 points, evenly distributed along its perimeter, which simulates an edge-based segmentation method. The robot hand is

also represented by a square marker, with hand `CF` in the center. The image measurements
are corrupted with Gaussian noise, where σ is increased from 0pixels to 1.0pixels. For each
noise level, ten repetitions were made. Results are given in Figure 7.5, in terms of mean
translational error over noise. It can be seen that the linear method gives plausible results
(below 10mm) for image noise below 0.5pixels. Up to 1pixels image noise, the nonlinear
method always converges to an accurate solution (< 1mm error).

To evaluate kinematic calibration, a total of eight reference points have been applied
to the links and the robot base. The marker positions were selected to approximate the
actual marker positions, as shown in Fig. 7.7. Kinematic calibration results over image
noise are given in Fig. 7.6. To evaluate calibration accuracy, a total of 4096 joint values
have been regularly sampled over the workspace. At each position, the translational error
between ideal and calibrated model has been evaluated. Again, for each noise level, the
mean error over ten repetitions is given.
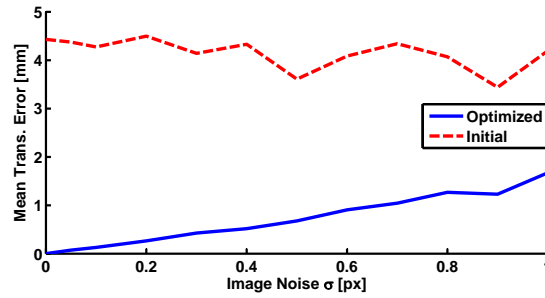
Figure 7.6: Mean translational error over image noise for the optimized kinematic calibra-
tion result (solid line) and the initial solution (dashed line).

The synthetic calibration experiment gave a residual position error of 0.12mm for hand-
eye calibration, and 0.13mm for kinematic calibration, under an image based localization
error of 0.1pixels. The calibration poses were selected, such that the points of interest
were visible in the camera, but not to maximize an observability measure.

7.5.2 Practical Experiment

A practical experiment has been conducted by mounting an industrial camera on the
end-effector of a Mitsubishi RV-1A robot arm (Fig. 7.7). The camera was calibrated
in an offline procedure using a method similar to Zhang [116]. A mirror (dimension
800×600mm^2) was placed in front of the manipulator. Rectangular markers (dimension
25×25mm^2) were placed on the front side of the mirror plate to estimate the plane normal.
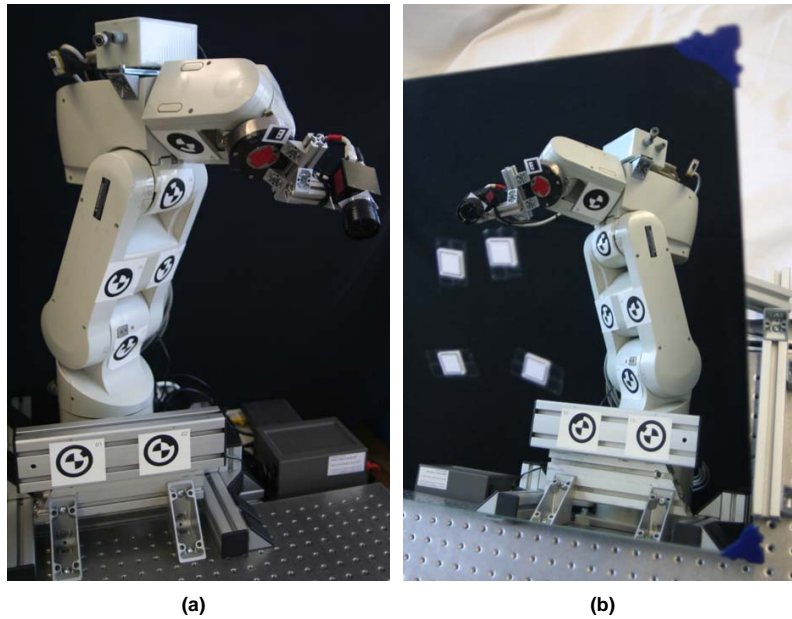
Figure 7.7: Robot calibration setup. A camera, mounted on a robot with attached reference markers (a) observes the robot through a planar mirror (b).

The applied mirror was a second surface mirror, which might introduce a translational error in the mirror plane due to refraction, a source of error which has not been further investigated.

For hand-eye calibration, a total of 16 robot poses were used. The robot flange pose was measured by fitting two ellipses in the image. One for detecting the flange perimeter and one to measure a drill-hole to constrain rotation around the hand center (see Fig. 7.8).

Kinematic calibration was performed from a total of 35 robot poses, the corresponding virtual camera views are shown in Fig. 7.9. The complete kinematic chain was evaluated against 32 independent camera poses calculated relative to a highly accurate reference target ($150 \times 150\text{mm}^2$ dot grid pattern, nominal accuracy $0.9\mu m$). These poses were compared with the blueprint robot on one side, and with the calibrated robot on the other side. Hand-eye transformation for the blueprint robot was estimated using the method of Tsai/Lenz. The blueprint robot gave a mean translational error in the camera pose was 3.88mm for the blueprint robot, and 1.54mm for the calibrated robot.
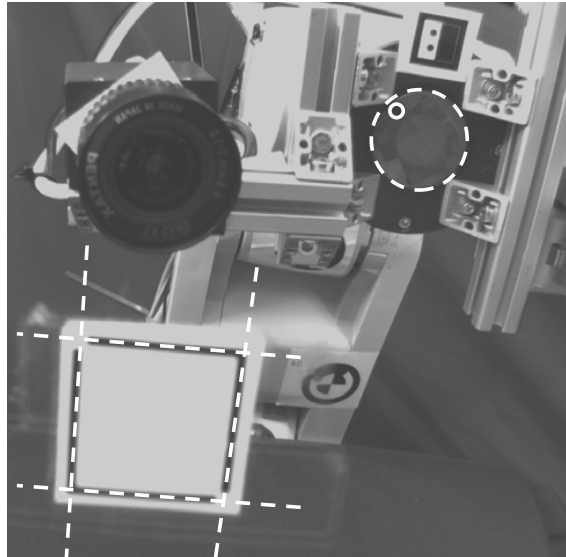
Figure 7.8: Hand-eye calibration. A hand mounted camera observes the hand flange (dashed ellipse) in a mirror. Rectangular markers (dashed lines) are attached to the mirror to determine the plane normal.
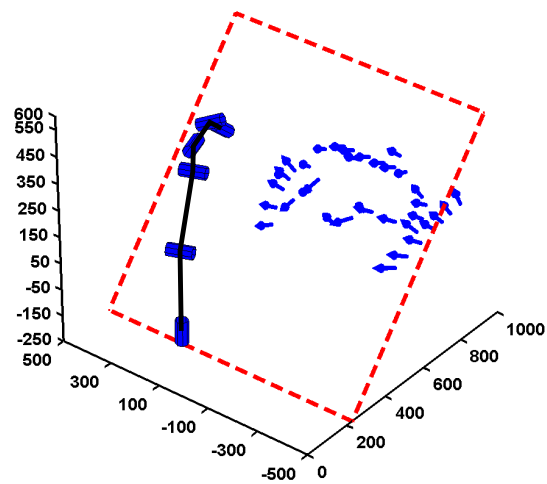
Figure 7.9: Robot model and mirrored camera poses. Moving the hand mounted camera, a number of virtual camera stations behind the mirror (dashed line) is generated, observing the robot. Axis units are given in [mm].

7.6   Conclusion

Robot kinematic- and hand-eye calibration is possible using a hand mounted camera and mirrors. In the context of this work, several cameras are anyway mounted to the robot hand for measurement purposes. The method is cheap, because no external reference apart from a planar mirror is required. One obvious drawback is the requirement that the hand flange must be visible. In many cases this is difficult to guarantee due to mechanical reasons. It is important to note that the hand-eye translational offset must be sufficiently large to generate a well constrained photogrammetric problem.

The feasibility of the proposed procedure was demonstrated in a laboratory experiment.

While no initialization is necessary for the hand-eye problem, the kinematic calibration procedure requires rough knowledge of the kinematic chain, hand-eye offset and location of the mirror plane. Kinematics and hand-eye offset may be initialized from construction plans, while the mirror plane location can be estimated e.g. from the mirror edges. In principle, the method can be applied to any type of kinematic manipulator, as long as links are visible by the camera through a mirror. It would also be possible to place the camera outside and observe the moving robot, but here one would lose flexibility in the range of motion and in selecting suitable calibration poses.

Further research would include calibration through multiple mirror reflections, as it has been done for 3D reconstruction in [29], and the tracking of robot links by natural features instead of coded markers.

Chapter 8

Applications and System Evaluation

Contents

In this chapter, the methodological findings of this work are brought together in an integrated system. A 3D computer vision sensor is constructed, consisting of four cameras and two projectors, mounted on an industrial manipulator, as described in Chapter 3. The prototype is used to solve a practical inspection problem, according to a functional specification as provided by an automation company. The author hereby aims to demonstrate flexibility and practical applicability of the entire measurement concept, and also of the specific methods which have been introduced so far.

8.1 Introduction

In this chapter we turn back to a similar problem as stated at the beginning (Section 1.1). An exhaust pipe, shown in Figure 8.1(a), which is approximately $50cm$ long, has one contact flange on each ending. It should be automatically checked if the relative position of both flanges is correct, whether the sealing surfaces are planar and the drill holes are compliant with defined quality constraints on position and diameter.

(a)

(b)

(c)

(d)
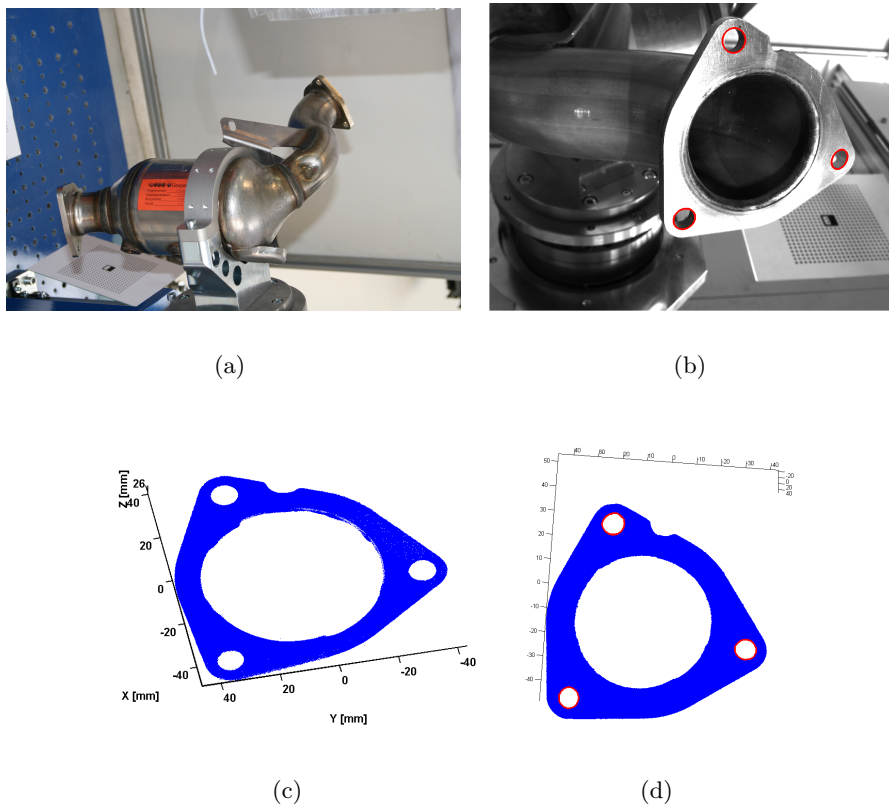
Figure 8.1: Exemplary measurement application. An exhaust pipe (a) has to be inspected for dimensional correctness of entrance flange (left) and exit flange (right). A measurement image of one flange with marked boreholes is shown in (b). The point-cloud resulting from a structured light reconstruction is shown in (c). Measured borehole circles on the point-cloud is shown in (d).

The measurement sequence hereby consists of moving the sensor to each contact flange, perform the local measurements, namely drill-hole measurement and planarity checks, and finally compute the relative displacement of both flanges.

8.2 Inspection Methodology

8.2.1 Drill Hole Inspection

Each contact flange consists of several drill holes (Figure 8.1(b)). Hole diameters and distances between them need to be inspected. In a first preprocessing step, the influence of stray light is suppressed by using difference images as described in Chapter 4.4.2. Because of the shiny metal surface, data redundancy is required to accurately measure all boreholes,

which is achieved by employing all cameras of the sensor setup. The drill hole borders
are coplanar with the flange plane, which facilitates data fusion from the different camera
views, as illustrated in Figure 8.2.1. In a first step, the flange plane π is reconstructed
using structured light reconstruction (Figure 8.1(c)) and subsequent least-squares plane
fitting.

(a)   (b)
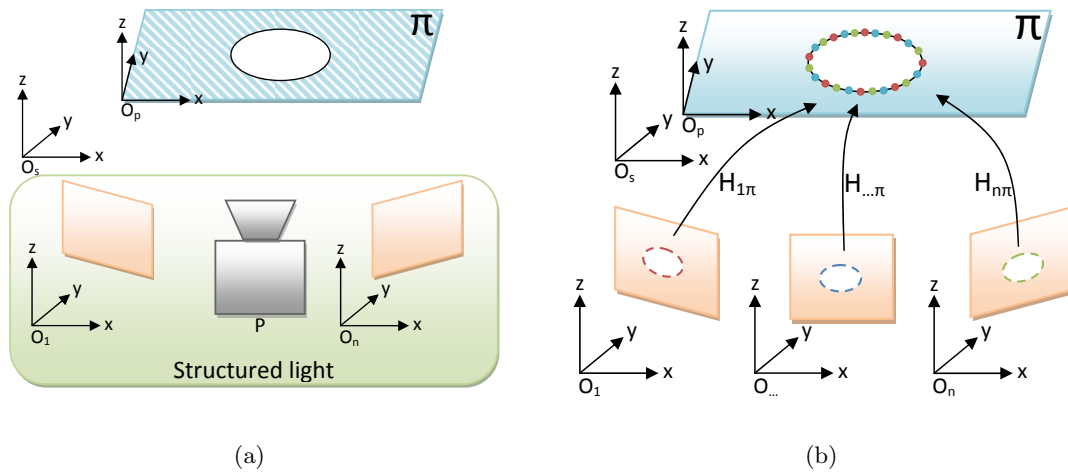
Figure 8.2: Principle of borehole inspection. First, the surface plane π is reconstructed
using structured light (a). Second, an edge detector is applied in each camera view $1 \ldots n$.
Edge pixels are back-projected to the plane π, where an ellipse-fitting method fuses the
results [70].

In a second step, images of the flange are acquired from all cameras. In each camera
view, the drill hole borders are detected separately by using an edge detector. The edge
pixels are then back-projected onto the known flange plane π, where an ellipse-fitting
algorithm is applied to estimate the borehole parameters (radius and center), as shown
in Figure 8.1(d). In this course, the information gathered by several cameras is elegantly
merged. The ellipse-fitting procedure, essentially a least-squares estimation problem, is
conducted in object space.

A different option would be to estimate the borehole ellipse separately in each camera
view, and finally estimate the corresponding 3D quadric. The proposed method has the
advantage that ellipse estimation takes place in object space. For nearly circular boreholes
this means that the estimation problem is not biased due to perspective distortion (see
Chapter 2.5), and measurements from all camera views, weighted with their confidence,
are elegantly merged.

8.2.2   Planarity Inspection

As a by-product of the drill-hole inspection problem, the flange plane is reconstructed, in form of a point cloud, as well as an analytic plane. The deviation of reconstructed points from the plane holds information about flange planarity. For better visualization, the height map of the sealing surface of the exhaust pipe depicted in Figure 1.1 is shown in Figure 8.3. A point-cloud of the second exhaust pipe is shown in Figure 8.1(c).
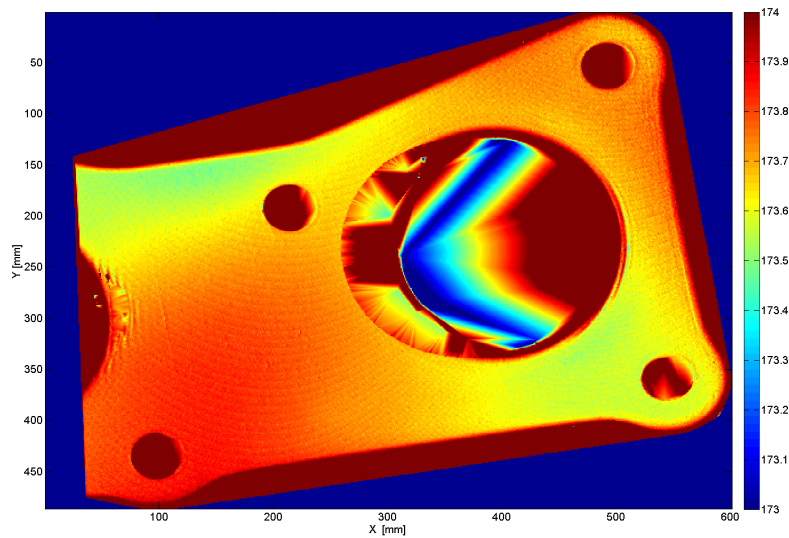
Figure 8.3: Height map of a flange surface. Using pattern adaption, even specular reflecting surfaces can be reconstructed with high spatial resolution [70].

8.2.3   Flange Inspection

The relative position of both flanges is derived from the local drill-hole measurements, the known eye-hand calibration and robot kinematic parameters. The working principle is sketched in Figure 8.4. A unique CF is assigned to each flange, using the drill hole centers as anchor points. The relative pose of the flange to the sensor is determined and further transformed to the robot base CF. When both flanges are given in the common base CF, their relative pose is easily derived.
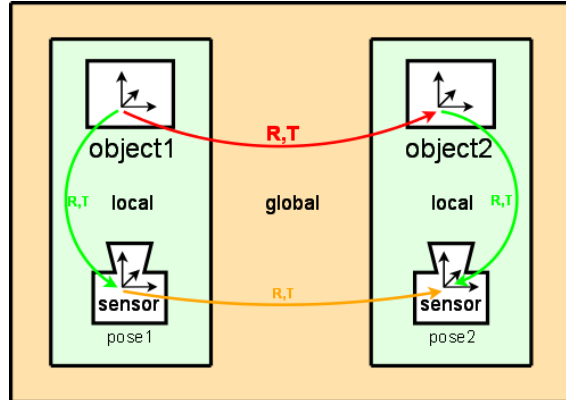
Figure 8.4: Principle of a global measurement. Two distinct local measurements are performed in the sensor `CF`. In each measurement, the pose of an object relative to the sensor is determined. With known sensor motion, both measurements are transformed in a common `CF` [70].

8.3 Evaluation

8.3.1 Local Measurement

The measurement results were evaluated against reference measurements from a `CMM`. The entry and exit flange planes were measured using tactile point measurements and plane-fitting. The interior walls of the drill holes were hereby sampled pointwise (10 measurements per hole). The resulting points were projected to the flange planes to obtain the upper hole boundary. An ellipse-fitting algorithm gave the hole centers and radii. The reference measurement was repeated twice. Because the result varied slightly, the vision system was evaluated against both `CMM` results, as shown in Tables 8.1 and 8.2, where the reference measurements are compared against each other, as well as the vision system results.

To evaluate the local measurement performance under more well-behaved ambient conditions, a synthetic flange was measured as well, consisting of a sheet of paper glued on glass, with boreholes (three black circles) printed on it. The same flange measurement procedure as before was performed. Figure 8.5 shows an example image. The green circles denote the reference centers of each hole and the red dots denote the measurement result.

Results of the synthetic flange measurement are summarized in Table 8.3. Circle radii, distances between circle centers and distances to the centroid are compared to the reference information, obtained indirectly through the printer resolution.

The synthetic flange measurement showed superior accuracy, compared to the real-

	E_1 [mm]	E_2 [mm]	E_{CMM} [mm]
Center hole1	0.28	0.3	0.036
Center hole2	0.24	0.25	0.034
Center hole3	0.18	0.2	0.035
Mean center	0.24	0.25	0.035
Flange centroid	0.0	0.0	0.0
Radius hole1	-	0.23	-
Radius hole2	-	0.34	-
Radius hole3	-	0.20	-

Table 8.1: Measurement results on entrance flange. On the entrance flange, shown in Figure 8.1(b), the borehole radii and center locations were measured. The Euclidean errors between measured parameters and the first CMM measurement (E_1), as well as the second CMM measurement (E_2) are given, as well as the Euclidean deviation between both CMM measurements (E_{CMM}) [70].

	E_1 [mm]	E_2 [mm]	E_{CMM} [mm]
Center hole1	2.39	2.03	0.91
Center hole2	3.8	3.22	1.58
Center hole3	2.93	2.39	1.07
Mean center	3.04	2.55	1.19
Flange centroid	3.01	2.49	1.13
Radius hole1	-	0.85	-
Radius hole2	-	0.19	-
Radius hole3	-	0.60	-

Table 8.2: Measurement results on entrance flange. On the exit flange, the borehole radii and center locations were measured. The Euclidean errors between measured parameters and the first CMM measurement (E_1), as well as the second CMM measurement (E_2) are given, as well as the Euclidean deviation between both CMM measurements (E_{CMM}) [70].

world example. On one side, this results from the idealized object properties, and on the other side from more accurate reference information. The real flange holes are slightly elongated, making an exact reference measurement difficult. Even two CMM measurements showed a considerable deviation from each other.
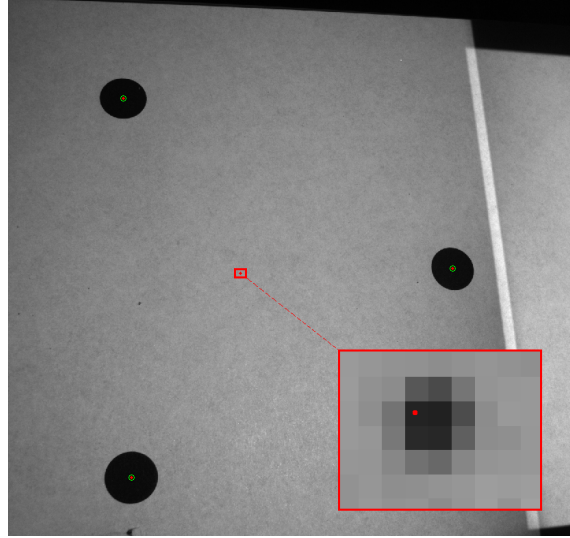
Figure 8.5: Synthetic flange with hand-annotated reference information (green circles) and measured `3D` points, reprojected to the image (red dots). The centroid of the circle centers is also shown, in comparison with the ideal centroid, which has been printed on the synthetic flange a s black dot [70].

	E_{rad} [mm]	E_d [mm]	E_{cc} [mm]
Hole1	0.0314	0.0526	0.0133
Hole2	0.0161	0.2242	0.1017
Hole3	0.0100	0.0251	0.0859
Mean	0.0192	0.1006	0.0670

Table 8.3: Metric error of synthetic flange measurement. Circles on a printed paper (see Figure 8.5) were inspected. The measured parameters are circle radii, distances between circle centers and distances of the circle centers to their centroid. The measured results are compared to ground truth information retrieved from the printer resolution. E_{rad} is the absolute radius error, E_d is the absolute error of center-center distance (w.r.t. the next hole), and E_{cc} is the error of the distance between a circle center to the centroid of all circle centers [70].

8.3.2   Global Measurement

The accuracy of global measurements is mainly affected by the robot positioning quality, which can be split into repeatability and absolute positioning accuracy. In this section, repeatability and absolute accuracy are evaluated for a specific manipulator, which has been used during measurement, namely an ABB IRB 2400/16 6 `DOF` robot.

8.3.2.1 Repeatability

Repeatability forms a lower bound on the reachable absolute positioning accuracy. As a consequence, in a first step it has been evaluated whether the robot complies to the nominal value of 60μm [1].

In accordance with the ISO9283 standard [50], the industrial robot manipulator of the type ABB IRB 2400/16, with the vision sensor mounted on its hand, was moved to random position in its workspace and afterwards placed in front of a static reference target, similar to the calibration procedure described in Chapter 5.2.3. The sensor pose relative to the target was measured and recorded. The procedure was repeated over roughly 70 minutes with different robot motions. Finally the pose difference over all measurements was evaluated, resulting in a mean absolute error of 28μm.

8.3.2.2 Absolute Positioning

To evaluate absolute positioning accuracy, a steel plate with eight drill holes, as shown in Figure 8.6, was constructed. The reference object was inspected by a `CMM` to get accurate ground truth information.
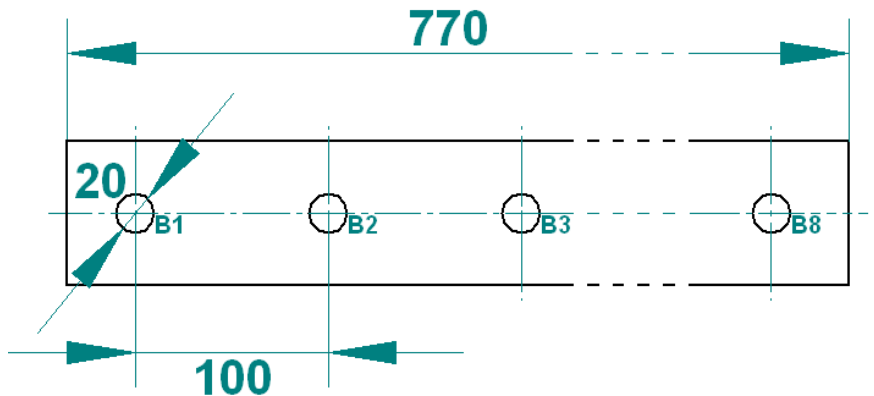
Figure 8.6: Reference object for evaluation of absolute positioning accuracy. Eight drill holes with a diameter of 20mm (named *B1, B2,..., B8*) are placed in intervals of 100mm over the length of the reference object.

The plate was placed in the robot workspace, as shown in Figure 8.7. The measurement system was used to determine the relative position of the drill holes. Because the length of the object exceeds the vision sensor's workspace, it had to be moved from one hole to

the next. The measurement results were merged afterward using the known robot motion and sensor-hand pose. The borehole measurement procedure was hereby identical to the one described in Section 8.3.1. The evaluated distances are sketched in Figure 8.7(b).
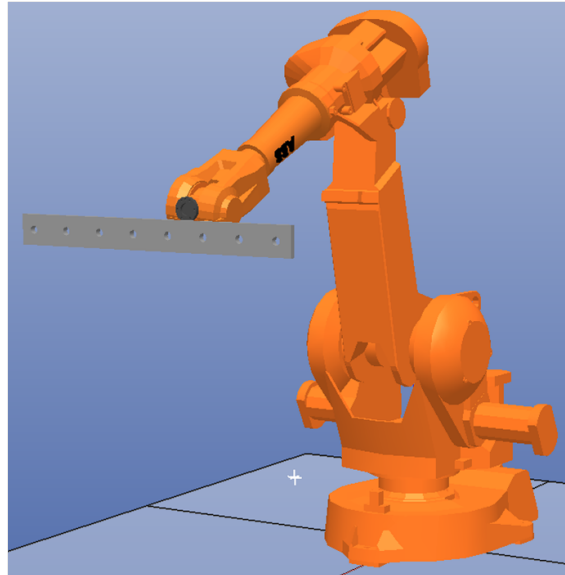
(a) The reference object is mounted 1400mm in front of the robot base, at a height of 850mm.

(b) Measured drill holes. The distances between the third hole and rest of the holes to the right were measured with the CMM and with the vision sensor.

Figure 8.7: Robot accuracy evaluation using a reference object.

A comparison to the CMM data gave a mean absolute positioning accuracy of 0.065mm over ten repetitions of the measurement procedure.

The outer borders of the reference plate were also reconstructed using structured light, to obtain a length measurement for the object (see Figure 8.8).

The measured plate length was 769.74mm, in comparison to the CMM measurement of 769.97mm. The difference of 0.24mm serves as a quality measure for the absolute positioning accuracy over the whole plate length.
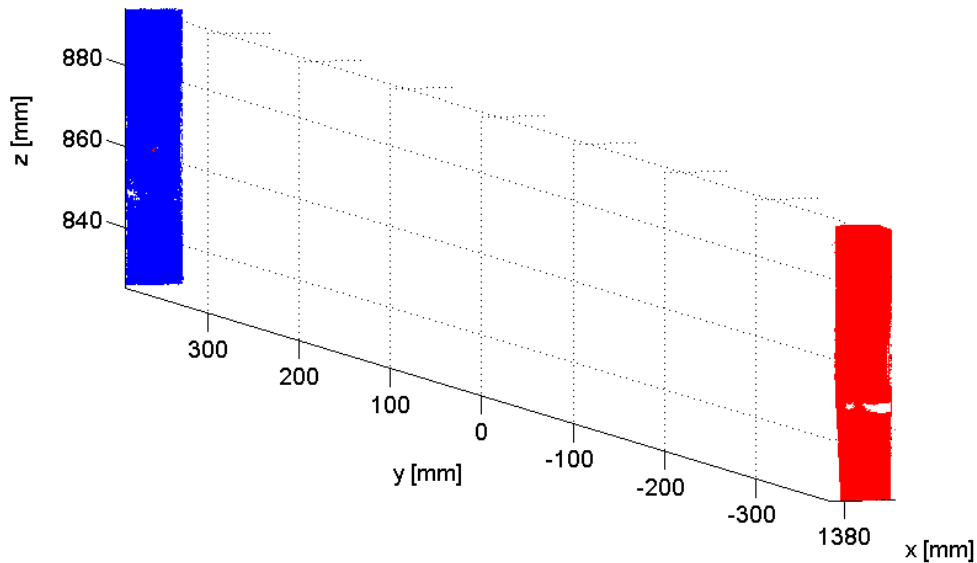
Figure 8.8: Measurement of object length. Both outer surfaces (read and blue) of the reference object were reconstructed using structured light and plane-fitting. The object length was derived by calculating the perpendicular distance of the center of the blue plane to the red plane [70].

Using the same procedure, plate thickness was measured, as shown in Figure 8.9. The thickness measurement was compared to reference data measured by a digital sliding caliper of the type *Mitutoyo CD-6"CX*, with a nominal tolerance of ± 0.02mm. A total of five reference measurements were conducted, in the same area as the vision sensor, resulting in an average thickness of 14.74mm. The difference to the vision based measurement is 0.33mm.

Finally, the flange measurement as described in Section 8.2 was conducted on a real exhaust pipe and compared to reference data from a `CMM`. Because two reference measurements with the `CMM` also differed considerably, a comparison between them and the results from the vision system is given in Table 8.4.

8.4   Conclusion

This work has been partly conducted in cooperation with an automation company, in the course of generating a highly flexible, accurate and easy to use measurement instrument for `3D` inspection. In this chapter, the applicability of the proposed methods to a real-world problem with real-world requirements has been demonstrated. Additional problems
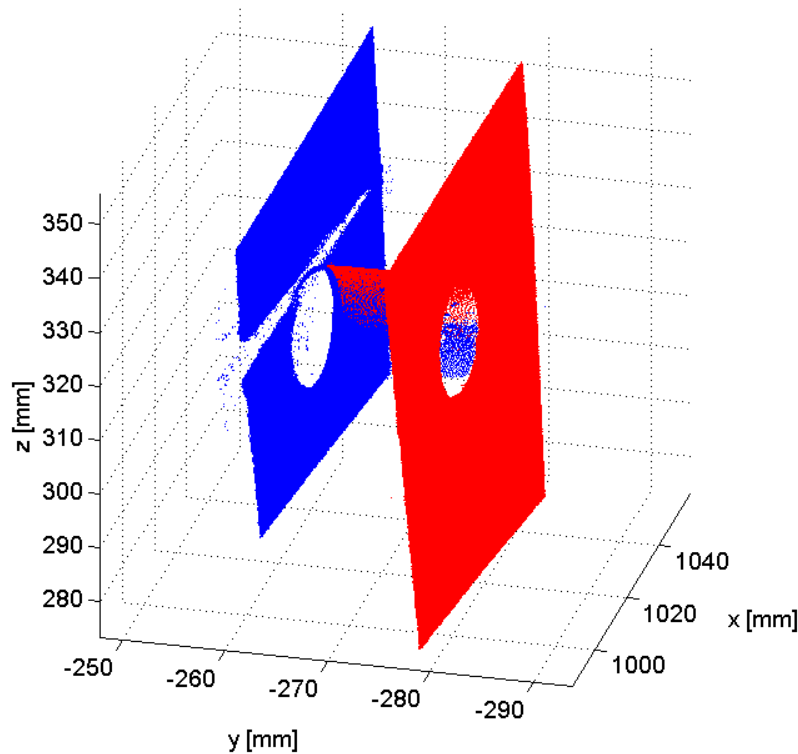
Figure 8.9: Measurement of object length. The thickness was estimated by reconstructing the front and the back side of the steel bar using structured light. To both point clouds (red and black dots) planes were fitted, and plate thickness was calculates as the perpendicular distance from the red plane centroid to the blue plane [70].

the sensor was able to solve include 2D image processing tasks like barcode reading, optical character recognition and detection of welding artifacts.

The exhaust pipe, used as an example throughout this work, has a complex shape. The difficulties we had during creation of reliable reference data using a CMM, show the complexity of the measurement problem. The global measurement consequently gave the best results on a simpler object, namely the ruler used to evaluate absolute positioning accuracy. It can also be observed, that the ruler thickness measurements show a larger error than the pose measurements of the drill holes. This is probably due to joint inaccuracies. The robot joint excitations for the upper joints were considerably higher in the thickness experiment, because the sensor had to be placed on opposite position and orientation. Errors in the eye-hand calibration are also more prominent in this configuration. An erroneous offset in the sensor viewing direction is essentially doubled when doing thickness measurements and cancels out when doing only translational measurements.

Comparison of global flange measurement to CMM data			
Entrance flange			
	E_1	E_2	E_{CMM}
Angle error [deg]	0.0	0.0	0.0
Ellipse center error [mm]	0.0	0.0	0.0
Exit flange			
	E_1	E_2	E_{CMM}
Angle error [deg]	0.63	0.76	0.14
Ellipse center error [mm]	3.01	2.55	1.07
Flange centroid measurement (entrance - exit flange)			
	CMM_1	CMM_2	Sensor
Distance [mm]	406.67	407.25	406.75

Table 8.4: Measurements of the sensor system, compared to two measurements of a CMM. Two reference measurements differed considerably, so the vision sensor is compared to both. The entrance flange was used as alignment plane and the centroid of the flange's drill holes was used as origin, therefore there is no error on the entrance flange. The centroid error states the translational error between the coordinate systems fitted to each flange.

In accordance with the literature in robot calibration, accuracy strongly depends on the manipulator used, so it is hard to give absolute measures for the reachable system accuracy. The striking feature of the proposed concept however is ease of use. Once set up, all calibration tasks, ranging from camera/projector calibration to eye-hand calibration and kinematic calibration, run automatically. This alone saves time during robot maintenance, sensor maintenance or replacement of broken parts. Further, pattern adaption on the measurement area also runs automatically. The image-based measurement is planned afterwards, on an optimal sensor placement and on illumination-optimized images. Selection of cameras, selection of the light source, data transfer and control are all solved in one generic package which can be applied to a variety of inspection problems.

Chapter 9

Conclusion

In this thesis, the problem of vision based three-dimensional quality inspection in industrial environments has been addressed, with a special focus on medium scale part inspection (dimension in the range of one meter). Within the huge variety of possible measurement concepts, a robot mounted projector-camera system has been proposed, motivated and evaluated. In the course of the work, the two main limiting factors of accuracy and robustness were further investigated: radiometric quality and geometric accuracy.

An investigation of the available measurement concepts showed, that their realizations are either limited in the range of applications, overly complicated or very expensive. The sensor proposed here mainly consists of industrial cameras, standard video projectors and a manipulator. In relation, e.g. to a CMM, all of these parts are comparably cheap. The parts are combined in one package, in a fixed geometric configuration, and are together able to solve a large variety of measurement problems. The applicability has been demonstrated on a real-world problem, where an exhaust pipe had to be inspected for a variety of different parameters.

The central factor of every robust vision system is image quality. On metallic objects, which are common in industrial quality inspection, dynamic range is a key issue. An investigation of different methods for dynamic range enhancement was conducted. Both the principle of HDR imaging and the principle of intensity optimization on the projector brought improvements in accuracy on difficult surfaces. Hereby the optimization of projector intensity is preferable, because excessive light intensity is hereby avoided in the first place, which also avoids imaging artifacts. Especially in quality inspection, where the same type of object is inspected repeatedly, the optimization pattern can be reused from object to object, which results in a considerable speedup, compared to HDR imaging. Currently,

the proposed pattern adaption method is based on knowledge of scene geometry, which leads to an ill-posed problem. Future work would be the investigation of methods which do not rely on scene knowledge or even do not rely on geometric sensor calibration.

With radiometrically optimized images of an object, the second problem is how to relate image based measurements to world coordinates. We hereby rely on position information from the robot and a triangulation based measurement principle. For triangulation, the correspondence problem is solved by a proven method, namely structured light. Assuming ideal image measurements, resulting 3D accuracy relies on correct modeling of robot and sensor geometry, as well as accurate knowledge of the model parameters. The modeling process touches different fields of expertise, from robotics over photogrammetry to computational geometry. It was shown that the resulting model can be calibrated in an automatic fashion, including camera geometry, projector geometry, and robot geometry. In this course, two novel contributions were proposed: robot calibration by self-inspection through a mirror and structure estimation through interpolation with generalized barycentric coordinates.

To complete the work with an outlook, one inevitably notices, that the hardware components by themselves are under a rapid development. Projection and camera systems are continuously miniaturized, with ever increasing resolution. This does not only have an effect on sensor miniaturization, but allows to increase redundancy in the number of projector and cameras. Especially the intelligent exploitation of redundant projectors is a topic which has not been addressed so far.

Another issue which was exempt from this work is user interaction. Many sensor capabilities require the application of geometrical "tricks", like the plane-based circle fitting, proposed in Chapter 8. These methods require a considerable amount of parametrization, which is hard to do if the underlying methods are not known to the engineer. On the other hand, most of these parameters are geometric in nature, and are intuitive, if presented in 3D, in camera image space, or directly on the object. This can also be realized with a projector-camera system. A guided task teaching system, which relies on the projector-camera system for user interaction, would hereby be a promising approach.

The proposed hardware system was developed to the state of a research prototype. Consequently, engineering issues like long-term stability or temperature sensitivity need to be investigated in detail.

Appendix A

Acronyms and Symbols

List of Acronyms

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
BRDF	bidirectional reflectance distribution function
CCD	charge coupled device
CCF	camera coordinate frame
CDS	correlated double sampling
CF	coordinate frame
CMM	coordinate measuring machine
CMOS	complementary metal oxide semiconductor
CPC	complete and parametrically continuous
CRF	camera response function
DH	Denavit Hartenberg
DLP	digital light processing
DLT	direct linear transform
DMD	digital micromirror device
DOF	degrees of freedom
DOM	degree of motion
DR	dynamic range
GigE	gigabit ethernet
GPU	graphics processing unit

HDR	high dynamic range
ICF	image coordinate frame
IR	infrared
IR-LED	infrared light emitting diode
ISO	International Organization for Standardization
LAN	local area network
LASER	light amplification by stimulated emission of radiation
LCD	liquid crystal display
LCoS	liquid crystal on silicon
LED	light emitting diode
LIDAR	light detection and ranging
LM	Levenberg-Marquardt
LVDT	linear variable displacement transducer
ML	maximum likelihood
MLE	maximum likelihood estimation
MRE	mean reprojection error
PDF	probability density function
PMD	photonic mixer device
PRF	projector response function
RMS	root mean square
RANSAC	random sample consensus
RPY	roll-pitch-yaw
SNR	signal to noise ratio
SOCP	second order cone programming
TOF	time of flight
VDI	Verein Deutscher Ingenieure
WCF	world coordinate frame

Bibliography

[1] ABB (2008). ABB Produkthandbuch IRB 2400 3HAC 7628-1 /M2000.

[2] Alici, G. and Shirinzadeh, B. (2003). Laser interferometry based robot position error modelling for kinematic calibration. In *Proc. International Conference on Intelligent Robots and Systems*, volume 4, pages 3588–3593 vol.3.

[3] Atkinson, K. B. (2001). *Close Range Photogrammetry and Machine Vision*. Whittles Publishing.

[4] Bai, Y., Zhuang, H., and Roth, Z. (2003). Experiment study of puma robot calibration using a laser tracking system. In *Proc. IEEE International Workshop on Soft Computing in Industrial Applications, SMCia*, pages 139–144.

[5] Bennett, D. and Hollerbach, J. (1991). Autonomous calibration of single-loop closed kinematic chains formed by manipulators with passive endpoint constraints. *IEEE Transactions on Robotics and Automation*, 7(5):597–606.

[6] Beraldin, J.-A., Rioux, M., Cournoyer, L., Blais, F., Picard, M., and Pekelsky, J. (2007). Traceable 3D imaging metrology. In Beraldin, J.-A., Remondino, F., and Shortis, M. R., editors, *Proc. SPIE Videometrics IX*, volume 6491.

[7] Beyer, H. (1992). Accurate calibration of CCD-cameras. In *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 96–101.

[8] Beyer, L. (2004). *Genauigkeitssteigerung von Industrierobotern*. PhD thesis, Fachbereich Maschinenbau der Helmut-Schmidt-Universität Universität der Bundeswehr Hamburg.

[9] Bimber, O. and Iwai, D. (2008). Superimposing dynamic range. In *SIGGRAPH Asia '08: ACM SIGGRAPH Asia 2008 papers*, pages 1–8, New York, NY, USA. ACM.

[10] Braeuer-Burchardt, C., Heinze, M., Munkelt, C., Kühmstedt, P., and Notni, G. (2006). Distance dependent lens distortion variation in 3D measuring systems using fringe projection. In Chantler, M. J., Trucco, E., and Fisher, R. B., editors, *British Machine Vision Conference*, volume 1, pages 327–336. British Machine Vision Association -BMVA-.

[11] Brenner, C., Boehm, J., and Guhring, J. (1999). Photogrammetric calibration and accuracy evaluation of a cross-pattern stripe projector. In *SPIE Conference on Videometrics*, volume 3641, pages 164–172, San Jose, California.

[12] Brink, W., Robinson, A., and Rodrigues, M. (2008). Indexing uncoded stripe patterns in structured light systems by maximum spanning trees. In *British Machine Vision Conference*, pages 575–584.

[13] Brown, D. C. (1964). An advanced plate reduction for photogrammetric cameras. *Air Force Cambridge Research Laboratories Report*, 40(64).

[14] Brown, D. C. (1966). Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3).

[15] Brown, D. C. (1971). Close range camera calibration. *Photogrammetric Engineering*.

[16] Brown, D. C. (1976). The bundle adjustment - progress and prospects. *Int. Archives of Photogrammetry*, 21:1–33.

[17] Carrihill, B. and Hummel, R. A. (1985). Experiments with the intensity ratio depth sensor. *Computer Vision, Graphics and Image Processing*, 32(3):337–358.

[18] Caspi, D., Kiryati, N., and Shamir, J. (1998). Range imaging with adaptive color structured light. In *Pattern Analysis and Machine Intelligence*, volume 20, pages 470–480, Los Alamitos, CA, USA. IEEE Computer Society.

[19] Chojnacki, W., Brooks, M., van den Hengel, A., and Gawley, D. (2000). On the fitting of surfaces to data with covariances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1294–1303.

[20] Chou, J. and Kamel, M. (1988). Quaternions approach to solve the kinematic equation of rotation, $A_a A_x = A_x A_b$, of a sensor-mounted robotic manipulator. In *Proc. IEEE International Conference on Robotics and Automation*, pages 656–662 vol.2.

[21] Clarke, T. and Fryer, J. (1998). The development of camera calibration methods and models. *Photogrammetric Record*, 16(91):51–66.

[22] Daniilidis, K. and Bayro-Corrochano, E. (1996). The dual quaternion approach to hand-eye calibration. In *Proc. International Conference on Pattern Recognition*, volume 1, page 318, Washington, DC, USA. IEEE Computer Society.

[23] Debevec, P. E. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH*, pages 369–378, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

[24] Denavit, J. and Hartenberg, R. S. (1955). A kinematic notation for lower-pair mechanisms based on matrices. *Transactions of the American Society of Mechanical Engineers, Journal of Applied Mechanics*, 23:215–221.

[25] Dusharme, D. (2004). 2004 metrology market survey. *Quality Digest Magazine*.

[26] Faugeras, O. and Luong, Q.-T. (2001). *The Geometry of Multiple Images*. MIT Press. ISBN: 0-262-06220-8.

[27] Floater, M. (2003). Mean value coordinates. *Computer Aided Geometric Design*, 20:19–27.

[28] Floater, M. S., Kos, G., and Reimers, M. (2005). Mean value coordinates in 3D. *Computer Aided Geometric Design*.

[29] Forbes, K., Voigt, A., and Bodika, N. (2004). Visual hulls from single uncalibrated snapshots using two planar mirrors. In *Proc. Fifteenth Annual Symposium of the Pattern Recognition Association of South Africa*.

[30] Fraser, C. S. (1997). Digital camera self-calibration. *ISPRS Journal of Photogrammetry and Remote Sensing*, 52(4):149–159.

[31] Fraser, C. S. and Shortis, M. R. (1992). Variation of distortion within the photographic field. *ISPRS Photogrammetric Engineering and Remote Sensing*, 6(58):851–855.

[32] Fraser, C. S., Shortis, M. R., and Ganci, G. (1995). Multisensor system self-calibration. *Videometrics IV*, 2598(1):2–18.

[33] Fratpietro, A. and Hayes, M. (2004). Relative measurement for kinematic calibration using digital image processing. In *Proc. CSME Forum*, pages 758–767.

[34] Fryer, J. and Brown, D. (1986). Lens distortion for close-range photogrammetry. *Photogrammetric Engineering and Remote Sensing*, 52:51–58.

[35] Fuller, W. A. (1987). *Measurement error models*. John Wiley & Sons.

[36] Gamal, A. E. (2002). High dynamic range image sensors. Tutorial International Solid State Circuits Conference.

[37] González, J. I., Gámez, J. C., Artal, C. G., and Cabrera, A. N. (2005). Stability study of camera calibration methods. In *Proc. VI Workshop en Agentes Fisicos, WAF 2005*.

[38] Goswami, A., Quaid, A., and M.Peshkin (1993). Identifying robot parameters using partial pose information. *IEEE Control Systems Magazine*, 13(5):6–14.

[39] Grossberg, M. D., Peri, H., Nayar, S. K., and Belhumeur, P. N. (2004). Making one object look like another: Controlling appearance using a projector-camera system. In *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 452–459.

[40] Gühring, J. (2001). Dense 3-D surface acquisition by structured light using off-the-shelf components. videometrics. In *Proc. Videometrics and Optical Methods for 3D Shape Measurement*, pages 22–23.

[41] Hartley, R. I. (1997). In defense of the eight-point algorithm. *Pattern Analysis and Machine Intelligence*, 19(6):580–593.

[42] Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518, second edition.

[43] Hayati, S., Tso, K., and Roston, G. (1988). Robot geometry calibration. *Proc. IEEE International Conference on Robotics and Automation*, 2:947–951.

[44] Heidl, W. (2006). A computer vision system for simultaneous calibration of a stereo camera rig and kinematic calibration of an articulated robot arm. Master's thesis, Institute for Computer Graphics and Vision, Graz University of Technology.

[45] Heikkila, J. and Silvén, O. (1996). Calibration procedure for short focal length off-the-shelf CCD-cameras. In *Proc. International Conference on Pattern Recognition*, page 166, Washington, DC, USA. IEEE Computer Society.

[46] Heikkila, J. and Silvén, O. (1997). A four-step camera calibration procedure with implicit image correction. In *IEEE Computer Vision and Pattern Recognition or CVPR*, page 1106, Washington, DC, USA. IEEE Computer Society.

[47] Hollerbach, J. M. and Wampler, C. W. (1996). The calibration index and taxonomy for robot kinematic calibration methods. *International Journal of Robotics Research*, 15(6):573–591.

[48] Horaud, R. and Dornaika, F. (1995). Hand-eye calibration. *International Journal of Robotics Research*, 14(3):195–210.

[49] Inokuchi, S., Sato, K., and Matsuda, F. (1984). Range imaging system for 3-D object recognition. In *Proc. International Conference on Pattern Recognition*, pages 806–808.

[50] ISO (1999). ISO 9283:1998 Industrieroboter Leistungskenngroeszen und zugehoerige Pruefmethoden.

[51] ISO10360-4 (2000). Geometrical product specifications (GPS): Acceptance and reverification tests for coordinate measuring machines (CMM).

[52] Je, C., Lee, S. W., and Park, R.-H. (2004). High-contrast color-stripe pattern for rapid structured-light range imaging. In *Proc. European Conference on Computer Vision*, volume 3021. Springer Berlin / Heidelberg.

[53] Ju, T., Liepa, P., and Warren, J. (2007). A general geometric construction of coordinates in a convex simplicial polytope. *Computer Aided Geometric Design*, 24(3):161–178.

[54] Ju, T., Schaefer, S., and Warren, J. (2005a). Mean value coordinates for closed triangular meshes. *ACM Transactions on Graphics*, 24(3):561–566.

[55] Ju, T., Schaefer, S., Warren, J., and Desbrun, M. (2005b). Geometric construction of coordinates for convex polyhedra using polar duals. In *Proc. Eurographics Symposium on Geometric Processing*.

[56] Judd, R. and Knasinski, A. (1990). A technique to calibrate industrial robots with experimental verification. *IEEE Transactions on Robotics and Automation*, 6(1):20–30.

[57] Kahl, F. and Hartley, R. (2008). Multiple view geometry under the l-infinity norm. *Pattern Analysis and Machine Intelligence*, 30(9):1603–1617.

[58] Kang, H.-J., Jeong, J.-W., Shin, S.-W., Suh, Y.-S., and Ro, Y.-S. (2007). Autonomous kinematic calibration of the robot manipulator with a linear laser-vision sensor. In *Proc. Advanced Intelligent Computing Theories and Applications. With Aspects of Artificial Intelligence*, pages 1102–1109.

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

[60] Kawasaki, H., Furukawa, R., Sagawa, R., and Yagi, Y. (2008). Dynamic scene shape reconstruction using a single structured light pattern. In *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 1–8.

[61] Keyence (2003). CCD laser-wegmesssensoren.

[62] Knapitsch, A. (2008). Adaptive structured light in a multi camera and multi projector setup. Master's thesis, Institute for Computer Graphics and Vision, Graz University of Technology.

[63] Koninckx, T. and Van Gool, L. (2006). Real-time range acquisition by adaptive structured light. *Pattern Analysis and Machine Intelligence*, 28(3):432–445.

[64] Koninckx, T. P., Peers, P., Dutré, P., and Gool, L. J. V. (2005). Scene-adapted structured light. In *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 611–618.

[65] Kopparapu, S. K. and Corke, P. (1999). The effect of measurement noise on intrinsic camera calibrationparameters. In *Proc. IEEE International Conference on Robotics and Automation*, volume 2, pages 1281–1286.

[66] Kumar, R., Ilie, A., Frahm, J.-M., and Pollefeys, M. (2008). Simple calibration of non-overlapping cameras with a mirror. *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 1–7.

[67] Kyle, S. A. (1993). Non-contact measurement for robot calibration. *Robot Calibration*, pages 78–100.

[68] Leedan, Y. and Meer, P. (2000). Heteroscedastic regression in computer vision: Problems with bilinear constraint. *International Journal of Computer Vision*, 37(2):127–150.

[69] Lei, S., Jingtai, L., Weiwei, S., Shuihua, W., and Xingbo, H. (2004). Geometry-based robot calibration method. In *Proc. IEEE International Conference on Robotics and Automation*, volume 2, pages 1907–1912 Vol.2.

[70] Lenz, M. (2008). Solution of an industrial measurement task with a robot assisted structured light sensor. Master's thesis, Graz University of Technology.

[71] Luhmann, T. (2003). *Nahbereichsphotogrammetrie*. Wichmann.

[72] Maas, H. (1997). Dynamic photogrammetric calibration of industrial robots. *Videometrics V*, 3174:106–112.

[73] Mallon, J. and Whelan, P. F. (2006). Which pattern? biasing aspects of planar calibration patterns and detection methods. *Pattern Recognition Letters*, 28(8):921–930.

[74] Mann, S. and Picard, R. (1995). Being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *Proc. IS&T 46th annual conference*, pages 422–428.

[75] Maruyama, M. and Abe, S. (1993). Range sensing by projecting multiple slits with random cuts. *Pattern Analysis and Machine Intelligence*, 15(6):647–651.

[76] Mayer, H. (2000). From surface roughness to surface displacements. In *Proc. IAPR Workshop on Machine Vision Applications, MVA2000*, pages 611–614.

[77] Meng, Y. and Zhuang, H. (2001). Self-calibration of camera-equipped robot manipulators. *International Journal of Robotics Research*, 20(11):909–921.

[78] Mooring, B. and Padavala, S. (1989). The effect of kinematic model complexity on manipulator accuracy. In *Proc. IEEE International Conference on Robotics and Automation*, pages 593–598.

[79] Motta, J., de Carvalho, G., and McMaster, R. (2001). Robot calibration using a 3D vision-based measurement system with a single camera. *Robotics and Computer-Integrated Manufacturing*, 11:487–497.

[80] Mühlich, M. and Mester, R. (1998). The role of total least squares in motion analysis. In *Proc. European Conference on Computer Vision*, pages 305–321.

[81] Nahvi, A., Hollerbach, J., and Hayward, V. (1994). Calibration of a parallel robot using multiple kinematic closed loops. In *Proc. IEEE International Conference on Robotics and Automation*, pages 407–412 vol.1.

[82] Narasimhan, S. G., Koppal, S. J., , and Yamazaki, S. (2008). Temporal dithering of illumination for fast active vision. In *Proc. European Conference on Computer Vision*, pages 830–844.

[83] Nayar, S. K., Krishnan, G., Grossberg, M. D., and Raskar, R. (2006). Fast separation of direct and global components of a scene using high frequency illumination. In *ACM SIGGRAPH*, pages 935–944, New York, NY, USA. ACM.

[84] Nayar, S. K. and Mitsunaga, T. (2000). High dynamic range imaging: Spatially varying pixel exposures. In *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 1472–1479.

[85] Oggier, T., Buettgen, B., and Lustenberger, F. (2006). Swissranger SR3000 and First Experiences based on Miniaturized 3D-TOF Cameras. Technical report, Swiss Center for Electronics and Microtechnology.

[86] Phong, B. T. (1975). Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317.

[87] Pollefeys, M. (1999). *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*. PhD thesis, Katholieke Universiteit Leuven.

[88] Posdamer, J. L. and Altschuler, M. D. (1982). Surface measurement by space-encoded projected beam systems. *Computer Graphics and Image Processing*, 18(1):1–17.

[89] Prenninger, J. P. and Gander, H. (1993). *Robot Calibration*, chapter Measuring dynamic robot movements in 6 DOF and real time, pages 124–153. Chapman & Hall.

[90] Remondino, F. and Fraser, C. (2006). Digital camera calibration methods: Considerations and comparisons. In *Proc. ISPRS Commission V Symposium 'Image Engineering and Vision Metrology'*.

[91] Renaud, P., Andreff, N., Lavest, J.-M., and Dhome, M. (2006). Simplifying the kinematic calibration of parallel mechanisms using vision-based metrology. *IEEE Transactions on Robotics*, 22(1):12–22.

[92] Roberts, K. S. (1988). A new representation for a line. In *IEEE Computer Vision and Pattern Recognition or CVPR*, pages 635–640.

[93] Robinson, P., Orzechowski, P., James, P., and Smith, C. (1997). An automated robot calibration system. In *Proc. IEEE International Symposium on Industrial Electronics*, volume 1, pages 285–290.

[94] Salvi, J., Armangué, X., and Batlle, J. (2002). A comparative review of camera calibrating methods with accuracy evaluation. *Pattern Recognition*, 35(7):1617–1635.

[95] Sciavicco, L. and Siciliano, B. (2005). *Modelling and Control of Robot Manipulators*. Advanced Textbooks in Control and Signal Processing. Springer.

[96] Shiu, Y. and Ahmad, S. (1989). Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form AX=XB. *IEEE Transactions on Robotics and Automation*, 5(1):16–29.

[97] Strobl, K. H. and Hirzinger, G. (2006). Optimal hand-eye calibration. *Proc. International Conference on Intelligent Robots and Systems*, pages 4647–4653.

[98] Sturm, P. and S.J.Maybank (1999). On plane-based camera calibration: A general algorithm, singularities, applications. In *IEEE Computer Vision and Pattern Recognition or CVPR*, volume 1.

[99] Sun, W. and Cooperstock, R. (2006). An empirical evaluation of factors influencing camera calibration accuracy using three publicly available techniques. *Machine Vision and Applications*, 17(1):51 – 67.

[100] Tajima, J. and Iwakawa, M. (1990). 3-D data acquisition by rainbow range finder. *Proc. International Conference on Pattern Recognition*, i:309–313 vol.1.

[101] Tsai, R. (1986). An efficient and accurate camera calibration technique for 3D machine vision. In *IEEE Computer Vision and Pattern Recognition or CVPR*.

[102] Tsai, R. (1987). A versatile camera calibration technique for high accuracy 3D machine vision using off-the-shelf tv cameras and lenses. *IEEE Journal of Robotics and Automation*.

[103] Tsai, R. and Lenz, R. (1989). A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Transactions on Robotics and Automation*, 5(3):345–358.

[104] VDI/VDE2617 (1989). Accuracy of coordinate measuring machines part 3.

[105] Veitschegger, W. and Wu, C.-H. (1988). Robot calibration and compensation. *IEEE Journal of Robotics and Automation*, 4(6):643–656.

[106] Wachspress, E. (1975). *A Rational Finite Element Basis*. Academic Press.

[107] Wang, C.-C. (1992). Extrinsic calibration of a vision sensor mounted on a robot. *IEEE Transactions on Robotics and Automation*, 8(2):161–175.

[108] Wang, J., Shia, F., Zhanga, J., and Liu, Y. (2007). A new calibration model of camera lens distortion. *Pattern Recognition*, 41(2):607–615.

[109] Warren, J. (1996). Barycentric coordinates for convex polytopes. *Advances in Computational Mathematics*, 6:97–108.

[110] Wei, S. and Cooperstock, J. (2005). Requirements for camera calibration: Must accuracy come with a high price? In *Proc. Application of Computer Vision, 2005. WACV/MOTIONS '05 Seventh IEEE Workshops on*, pages 356–361.

[111] Weise, T., Leibe, B., and Gool, L. J. V. (2007). Fast 3D scanning with automatic motion compensation. In *IEEE Computer Vision and Pattern Recognition or CVPR*.

[112] Wozniak, A. (2007). New method for testing the dynamic performance of cmm scanning probes. *Instrumentaiton and Measurement, IEEE Transactions on*, 56(6):2767–2774.

[113] Xu, Y. and Aliaga, D. G. (2007). Robust pixel classification for 3D modeling with structured light. In *Graphics Interface*, pages 233–240.

[114] Zhang, Z. (1997). Parameter estimation techniques: A tutorial with application to conic fitting. *Image and Vision Computing*, 15(1):59–76.

[115] Zhang, Z. (1998). On the optimization criteria used in two-view motion analysis. *Pattern Analysis and Machine Intelligence*, 20(7):717–729.

[116] Zhang, Z. (1999). Flexible camera calibration by viewing a plane from unknown orientations. In *Proc. International Conference on Computer Vision*, volume 1, pages 666–673.

[117] Zhuang, H., Roth, Z., and Hamano, F. (1992). A complete and parametrically continuous kinematic model for robot manipulators. *IEEE Transactions on Robotics and Automation*, 8(4):451–463.

[118] Zhuang, H. and Roth, Z. S. (1995). *Camera Aided Robot Calibration*. CRC Press, Inc., Boca Raton, FL, USA.

[119] Zhuang, H., Wang, K., and Roth, Z. (1995). Simultaneous calibration of a robot and a hand-mounted camera. *IEEE Transactions on Robotics and Automation*, 11(5):649–660.

[120] Zuang, H. and Shiu, Y. C. (1993). A noise-tolerant algorithm for robotic hand-eye calibration with or without sensor orientation measurement. *IEEE International Conference on Systems, Man and Cybernetics*, 23(4):1168–1175.