3D Scene Reconstruction by Integration of Photometric and Geometric Methods

Pablo d'Angelo

July 2007

Dissertation zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.)

Dipl.-Ing. (FH) Pablo d'Angelo
Environment Perception
Group Research
DaimlerChrysler AG
email: pablo.dangelo@web.de

Abdruck der genehmigten Dissertation zur Erlangung des akademischen Grades
Doktor der Ingenieurwissenschaften (Dr.-Ing.).
Der Technischen Fakultät der Universität Bielefeld
am 2. April 2007 vorgelegt von Pablo d'Angelo,
am 12. Juli 2007 verteidigt und genehmigt.

Gutachter:
Prof. Dr.-Ing. Franz Kummert, Universität Bielefeld
Dr. rer. nat. Christian Wöhler, DaimlerChrysler AG

Prüfungsausschuss:
Prof. Dr.-Ing. Franz Kummert, Universität Bielefeld
Dr. rer. nat. Christian Wöhler, DaimlerChrysler AG
Prof. Dr.-Ing. Holger Theisel, Universität Bielefeld
Dr. Peter Steffen, Universität Bielefeld

Dedicated to Jingping Liu

Acknowledgements

I would like to sincerely thank the many persons who, through their continuous support, encouragement and advice have helped to complete this work.

First and foremost, I would like to thank my advisor, Dr. rer. nat. Christian Wöhler for being an excellent mentor and teaching me how to do research. A bit more than 3 years ago, he convinced me to start working on this thesis and his focus on research has enabled me to finish this thesis.

Prof. Dr.-Ing. Franz Kummert, my doctoral advisor, and Prof. Dr.-Ing. Gerhard Sagerer have provided valuable feedback. Prof. Dr.-Ing. Rainer Ott has carefully read the draft and provided important comments which have lead to significant improvements.

The environment perception group at the DaimlerChrysler Research Centre has been a very stimulating environment. I would like to thank Dr.-Ing. Ulrich Kressel, Dipl. Inf. Annika Kuhl, Dipl. Inf. Lars Krüger, Dipl. Ing. (FH) Kia Hafezi, Dipl. Ing. Marc Ellenrieder and Dipl. Ing. Frank Lindner for many inspiring discussions, help and generally providing an enjoyable and friendly atmosphere.

Finally, I am forever indebted to my family and especially my wife Jingping for their understanding, endless patience and encouragement when it was most required.

Contents

Abstract

In this thesis, we have developed a framework for image-based 3D reconstruction of sparse point clouds and dense depth maps. The framework is based on self-consistent integration of geometric and photometric constraints on the surface shape, such as triangulation, defocus and reflectance. The reconstruction of point clouds starts by tracking object features over a range of distances from the camera with a small depth of field, leading to a varying degree of defocus for each feature. Information on absolute depth is obtained based on a Depth from Defocus approach. The parameters of the point spread functions estimated by Depth from Defocus are used as a regularisation term for Structure from Motion. The reprojection error obtained from bundle adjustment and the absolute depth error obtained from Depth from Defocus are simultaneously minimised for all tracked object features. The proposed method yields absolutely scaled 3D coordinates of the scene points without any prior knowledge about either scene structure or the camera motion. Another part of the framework is the estimation of dense depth maps based on intensity and polarisation reflectance and absolute depth data from arbitrary sources, eg. the Structure from Motion and Defocus method. The proposed technique performs the analysis on any combination of single or multiple intensity and polarisation images. To compute the surface gradients, we present a global optimisation method based on a variational framework and a local optimisation method based on solving a set of nonlinear equations individually for each image pixel. These approaches are suitable for strongly non-Lambertian surfaces and those of diffuse reflectance behaviour and can also be adapted to surfaces of non-uniform albedo. We describe how independently measured absolute depth data is integrated into the Shape from Photopolarimetric Reflectance (SfPR) framework in order to increase the accuracy of the 3D reconstruction result. We evaluate the proposed framework on both synthetic and real-world data. The Structure from Motion and Defocus algorithm yields relative errors of absolute scale of usually less than 3 percent. In our real-world experiments with SfPR, we regard the scenarios of 3D reconstruction of raw forged iron surfaces in the domain of industrial quality inspection and the generation of a digital elevation model of a section of the lunar surface. The obtained depth accuracy is better than the lateral pixel resolution.

1 Introduction

Three-dimensional object and surface reconstruction from images is an important topic in various application areas, such as quality inspection, reverse engineering, robotics, geography and archaeology.

In the domain of quality inspection, a large number of inspection tasks depend on 3D reconstruction techniques. Examples are the detection of defects such as small dents on a variety of surfaces, for example on forged or cast metallic surfaces. Tasks of this kind usually require the accurate measurement of depth on small surfaces. Other tasks depend on the precise measurement of a sparse set of well defined points, for example to determine if an assembly process has been completed with the required accuracy, or measurement of the relative movement between important parts during a crash test.

In the field of cartography and astrogeology, images captured from air- or spacecraft are used to reconstruct the ground topography of the earth or other planets with high detail. 3D reconstruction plays an important role in autonomous robotic systems, for example during exploration of unknown terrain. The 3D reconstruction of archaeological excavations and historic objects is also an important application area in the field of archaeology.

Many methods for 3D reconstruction from images exist, they can be categorized into geometric methods, which are based on the modelling of the geometric aspects of image creation, and photometric methods, which are primarily based on photometric modelling. The various application scenarios have different requirements on the reconstruction. For some tasks, it is sufficient to produce a sparse set of 3D points, where 3D information is available only for a very small number of pixels in the input images, while others require a dense reconstruction, with 3D information available for every pixel in the input images. Other important factors include the size, shape and material of the objects, the number of required images, requirements on positions of the cameras or light sources, and the time allowed for image capture and reconstruction.

Reconstruction methods need be chosen carefully considering the requirements of the reconstruction task. For some tasks, no existing method might be applicable and new methods need to be developed.

1.1 Aim and scope of this thesis

Well known geometric approaches to 3D reconstruction include passive methods such as Stereo, Structure from Motion (SfM) and Depth from Defocus (DfD), which do not require a specific, structured illumination. Active methods include the projection of structured light and triangulation of laser lines. While active methods can result in dense and precise measurements, they require a higher instrumentation effort, resulting in high equipment costs. The acquisition time for scanning methods such as laser line triangulation is also a factor that needs to be taken into account. If a dense reconstruction is desired, passive methods such as Stereo, Depth from Defocus, and Structure from Motion often require structured illumination to artificially produce texture required for a dense reconstruction of the surface. Additionally, Structure from Motion can only recover 3D shape and camera poses up to a scale factor, if no information about scene or camera motion is available.

Examples for photometric reconstruction algorithms are Shape from Shading (SfS) and Shape from Polarisation. In contrast to passive geometric approaches, they can be used for dense reconstruction of smooth, textureless surfaces without structured illumination. Despite Shape from Shading is an appealing method, it has been applied with limited success to real world reconstruction problems with non-Lambertian reflectance.

Often, methods based on different principles (triangulation, defocus, shading etc.) can be used to solve a reconstruction problem, and each method has its strength in different areas. A combined method which is based on multiple principles has the potential to combine these strengths to archive a better reconstruction result. The development and analysis of such combined methods is the main topic of this thesis.

The main points investigated in this thesis are:

- Can Structure from Motion be extended to recover absolutely scaled coordinates without any knowledge about scene structure or camera motion?
- Does an extended reflectance model which includes polarisation effects during reflection of light at a surface improve the photometric reconstruction?

- How can independently obtained depth information, for example obtained using Stereo or Structure from Motion, be used to improve the accuracy of Shape from Shading and polarisation?

Applications of dense photometric surface reconstruction include the quality inspection of rough metallic surfaces for small defects. This is a particularly hard problem due to the specular reflections and the roughness of the surface. Passive geometric methods such as Stereo and Structure from Motion can only reconstruct a very sparse set of points reliably since the determination of corresponding points on strongly specular surfaces is a very hard problem.

1.2 Notational conventions

If possible, the conventions of the major cited works are used to avoid confusion and allow the reader familiar with the respective literature to grasp the concepts quickly. In case several different coordinate systems are used simultaneously, the notation by Craig (1989) will be used to clearly state the coordinate system in which a point is defined. Using this notation, a point $\mathbf{x}$ defined in the coordinate system C is written as ${}^{C}\mathbf{x}$.

1.2.1 Abbreviations

Several abbreviation are used throughout the whole text. They are defined at the place a term is first introduced, but are also listed here for reference.

DfD	Depth from Defocus
DfF	Depth from Focus
KLT	Kanade Lucas Tomasi Tracker
PDE	partial differential equation
PSF	point spread function
ROI	region of interest
RMSE	root mean square error
SfM	Structure from Motion
SfS	Shape from Shading
SfPR	Shape from Photopolarimetric Reflectance

1.3 Section overview

The state of the art in object and surface reconstruction relevant for this thesis is discussed in Part I. It briefly describes the major 3D reconstruction methods related to this thesis.

Part II presents a flexible object and surface reconstruction system, which consists of several novel surface reconstruction methods. The reconstruction of point clouds with absolute scale by integration of Structure from Motion and Depth from Defocus is developed in chapter 7. This method estimates the scene points and camera motion by minimising a combined Structure from Motion and defocus error term.

Shape from Photopolarimetric Reflectance (SfPR), a surface reconstruction method based on Shape from Shading and Polarisation, is developed in chapter 8. By modelling not only the intensity but also the polarisation reflectance behaviour, additional constraints on the surface shape are provided. This method is extended in chapter 9 to include a independently acquired depth information, such as a sparse point cloud created by Stereo or the Structure from Motion and Defocus proposed in this thesis.

The developed methods are evaluated in Part III with synthetic and real data from various application areas.

Part I

State of the art

2 Geometric methods

Reconstruction of shape from two-dimensional images has been an important topic since the invention of photography by Niepce and Daguerre in 1839. Laussedat and Meydenbauer developed the first photogrammetric methods in the mid 19th century (Luhmann 2003), used for mapping and reconstruction of buildings. The early photogrammetric methods were based on the geometric modelling of the image formation, exploiting the perspective projection of the 3D scene onto a flat 2D image plane.

The camera model used by most photogrammetric and computer vision approaches is the pinhole camera (cf. Fig. 2.1). The projection of a 3D point given in the camera coordinate system C, ${}^{C}\mathbf{x} = [x_1, y_1, z_1]$ into ${}^{J}\mathbf{x} = [u_1, v_1]$ in image coordinates can be denoted by the projection function $\mathcal{P}$:

$$ {}^{J}\mathbf{x} = \mathcal{P}(K, {}^{C}\mathbf{x}) \tag{2.1} $$

The parameter K defines the internal (focal length, lens distortion parameters) camera orientation. The projection function of a pinhole camera is given by

$$ \begin{aligned} u_1 &= -f\frac{x_1}{z_1} \\ v_1 &= -f\frac{y_1}{z_1}, \end{aligned} \tag{2.2} $$

where f is the distance between pinhole and image plane. Once multiple cameras are considered, it is practical to introduce a world coordinate system W, and specify the orientation T_i of each camera relative to this world coordinate system. Then the projection function of a point in the world coordinate system needs to be transformed into the camera coordinate system of the ith camera, C_i using a camera orientation T_i. In this case the projection function depends on both internal orientation K_i and external orientation T_i: ${}^{J_i}\mathbf{x} = \mathcal{P}(T_i, K_i, {}^{C}\mathbf{x})$.

In the computer vision community the internal camera orientation parameters are known as intrinsic camera parameters, while the external orientation parameters are known as extrinsic camera parameters. The projected points are then captured by a light sensitive

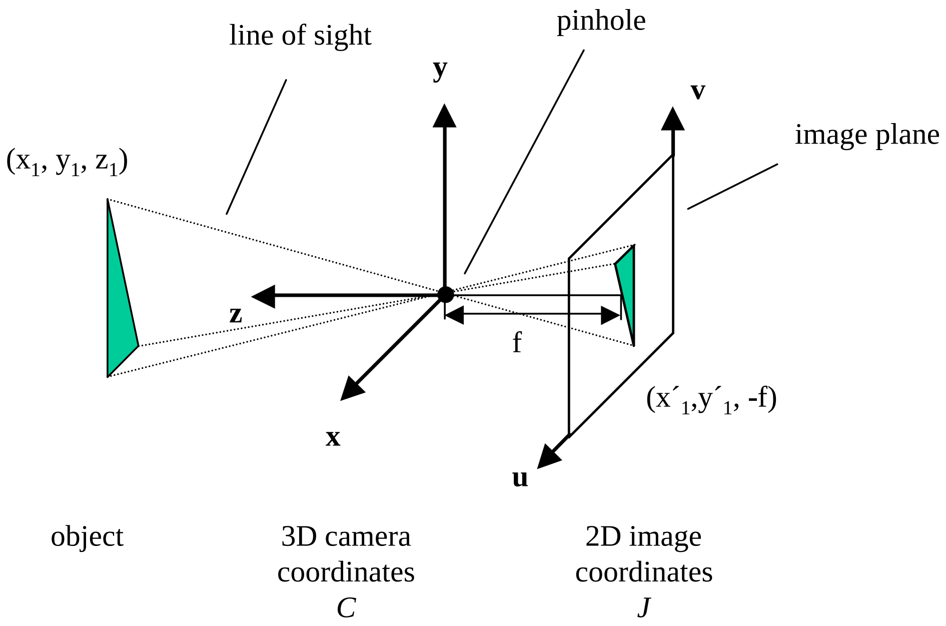


Figure 2.1: Pinhole projection

device, typically a film or digital sensor. In the case of a digital sensor, the light sensitive area is sampled and the light intensity is measured at each sample point (Luhmann 2003).

Note that the 3D point in camera coordinates ${}^{C}\mathbf{x}$ cannot be determined uniquely given camera parameters K and image point ${}^{J}\mathbf{x}$, since they only define a ray in C on which ${}^{C}\mathbf{x}$ is located.

2.1 Projective Geometry

Projective geometry is a powerful mathematical tool for the analysis of 3D reconstruction from projective images, it is however not used extensively in this thesis, therefore only a brief overview is given. Birchfield (1998) provides a readable and easily understandable introduction to projective geometry, while Faugeras (1993) and Hartley and Zisserman (2004) give a thorough analysis and in-depth details on the application of projective geometry to various computer vision problems.

In the above section, Euclidean geometry is used to describe the projection of the 3D world coordinates into 2D image coordinates, resulting in a nonlinear projection function $\mathcal{P}$ (cf. Eq. (2.2)), a transformation which does not preserve invariants such as length, angles and parallelism. In the Euclidean geometry, rotation and translation preserve these invariants.

Projective geometry allows a much larger class of invariant preserving transformations than Euclidean geometry. Among its invariant transformations are scaling and shear, as well as perspective projection. This makes it a suitable tool for the analysis of images captured by pinhole cameras. Euclidean geometry is actually a subset of projective geometry (Birchfield 1998).

Projective geometry uses homogeneous coordinates, for example a point (x, y, z) in three dimensional Euclidean geometry is represented in the 3D projective space by homogeneous coordinates $(X, Y, Z, W) = (x, y, z, 1)$. Overall scaling is unimportant, $(X, Y, Z, W) = (\alpha X, \alpha Y, \alpha Z, \alpha W)$ for any nonzero α. To transform a point in 3D projective space into euclidean coordinates, we simply divide it by the last coordinate, W. The general transform in 3D projective space is a matrix multiplication by a 4×4 matrix: $\mathbf{x}' = \mathbf{T}\mathbf{x}$. For a projection into a plane, a 3×4 matrix is sufficient, the projection of a point onto a plane located at the origin, spanning the x and y directions (cf. Eq. (2.2)) is given by

$$ {}^{J}\mathbf{x} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} {}^{W}\mathbf{x}. \tag{2.3} $$

This formulation of the projection is often used in both computer vision and computer graphics (Foley et al. 1993). Apart from simply applying projective transformations, the image formation can be modelled in a consistent way using linear algebra. An important projective transform is the essential matrix, which contains the extrinsic camera parameters of two pinhole cameras observing the same scene. The fundamental matrix is a generalisation of the essential matrix and additionally contains the intrinsic camera parameters. These matrices can be estimated directly using point correspondences in the image plane, the estimation of the corresponding world coordinates ${}^{W}\mathbf{x}$ of the points is not required.

2.2 3D reconstruction

Most geometric methods for 3D reconstruction are based on corresponding points in multiple images. For a single scene point ${}^{W}\mathbf{x}$, which is captured on N images, the corresponding image points ${}^{J_i}\mathbf{x}$ in each image i can be determined, either manually or by automatic correspondence search methods. Automatic detection of corresponding points arbitrary scenes is a challenging problem and an active research area (Shi and Tomasi 1994; Moreels and

Perona 2005). Together with the external and internal camera orientation, each point ${}^{J_i}\mathbf{x}$ forms a ray and ideally, all N rays intersect in the scene point ${}^{W}\mathbf{x}$.

The first general scene reconstruction methods that can cope with an arbitrary motion of the camera were developed by photogrammetrists in the late 19th and early 20th century (Finsterwalder 1899; Kruppa 1913), see also (Luhmann 2003; Åström 1996). These methods seek to determine internal and external camera orientation and 3D coordinates of the scene points.

Kruppa (1913) describes an analytical solution to recover relative camera orientation in an image pair, using 5 corresponding image points. It is an example of a minimal case method, as it solves for scene and camera parameters using a minimal number of point correspondences. Recently, computationally efficient and stable 5 point algorithms have been proposed by Stewénius et al. (2006).

2.3 Bundle adjustment

In classical bundle adjustment (Brown 1958; Luhmann 2003; Lourakis and Argyros 2004), scene points (structure) and camera orientation (motion) are recovered jointly and optimally from corresponding image points. The bundle adjustment error term

$$E_{\mathrm{B}}\left(\{T_i\},\{X_j\}\right) = \sum_{i=1}^{N}\sum_{j=1}^{M}\left[\mathcal{P}\left(T_i, K_i, X_j\right) - \mathbf{x}_{ji}\right]^2 \tag{2.4}$$

can be used to minimise the re-projection error E_{B} with respect to the unknown N internal camera orientations K_i, external camera orientations T_i and the M scene points X_j. Here, $\mathbf{x}_{ji}$ denotes the given 2D pixel coordinates (u_{ji}, v_{ji}) of feature j in image i.

Bundle adjustment is a very flexible method, depending on the reconstruction task, values for all or some of parameters K_i, T_i and X_j might be unknown. By minimising equation (2.4) with respect to the unknown parameters, the bundle adjustment method can be used for calibration of internal and/or external camera parameters as well as pose estimation of objects. The method can be applied to image sequences acquired by the same camera, or to images acquired simultaneously by multiple cameras. It is also possible to use cameras with different projection functions $\mathcal{P}$, for example pinhole and fish-eye cameras, in the same reconstruction task. If additional information about the scene is available (for example, the position of some 3D points in world coordinates), additional terms can be added to

Eq. (2.4). Measurement uncertainties of the known variables can be used to estimate the uncertainty of the estimated parameters.

Eq. (2.4) is a nonlinear equation and usually minimised using the Levenberg-Marquardt or Gauss-Newton algorithm. Even bundle adjustment tasks with many unknowns can be optimised efficiently, since the reprojection error of the jth point in view i only influences T_i, K_i in frames where the point j could be tracked as well as X_j. This leads to a sparse set of equations, which can be exploited by the optimisation algorithm. (Brown 1958; Lourakis and Argyros 2004). Since Eq. (2.4) is usually a strongly non-linear function with many local minima, suboptimal solutions may be found by the minimisation algorithm, if no reasonable initial guesses for the parameters are available. In some applications, like aerial photogrammetry, initial values are readily available. If they are not, initial values have to be computed using another algorithm, for example by repeated application of a minimal point method, or a linear Structure from Motion algorithm such as the factorisation algorithm by Tomasi and Kanade (1992). Bundle adjustment can only recover the scene up to an unknown scale factor, but this factor can be calculated if additional information about the scene, e.g. a distance between two scene points is known.

In general bundle adjustment provides accurate reconstruction of scene points for which correspondences could be established. Problems also occur when the correspondences contain outliers that do not comply with the assumption of a gaussian reprojection error distribution. In that case the estimated parameters can contain gross errors that are not directly apparent in the statistics of the estimated parameters. Ways to work around outliers are based on screening the data for outliers, for example using RANSAC (Fischler and Bolles 1981) together with a minimal case 5 point algorithm (Nister 2004), or using a M-Estimator[1] while minimising Eq. (2.4). Usually correspondences can only reliably be extracted in high contrast image areas, resulting in a sparse 3D reconstruction, where areas with uniform or repetitive texture cannot be reconstructed.

2.4 Stereo vision

If the scene is observed by two cameras with known internal and external orientation, the epipolar constraint can be used. This setup is exploited in the stereo vision approach to 3D reconstruction. The epipolar constraint simplifies the correspondence search problem,

---

[1] When using an M-Estimator, the L-2 norm in Eq. (2.4) is replaced by a norm that is less sensitive to large errors (which are assumed to be outliers).

because the epipolar constraint limits the correspondence search region for a given point in one image to a single line in the other image. Additionally, each 3D point can be calculated directly through triangulation such that no bundle adjustment is required. Due to these simplifications, stereo vision is a widely used technique in close range 3D reconstruction. In most Stereo systems two views with known internal and external camera orientation are used. In a typical Stereo application, a scene is simultaneously monitored by two cameras whose centres of projection are separated by the baseline distance. In many practical applications, the optical axes of the two cameras are parallel, and the images are taken with the same focal length. This is often called the standard stereo geometry and leads to epipolar lines oriented parallel to image rows or columns, where the correspondences can be found efficiently. It is possible to transform images from an arbitrary camera setup into images with horizontal or vertical epipolar lines, using a process known as stereo rectification (Ayache 1991; Bouguet 1997; Krüger et al. 2004).

Several surveys (Barnard and Fischler 1982; Dhond and Aggarwal 1989; Scharstein and Szeliski 2002; Zhang 2003; Brown et al. 2003) provide an exhaustive overview of the different Stereo methods. Given the internal parameters (focal length, distortion parameters, location of the principal point) and external parameters (position and orientation) of the cameras, the distance of objects or object parts in the scene is obtained by computing the disparity map, i. e. the offset between the pixels in both images that belong to the same scene point, respectively, and transforming them into world coordinates. The internal and external camera parameters can be determined using camera calibration (Bouguet 1997; Krüger et al. 2004).

Robust determination of the corresponding points and thus of disparity is the central problem to be solved by stereo vision algorithms. An early survey by Barnard and Fischler (1982) reports the use of block and feature matching. Block matching approaches compare a small area in one image with potentially matching areas in the other image. Often cross-correlation or the sum of squared differences are used as matching criteria. This assumes structures parallel to the image plane (fronto-parallel structures). At depth discontinuities or tilted areas, a block will contain pixels from different depths leading to less reliable matching results. Additionally uniform image areas cannot be matched reliably. Feature matching approaches extract suitable features like edges or curves (Wöhler and Krüger 2003) and match these by computing suitable similarity measures. Since these features are usually well localized, feature based methods handle depth discontinuities better, but might provide a sparse disparity map, compared to block matching.

The robustness and accuracy of the disparity estimates can be improved by considering additional constraints during the matching process. For example the smoothness constraint states that the disparity should vary smoothly, this is especially useful for uniform areas where no correspondences can be established. The ordering constraint states that for opaque surfaces the order of correspondences is always preserved. Many Stereo algorithms (Intille and Bobick 1994; Cox et al. 1996) use dynamic programming (Cormen et al. 1990) to efficiently and optimally calculate the disparity values of a complete scanline while considering the ordering constraint. Constraints over the whole image, across several scanlines, are hard to integrate into the dynamic programming framework. Algorithms based on graph cuts (Cormen et al. 1990; Boykov and Kolmogorov 2004) can use the constraints globally during the reconstruction and are among the best performing Stereo algorithms (in terms of reconstruction quality) (Scharstein and Szeliski 2002).

3 Real aperture methods

The geometric methods described in the previous section are all based on an ideal camera, which projects scene points into image points perfectly. However, a real camera system uses a lens of finite aperture, which results in images with a limited depth of field. The depth dependent blurring is not considered by the geometric methods and usually decreases the accuracy of the correspondence search methods.

The depth dependent defocussing is illustrated in Fig. 3.1, where a scene point at distance d_0 is in focus (projected onto a single point in the image plane located at distance v), while points at other distances $d \neq d_0$ to the camera are spread onto a larger area, leading to a blurred image. If the light rays are traced geometrically, object points that are out of focus will be imaged to a circular disk. This disk is known to photographers as the circle of confusion. Using the lens law

$$\frac{1}{v} + \frac{1}{d} = \frac{1}{f}, \tag{3.1}$$

its diameter C can be approximated (Pentland 1987) by

$$C = Dv \left(\frac{1}{f} - \frac{1}{v} - \frac{1}{d} \right), \tag{3.2}$$

where f is the focal length and D is the diameter of the lens aperture. The depth dependent term $1/d$ approaches zero for larger values of d while the other terms stay constant, resulting in little change of the blur radius for objects at a large distances d. This limits the real aperture methods to close range scenarios, where two different depth values result in a measurable change of C.

In optical systems, the image projected onto the image plane can be described by a convolution of the ideal image with a point spread function (PSF). The PSF thus describes the response of the imaging system to a single object point. The geometric modelling in Eq. 3.2 suggests that the PSF should be a circular filter with a diameter C depending on the depth d. This type of PSF is often called a pillbox blur in the computer vision literature. The shape of the PSF is additionally influenced by diffraction at the lens aperture.

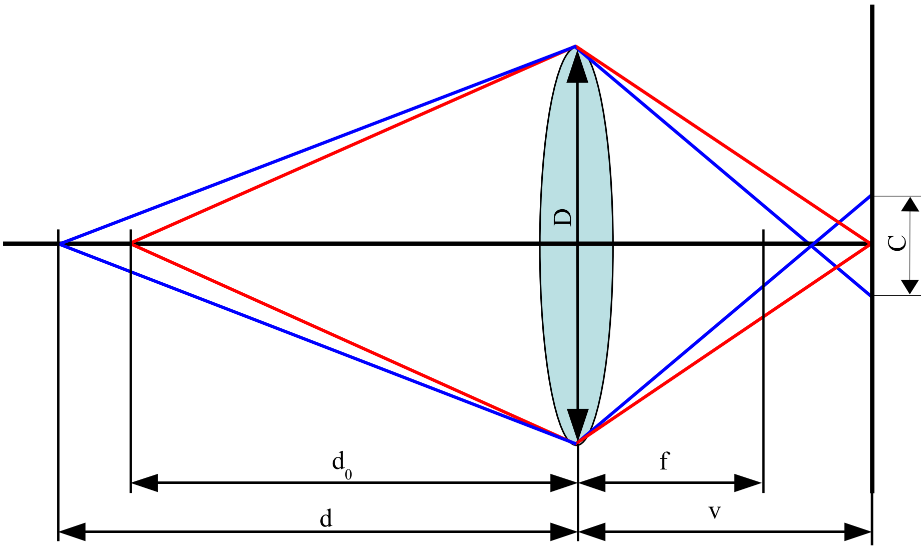


Figure 3.1: Real aperture lens model used for Depth from Defocus.

The shape of the diffraction related contribution to the PSF is described by a wavelength dependent Bessel function of the first kind, see (Born and Wolf 1965) for more details. The "‘rings"’ produced by the diffraction vary in amplitude, width, and position with different states of focus and wavelengths. As wave length varies, the positions of the rings change by as much as 90°, leading to a superposition of minima and maxima of different wavelength. This applies especially if white light with a continuous spectrum from blue to red is assumed. If additional imaging system artefacts like chromatic abberation and digital sampling are considered, the PSF of a real aperture lens is almost certainly best described by a two-dimensional Gaussian (Pentland 1987)

$$h(\sigma) = \frac{1}{2\pi\sigma^2} exp\left(\frac{-(x^2+y^2)}{2\sigma^2}\right). \tag{3.3}$$

The value of σ is proportional to the radius of the circle of confusion: $\sigma = \rho C$. The constant ρ depends on the particularities of the optics and imaging sensor used and is usually determined in a calibration procedure (Pentland 1987). The defocussed image G can be expressed by a space variant convolution with the depth dependent PSF $h(\sigma(d(u,v)))$ of the ideally focused image image F:

$$G(u,v) = h(\sigma(d(u,v))) * F(u,v) \tag{3.4}$$

3.1 Depth from Focus

When the image is in focus, knowledge of the camera parameters f and v can be used to calculate the depth d of the object. In *Depth from Focus* (DfF), a sequence of images of a scene is obtained by continuously varying the distance v between the lens and the image detector (Subbarao and Choi 1995). This leads to a series of differently blurred images. For each image, a sharpness measure is computed at each pixel in a local window. Since the PSF is a depth dependent low pass filter, most focus measures proposed in the literature are based on the high frequency content of an image. For each pixel position the image with the maximum focus measure is determined.

Solving the lens law for d leads to

$$d = \frac{f\, v}{f - v}. \tag{3.5}$$

Using the known camera parameters f and v, the depth d of a well focussed object point can be calculated.

The main difference between the different Depth from Focus methods proposed in the literature is the choice of the focus measures (Krotkov 1988), common measures are based on the strength of high frequency components in the amplitude spectrum. A particularly simple way is to use the image intensity variance of a local region. With a suitable criterion, the maximum of the sharpness measure can be interpolated, resulting in an improved depth resolution (Nayar and Nakagawa 1994).

Depth from Focus is a comparably simple method, only one camera position is involved and the computational cost for depth recovery is quite low. Additionally, there is no correspondence problem and the accuracy of the method is relatively high. As the Depth from Focus method relies on high frequency image content it can only estimate the depth for surfaces with image texture. Like all methods based on real aperture effects, it is only applicable to close range scenarios, where the depth of field is small, compared to the object depth range. Sources of measurement errors include edge bleeding and the assumption of a constant depth of each window. A fundamental drawback is the requirement of a whole image focus series, a non-interpolating approach requires one image for each desired distance D.

3.2 Depth from Defocus

The main drawback of Depth from Focus is the necessity of an image series captured with multiple camera focus settings that scans the whole depth measurement range. Depth from Focus uses the camera parameters of the sharpest frame to determine the object depth. However, according to Eq. (3.2), the radius of the circle of confusion is a function of the camera parameters and the depth of a scene point. Hence, the amount of blur observable in a defocussed image contains information about the depth of a scene point. Using the assumption of a Gaussian PSF (cf. Eq. (3.3)) and $\sigma = \rho C$, the parameter σ is used as a defocus measure.

Pentland (1987) shows that a-priori information about the image intensity distribution, e.g. the presence of sharp discontinuities (edges), allows the computation of the blur parameter σ based on a single defocussed image G. If no scene information is available, the ideally sharp image F in Eq. (3.4) can be an approximated by an image acquired with a very small (pin-hole) aperture. Measuring σ is the most important part of the depth estimation. The classical *Depth from Defocus* (DfD) approach by Subbarao (1988) uses two possibly defocused images G_1 and G_2 of the same scene taken at two different focal settings. Let $\mathfrak{G}_1(\omega, \nu)$ and $\mathfrak{G}_2(\omega, \nu)$ be the amplitude spectra of G_1 and G_2. By dividing $\mathfrak{G}_1$ by $\mathfrak{G}_2$ (cf. Eq. (3.4)), the unknown ideally focused image F can be eliminated:

$$\frac{\mathfrak{G}_1(\omega, \nu)}{\mathfrak{G}_2(\omega, \nu)} = \exp\left[-\frac{1}{2}(\omega^2 + \nu^2)(\sigma_1^2 - \sigma_2^2)\right] \tag{3.6}$$

Solving the above equation for $(\sigma_1^2 - \sigma_2^2)$, and substituting the result into Eq. 3.2 and $\sigma = \rho C$ yields a quadratic equation for the depth d. By applying this technique on a local window, depth values can be computed for each pixel of an image. In general the recovery of depth from defocussed images is a space-variant blur identification problem.

DfD can recover the depth using at least two defocussed images. DfD is more sensitive to inaccurate camera and blur models than Depth from Focus. Some commonly used lenses show non-gaussian PSFs, sometimes depending whether the image is focused in front or behind the image plane. Like Depth from Focus, textured or high frequency image content is required. Windowing for local analysis assumes a constant depth of all scene points captured in the window, tilted surfaces or depth discontinuities will result in increased measurement errors. Additionally, there is interaction between neighbouring windows, due to the non-local PSF scene points “spread” into neighbouring windows. A general property of the DfD approach is that the computed depth map is dense, i.e. for each pixel a depth value is determined, but the resulting depth values tend to display a considerable scatter. Chaudhuri et al. (1999) introduce algorithms that consider the window effects and provide an in-depth review of DfD and related real-aperture methods.

4 Photometric methods

In contrast to geometric methods, which are mainly based on the geometric aspect of image formation, photometric methods use the image irradiance (amount of light power incident onto a surface area on the image plane) as their main source of information. This requires modelling of the photometric aspect of image formation.

Fig. 4.1 illustrates the geometry of the considered image formation model for a single light source. A surface area defined by the surface normal $\vec{n}$ is illuminated by a point light source and viewed by a camera in the directions $\vec{s}$ and $\vec{v}$. The incidence angle θ_i is defined as the angle between surface normal $\vec{n}$ and illumination direction $\vec{s}$, the emission angle θ_e as the angle between surface normal $\vec{n}$ and viewing direction $\vec{v}$, and the phase angle α as the angle between illumination direction $\vec{s}$ and viewing direction $\vec{v}$. Then the image irradiance I at a point (u, v) in the captured image is defined by the reflectance function

$$I(u, v) = R_I\left(\vec{n}, \vec{s}, \vec{v}\right). \tag{4.1}$$

The reflectance function R_I may depend on further, e. g. material-specific, parameters which possibly in turn depend on the pixel coordinates (u, v), such as the surface albedo $\rho(u, v)$. The surface albedo specifies the intrinsic brightness of the surface, and often includes the radiant power incident onto the surface patch imaged at pixel (u, v).

Often, a physically plausible reflectance function, such as the reflectance of a Lambertian surface is used. The Lambertian model describes diffuse reflectance and only depends on the incidence angle θ_i. The reflectance of a Lambertian surface is thus independent from the viewing direction $\vec{v}$ and is given by

$$R_I(\theta_i) = \rho cos(\theta_i). \tag{4.2}$$

If $\vec{n}$ and $\vec{s}$ are unit vectors, the Lambertian model is given by $R_I(\vec{n}, \vec{s}) = \rho\vec{n} \cdot \vec{s}$. Due to its simple mathematical form, it is commonly used in the computer vision literature, however it is limited to a subset of smooth, diffuse surfaces (Zhang et al. 1999). The reflectance of possibly rough specular surfaces can be modelled with the Torrance and

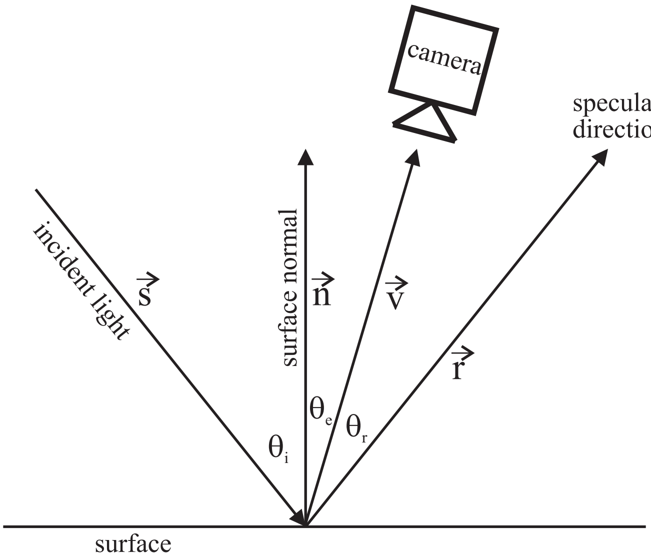


Figure 4.1: Image formation model used by most photometric surface reconstruction methods. In the general case, the vectors $\vec{s}$, $\vec{n}$ and $\vec{v}$ are not coplanar.

Sparrow reflectance function (Torrance and Sparrow 1967), a reflectance function based on geometrical modelling of surfaces consisting of randomly oriented, specular reflecting microfacets. Due to its higher mathematical complexity, only a few computer vision researchers use the Torrance and Sparrow or even more complex models, like the model proposed by Beckmann and Spizzichino (1963) which is based on electromagnetic wave theory.

The reflectance function of a given surface can be determined by measuring the reflectance under the given illumination conditions for various surface orientations, either by capturing many images of a flat part with different light source orientations, or by using a calibration object, for which the surface orientation at each pixel can be easily calculated, for example a sphere. The measured reflectance values can be used to determine the parameters of a parametric reflectance function, like the Torrance and Sparrow reflectance, or used directly.

Most approaches based on photometric properties consider scenes with a single object and uniform reflectance behaviour. Objects with complicated or spatially varying reflectance behaviour usually require additional information about the object shape, or multiple images with different imaging geometry. For complex scenes with multiple objects, different reflectance functions and large depth discontinuities (e.g. cluttered room or city street scenes), 3D reconstruction based on photometric methods alone is not feasible with currently known approaches.

4.1 Shape from Shading

The goal of *Shape from Shading* (SfS) is to reconstruct the three-dimensional surface shape using one intensity image of a scene. The first approaches to recover shape using the shading of surfaces were done in both the machine vision and the astrogeology communities. In the latter the method is called photoclinometry. Historically, photoclinometry has focused on the recovery of one-dimensional surface profiles, whereas Shape from Shading is concerned with the recovery of the surface shape of a whole area, a problem that is also known as area-based photoclinometry.

Most Shape from Shading algorithms assume that the scene is illuminated by one light source and do not consider multiple reflections on the surface. The general Shape from Shading problem is ill-posed (Prados and Faugeras 2005). For Shape from Shading with Lambertian surfaces and unknown light source direction, Belhumeur et al. (1999) show that the surface can only be recovered up to an unknown bas-relief ambiguity. Therefore the image formation model is often simplified using the following assumptions:

- Lambertian surfaces, often with uniform surface albedo,
- known light source and camera directions $\vec{s}$ and $\vec{v}$
- parallel incident light and orthographic projection, leading to constant light source and camera direction over the whole surface.

For the remainder of this section, these assumptions are used unless noted otherwise. While these assumption remove the generalized bas-relief ambiguity, they do not lead to a well posed reconstruction problem (Prados and Faugeras 2005).

The surface shape is described by a depth map $z(u, v)$, that stores the depth at each pixel position. In many Shape from Shading schemes, surface gradients $p = \frac{\partial z}{\partial u}$ and $q = \frac{\partial z}{\partial v}$ are used to express the surface normal. Assuming $\vec{s}$ and $\vec{v}$ are known and constant over the whole surface, Eq. (4.1) only depends on the surface gradient p and q:

$$I(u, v) = R_I\left(p(u, v), q(u, v)\right) \tag{4.3}$$

Methods to solve the SfS problem can be categorised into three groups: propagation approaches, minimisation approaches and local approaches. A survey by Zhang et al. (1999) provides a good overview of the methods developed by the computer vision community and includes an evaluation of algorithms on synthetic and real data.

4.1.1 Propagation approaches

Horn (1975) introduced a propagation algorithm to reconstruct profiles without assumptions about the local surface orientation, based on so-called characteristic strips. A characteristic strip is a line or curve in the image which propagates gradient information from a singular point with known surface orientation and depth. Multiple characteristic strips can be evolved in parallel to reconstruct a dense depth map. A singular point is a point where the surface gradient can be uniquely determined from the image irradiance. For example, if the light source is located at the viewer ($\vec{s}=\vec{v}$) many reflectance functions R_I have a unique global maximum at $\vec{n} = \vec{v}$. Non-Lambertian reflectance functions can be used, as well as perspective projection.

Solving Eq. 4.3 leads to a first order partial differential equation (PDE) (Prados 2004). In general, no unique solution exists for these problems. Existing numerical solvers require boundary conditions, typically the height of the solution on the image borders (so called Dirichlet boundary conditions). A particularly simple form is obtained for the SfS problem where light source and viewer direction coincide. This leads to the Eikonal equation (Rouy and Tourin 1992)

$$|\nabla z(u,v)| = \sqrt{\frac{1}{I(u,v)^2} - 1}. \tag{4.4}$$

Algorithms based on solving the PDE associated with the different formulations for the SfS problem have been proposed by Rouy and Tourin (1992) for the eikonal equation and Oliensis and Dupuis (1994) for arbitrary light source positions. While these algorithms are computationally more efficient than algorithms based on minimisation, their main limitation for practical applications is the requirement of boundary conditions, especially the height of the surface at the image boundaries. Such information is not available for many reconstruction problems of practical relevance. However, Prados and Faugeras (2005) have published a scheme that can work without boundary conditions for a special case where the light source is located at the viewer and the $1/r^2$ attenuation of the light intensity is considered.

4.1.2 Minimisation approaches

Another approach to surface reconstruction is based on minimising an error function. Since the measured image irradiance $I(u,v)$ of a real image will always be subject to measurement noise, the surface reconstruction method should be stable with respect to this noise. The surface gradients $p(u,v)$ and $q(u,v)$ can be computed by minimising the squared image brightness error

$$e_I = \sum_{u,v} \left[I(u,v) - R_I\left(\rho(u,v), p(u,v), q(u,v)\right)\right]^2 . \tag{4.5}$$

In many scenarios, the surface albedo ρ is assumed to be constant across the whole object. As the pixel intensity information alone is not necessarily sufficient to provide an unambiguous solution for the surface gradients $p(u,v)$ and $q(u,v)$, additional information is required. Horn (1989b) propose a "departure from smoothness" term, which requires smoothness of the surface, i. e. for example small absolute values of the directional derivatives of the surface gradients

$$e_s = \sum_{u,v} \left[\left(\frac{\partial p}{\partial x}\right)^2 + \left(\frac{\partial p}{\partial y}\right)^2 + \left(\frac{\partial q}{\partial x}\right)^2 + \left(\frac{\partial q}{\partial y}\right)^2 \right] . \tag{4.6}$$

The surface is reconstructed by looking for a minimum of

$$e = e_s + \lambda e_I \tag{4.7}$$

with respect to the surface gradients p and q. In this function λ is not a Lagrange multiplier, but a factor that balances the relative contributions of the brightness and the departure from smoothness error terms. This factor has to be chosen and is not estimated during the minimisation (Horn 1989b). Eq. (4.7) is then minimised by setting the derivatives of the error term e with respect to p and q to zero, leading to an iterative update rule for the estimated surface gradients $\tilde{p}$ and $\tilde{q}$:

$$\tilde{p}_{n+1} = \bar{p}_n + \lambda' \left(I - R_I(\bar{p}_n, \bar{q}_n)\right) \frac{\partial R_I}{\partial p} , \tag{4.8}$$

where n denotes the iteration index and $\bar{p}(u,v)$ is the local average over the four nearest neighbours of $\tilde{p}(u,v)$. A corresponding expression for $\tilde{q}$ is obtained in an analogous manner. This derivation is described in more detail by Horn (1989b) and Jiang and Bunke (1996). Since a large number of solutions for $\tilde{p}$ and $\tilde{q}$ exists, the initial values $p_0(u,v)$ and $q_0(u,v)$ should be provided based on a-priori knowledge about the surface.

The 3D surface profile $z(u,v)$ can be obtained by computing a surface of maximum consistency with the generally non-integrable gradient field $(\tilde{p}(u,v), \tilde{q}(u,v))$ through minimisation of

$$e_{\text{int}} = \sum_{u,v} \left[\left(\frac{\partial z}{\partial x} - \tilde{p} \right)^2 + \left(\frac{\partial z}{\partial y} - \tilde{q} \right)^2 \right]. \tag{4.9}$$

with respect to z. This leads to the Poisson equation $\Delta z = p_x + q_y$, which can be solved for z using variational (Horn 1989b) or direct analytical methods (Simchony et al. 1990). The latter method has a significantly lower computational complexity.

The departure from smoothness error leads to over-smoothing of wrinkled surfaces, and can be replaced by a departure from integrability error term Eq. (4.9) as discussed in detail by Horn (1989b). This scheme has the additional advantage that the surface height z is estimated together with p and q. While these are favourable properties, it needs to be initialised closely to the solution and is computationally expensive. To increase the convergence radius, Horn (1989b) combines both smoothness and integrability constraint and reduce the smoothness weighting factor during the surface reconstruction.

The minimisation approach described above is very flexible, since it allows arbitrary (but known) placement of the light source and is not restricted to Lambertian surfaces. The drawback is that the λ parameter, and a suitable stopping criteria for the iterative update (Eq. (4.8)) needs to be chosen, and usually a good initialisation and strong boundary conditions are required.

4.1.3 Local approaches

The local surface orientation then has two degrees of freedom, p and q, whereas Eq. (4.3) provides only one constraint. In traditional profile-based photoclinometry strong assumptions on the surface geometry are made. For example, when calculating a profile along a mountain ridge, the cross track slope can be assumed to be zero, providing an additional constraint on p and q (Squyres 1981). Depending on the specific reflectance function R_I, which can be non-Lambertian, one or more solutions for p and q can be found at each pixel on the profile, in case of multiple solutions, one solution has to be chosen. Once p and q are known, the height can be calculated by integration along the profile.

In the computer vision community, Pentland (1984) and Lee and Rosenfeld (1985) have used strong assumptions about the local surface shape to compute the surface orientation locally at each pixel. These algorithms assume a spherical surface shape at every point to be able to compute the surface orientation at each pixel locally, using the first two partial derivatives of the brightness. These algorithms are sensitive to noise and limited to approximately spherical surfaces.

4.2 Photometric stereo

Shape from Shading can be extended to use multiple images, each acquired with a different light source direction, this extension is known as *Photometric stereo*. Each image provides a brightness constraint (Eq. (4.1)), leading to a system of equations. This reduces or completely eliminates the ambiguities encountered in Shape from Shading. The initial work by Woodham (1980) dealt with Lambertian scenes, known light source directions and orthographic projection. If the surface orientation is given by a unit surface normal $\vec{n}$, the brightness constraint for each image i is given by:

$$I_i = \rho \frac{\vec{s} \cdot \vec{n}}{\|\vec{n}\|}. \tag{4.10}$$

For two light source directions and known albedo, two solutions for the surface normal can be found at each pixel, without boundary conditions or global constraints such as smoothness or integrability. For three or more light source directions, both surface normal and albedo can be estimated separately for each pixel.

Woodham (1994) has introduced a method based on three images that does not require Lambertian surfaces. In this approach, the scene is illuminated by one red, one green and one blue light source simultaneously and imaged with a colour camera, allowing to capture the three light source images in a single RGB image. His implementation is using table lookups and can operate in real time. This is a practical method, but it only works for grey objects, and the lamp colours have to chosen carefully based on the spectral response of the camera.

More recently, a method that can deal with materials with anisotropic reflectance functions, such as velvet, have been proposed by Hertzmann and Seitz (2005). Their method required a calibration object and does not make any assumptions on the reflectance of the material. At each image position, the intensity or colour measurements acquired using multiple light sources are stacked into an observation vector. On the calibration object, each observation vector is associated with a surface orientation. The surface orientation of points on the target object can be determined by searching for the best matching observation vector on the calibration object. The approach has been extended to deal with multiple materials, using a linear combination of two calibration objects. The number of images depends on the complexity of the reflectance function, Hertzmann and Seitz (2005) have used 8 to 12 images, depending on the complexity of the reflectance function. Despite being applicable to a very large class of surfaces, their drawback is the requirement for a calibration object and a large number of light sources to avoid ambiguities in the observation vectors. Also, the nearest neighbour search is computationally complex, even if fast, approximate methods are used.

Instead of locally estimating the surface properties at each pixel, a global minimisation approach similar to the one discussed in the previous section can be used to reconstruct a surface subject to some constraints. Given a larger amount of images, the reconstruction of surfaces with non-Lambertian and spatially varying reflectance functions and unknown light source directions becomes feasible (Goldman et al. 2005), but is computationally complex.

4.3 Shape from Polarisation

Light is a transverse electromagnetic wave, which is described by a polarisation state of the electric field. A thorough description of the polarisation of electromagnetic waves is given by Wolf and Born (1959). An electromagnetic wave consists of electric and magnetic fields whose vectors are at right angle to the direction of propagation and oscillate in phase. The electric and magnetic field vectors are orthogonal, for the description of polarisation it is sufficient to describe the electromagnetic wave using its electrical component. The variation along the direction of propagation varies sinusoidally and depends on the wavelength and energy of the wave. The polarisation describes the orientation of the electric vector perpendicular to the direction of propagation over time.

In general, superposition of multiple waves can be described by a linear combination of two perpendicular electric field vectors E_x and E_y. Linear polarisation will occur when E_x and E_y are in phase, if they are out of phase, elliptical or circular polarisation will occur, the resulting electric field vector $E(t) = E_x(t) + E_y(t)$ will rotate over time. Until now, we have assumed that the phase difference between two field vectors is constant, leading to a deterministic behaviour of the resulting field vector E. In many cases, the phase of the E_x and E_y is not deterministically synchronised and varies randomly over time, resulting in unpolarised light. Such light is produced by many common light sources such as light bulbs or light emitting diodes. A combination of unpolarised light and polarised light results in partially polarised light.

Most imaging sensors measure the energy of the incoming light and not its polarisation state. When observed through a rotating linear polarisation filter, the measured irradiance oscillates sinusoidally as a function of the orientation of the polarisation filter between a maximum $I_{\max}$ and a minimum $I_{\min}$. The polarisation angle $\Phi \in [0°, 180°]$ denotes the orientation under which maximum transmitted radiance $I_{\max}$ is observed. The polarisation degree is defined by $D = (I_{\max} - I_{\min})/(I_{\max} + I_{\min}) \in [0, 1]$.

The use of polarisation in the computer vision literature is relatively small, and except for the early work by Koshikawa (1979) only the linear polarisation state has been used in the computer vision community, since it is sufficient for most applications where unpolarised light is reflected by a surface (Wolff and Boult 1991).

Wolff (1991) assumes that surfaces are lit with unpolarised light, and uses a Fresnel reflectance model which predicts the polarisation state of light reflected from dielectric and metallic surfaces. This model is based on the Fresnel reflection coefficients $F_{\perp}$, which specifies the attenuation of the light polarised perpendicular to the plane of reflection, and $F_{||}$, which defines the attenuation of light polarised parallel to the plane of reflection. If these coefficients differ from each other, reflection of unpolarised light results in polarised reflected light. The values of the coefficients depend on the type of material (dielectric or conductive), its index of refraction, the type of reflection (diffuse or specular) and the angle of incidence or emittance (Wolff 1991). For specular reflection of smooth, dielectric surfaces, $F_{\perp} >= F_{||}$, resulting in linear polarisation perpendicular to the plane of reflection. The degree of polarisation depends on the incidence angle θ_i, and reaches 1 at the Brewster angle, $arctan(n)$, where n is the index of refraction, and approaches 0 at $\theta_i = 0$ and $\theta_i = 90°$. Diffuse, or body reflection results is largely unpolarised light (Wolff and Boult 1991), except for large large viewing angles θ_e. For this case, the polarisation direction is parallel to the emittance plane defined by the surface normal vector $\vec{n}$ and the viewing vector $\vec{v}$.

The dependencies of the Fresnel coefficients on the material type can be used for material classification by computing the ratio of the Fresnel coefficients of specular reflections. Wolff (1991) proposed further applications of the Fresnel reflectance model including classifications of edges according to their origin (occluding boundary, specularities, albedo and shadow edges), separation of diffuse and specular reflectance components, and the estimation of surface normals using specular reflections on dielectric surfaces.

Since the polarisation state of reflected light is a function of the orientation of the surface, polarisation measurements can be used for estimation of surface orientation. For specular reflection from smooth dielectric surfaces, the polarisation angle ϑ defines a plane in which the surface normal is located. Together with the specular angle of incidence, the surface normal can be determined. Estimation of the specular angle however requires knowledge index of refraction and is subject to a two way ambiguity, except at the Brewster angle, where the degree of polarisation reaches 1 (Wolff 1991). In other work by Miyazaki et al. (2003) a spherical surface normal distribution is assumed and the 3D surface shape and intensity reflectance properties are estimated from a single image lit with multiple light sources. Miyazaki et al. (2004) propose a related method for reconstruction of transparent objects, where the object is illuminated with light from all directions, producing specular reflection over the whole surface. The two way ambiguity is resolved by using a second image with slightly rotated object. A Fresnel reflectance model with complex index of refraction has been used by Morel et al. (2005) for the reconstruction of very smooth, mirror like metallic surfaces. In a later publication (Morel et al. 2006), the specular angle ambiguity is resolved by varying the illumination.

Note that all surface estimation algorithms above are limited to reconstruction of smooth, dielectric and metallic surfaces without interreflections. Since the polarisation of the diffuse reflectance component of dielectrics is very low, most approaches assume specular reflections, and thus require a uniform, spherical illumination. Polarisation based 3D surface reconstruction of rough dielectric and metallic surfaces has not been considered in the literature.

5 Combined approaches

The reconstruction algorithms introduced in the previous sections are based on properties of the object such as texture or shading. Given objects with sufficient texture, Stereo and DfD can be used to compute a dense surface map, with depth measurements available for each pixel. Algorithms based on Shape from Shading also produce a dense surface description, but work best on objects without texture, whereas approaches based on polarisation upon reflection require smooth surfaces. Since the surface properties of many objects vary, a single method will only yield useful results on areas where features required by the particular method are available. Reconstruction methods based on multiple, complementary features, such as Stereo on textured surface areas and shading in textureless areas, allow reconstruction of dense surface shape in both areas.

5.1 Shape from Shading and geometric approaches

A combined reconstruction based on geometric and photometric reconstruction methods is desirable, since both approaches complement each other. For example, Stereo and Shape from Motion require surface texture to establish correspondences, whereas most Shape from Shading perform best on textureless surfaces. A number of approaches to combine Stereo and Shape from Shading have been proposed in the literature. Cryer et al. (1995) fuse low-pass filtered stereo depth data and high-pass filtered SfS depth data. Their approach, however, requires dense depth data and fuses the independently obtained results of two separate algorithms. Horovitz and Kiryati (2004) propose a method that enforces sparse depth points during the surface gradient integration step performed in many SfS algorithms, involving a heuristically chosen parametrised weight function governing the local influence of a depth point on the reconstructed surface. They propose a second approach, suggesting a subtraction of the large-scale deviation between the depth results independently obtained by Stereo and SfS, respectively, from the SfS solution. For sparse Stereo data, the large-scale deviation is obtained by fitting a sufficiently smooth parametrised surface model to the depth difference values. Both approaches fuse independently obtained results of two separate algorithms. While the fusion of Stereo and SfS results requires no changes to the Stereo and SfS algorithms, it does not provide a self consistent fusion of the information obtained from both shading and polarisation properties, since the stereo and shading constraints are not considered during the final combination step.

Samaras et al. (2000) introduce a surface reconstruction algorithm that performs stereo analysis of a scene and uses a minimum description length metric to selectively apply SfS to regions with weak texture. A surface model described by finite elements is adjusted to minimise a combined depth, multi-image SfS, and smoothness error. The influence of a depth point on the surface, however, is restricted to a small local neighbourhood of the corresponding finite element, favouring the use of dense stereo depth data. A related approach by Fassold et al. (2004) integrates stereo depth measurements into a variational SfS algorithm and estimates surface shape, light source direction, and diffuse reflectance map. In their approach, the influence of a depth point is restricted to a small local neighbourhood of the corresponding image pixel.

Another approach by Lim et al. (2005) is based on photometric stereo, but moves the camera instead of the light source. For Lambertian surfaces, the reflectivity of the surface normal is independent of the viewing direction, cf. Eq. (4.10), the only difference is that due to the different viewpoints, the correspondence problem needs to be solved as well. They use Shape from Motion to reconstruct a small number of 3D points and use interpolation to arrive at a coarse dense depth map. This approximate depth map is used to “solve” the correspondence problem and perform Photometric stereo on assumed corresponding pixels. The correspondence and Photometric stereo steps are repeated until convergence.

5.2 Shape from Polarisation and geometric approaches

As described in Section 4.3, the polarisation angle measured for specular reflection on dielectic surfaces provides one constraint on the surface normal. This ambiguity can be avoided by using a stereo setup (Wolff 1991), but the correspondence problem has to be solved as well. Rahmann and Canterakis (2001) propose a global surface reconstruction scheme, which estimates a parametric surface description using three or more polarisation angle images of a specular object captured from different viewpoints. Their approach does not require surface texture. Atkinson and Hancock (2005) recover surface shape using two polarisation images of an object placed on a turntable, which is rotated by a known angle.

They recover correspondences and their associated surface normals based on polarisation angle and degree, and the turntable rotation. Their approach also considers the polarisation of the diffuse reflectance component.

Part II

Developed algorithms for scene reconstruction

6 System design and overview

Three dimensional scene reconstruction can be done using a variety of methods, cf. Chapter I. Depending on the requirements of the reconstruction task (for example: sparse vs. dense, accuracy) and characteristics of the scene (properties of the object, number of images and camera position, lighting), a reconstruction method needs to be chosen. Often, multiple methods based on different principles (triangulation, defocus, shading etc.) can be used to solve a reconstruction problem, and each method has its strength in different areas. For example, methods based on triangulation, such as Stereo or Structure from Motion, require corresponding points, which can only be established reliably on textured or structured surfaces or by using structured illumination. On the other hand, methods based on photometric properties usually work best on surfaces without surface texture. To archive a dense three dimensional surface reconstruction of an object that contains both textured and non-textured areas, a combined reconstruction scheme that takes into account both geometric and photometric constraints is advantageous.

Multiple approaches for the fusion of different reconstruction methods exist. One approach is the fusion of results obtained by each method, for example by averaging the reconstruction results. During the combination of the results, the fundamental constraints of each method are not considered. For example the Stereo method is based on the epipolar constraint and Shape from Shading on brightness constraint. Since the final reconstruction step does not consider the constraints of the individual methods, its result is not necessarily consistent with the original measurements (for example, disparity, and irradiance). Alternatively, one method can be extended to consider additional constraints, based on the result produced by other methods (for example, a point cloud or depth map). Then at least one constraint related directly to the scene measurement is considered during the final reconstruction phase. A self-consistent framework, where all constraints are considered simultaneously is the preferred method, since it will result in a reconstruction that is consistent with all measurements.

In this thesis, a modular surface reconstruction system is proposed, which consists of the individual methods shown in Fig. 6.1. The geometric methods considered include Depth

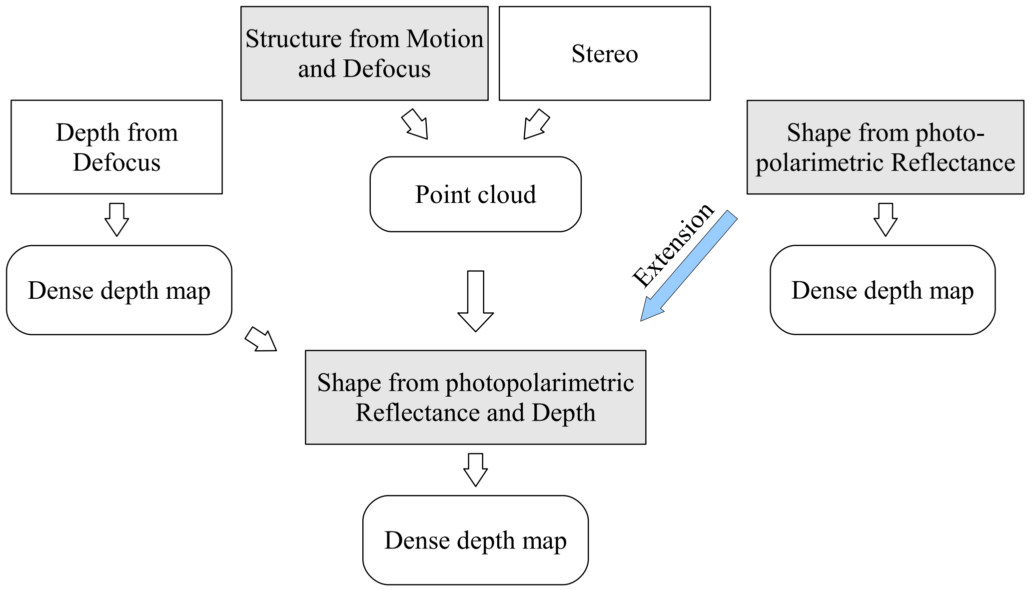


Figure 6.1: System overview. The 3D reconstruction methods (rectangles) considered in this thesis, their output (rounded rectangles), and possible combinations. Novel methods developed in this thesis are indicated by grey rectangles.

from Defocus (DfD), Structure from Motion and Defocus, and Stereo. The considered DfD algorithm is based on two images acquired with different aperture settings. The Structure from Motion and Defocus method is a novel combination of Depth from Defocus and Structure from Motion (SfM) into a self-consistent framework, it is described in detail in Chapter 7. The result of this method is a 3D point cloud, whose density mainly depends on the amount of texture and the surface reflectance properties of the object. Other reconstruction methods such as Stereo or laser triangulation can also be used to create a point cloud.

The photometric and polarimetric reflectance properties of the object surface are used by the Shape from Photopolarimetric Reflectance method to recover a dense depth map of the object surface. This method is based on Shape from Shading (cf. section 4.1), but additionally considers the linear polarisation state of the reflected light. In contrast to existing Shape from Polarisation methods (c.f. Section 4.3), the new method is also applicable to rough, metallic surfaces, whose reconstruction is frequently required in practical quality inspection scenarios. The novel Shape from Photopolarimetric Reflectance and Depth method can be used to integrate depth information, for example a point cloud created by Stereo or Structure from Motion and Defocus, into the photopolarimetric reconstruction process. This is especially useful if large surface areas are reconstructed, since the gradient based photometric reconstruction methods often lead to less accurate measurements due to the required integration step.

7 Structure from Motion and Defocus

Structure from Motion (SfM) recovers the spatial scene structure using corresponding points in at least two images acquired sequentially from different camera positions. Without prior knowledge about scene or camera motion, feature point positions and camera pose are only recovered up to an unknown scale factor (cf. Section 2.3). Depth from Defocus methods rely on the fact that a real lens blurs the observed scene before the imaging device records it. The amount of blurring depends on the actual lens, but also on the distance of the observed object to the lens (c.f. Section 3.2). Contrary to Structure from Motion, Depth from Defocus produces absolute, but noisy depth values.

So far, no attempt has been made to combine the precise relative scene reconstruction of Structure from Motion with the absolute depth data of Depth from Defocus. A related work was published by Myles and da Vitoria Lobo (1998), where a method to recover affine motion and defocus simultaneously is proposed. However, the spatial extent of the scene is not reconstructed by their method, since it requires planar objects.

To facilitate the integration of defocus information into the Structure from Motion framework, the image sequences are acquired such that the object is blurred in the first image of the sequence, becoming increasingly focused in the middle and blurred again in the last images. The focal settings of the camera are adjusted according to the maximal and minimal distance of the object. It may be necessary to fully open the aperture in order to obtain a small depth of field.

Figure 7.1 provides an overview of the Structure from Motion and Defocus method. Salient feature points are extracted and tracked across the sequence. The amount of defocus and the noisy depth estimate of each point is estimated using the novel Depth from Defocus by motion method. The final reconstruction of 3D points and camera poses is performed by minimisation of a combined bundle adjustment and defocus error term, resulting in a metric reconstruction at absolute scale.

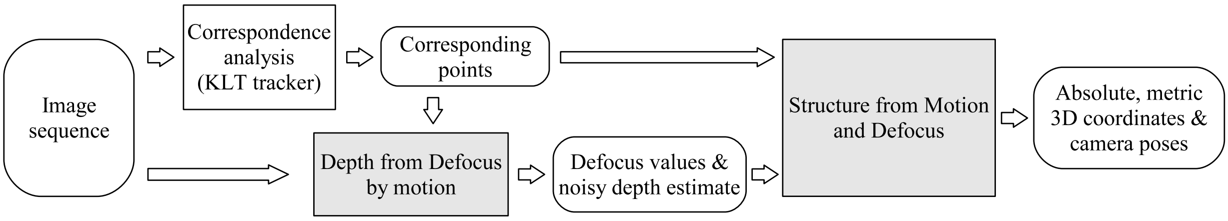


Figure 7.1: Overview of the Structure from Motion and Defocus method. The processing steps (rectangles) and their output (rounded rectangles). Novel methods developed in this thesis are indicated by grey rectangles.

7.1 Depth from Defocus by motion

Depth from Defocus directly recovers the spatial scene structure using a monocular camera. The depth d of the tracked feature points is calculated by measuring the amount of defocus, expressed e.g. by the standard deviation σ of the Gaussian-shaped point spread function (PSF) that blurs the image. Following the argument in Section 3, σ will be referred to as the “radius” of the PSF. As described in Section 3.2, two images of the same object taken at two different focal settings or a-priori information about the image intensity distribution, e.g. the presence of sharp discontinuities (edges), allow the computation of the PSF radius σ. Since images captures with different focal settings or a-priori information about sharp discontinuities in the scene are not available in the considered scenario, while multiple, differently defocused areas are, another approach is required.

We assume that local features in the scene are tracked across a sequence of images and that for each feature the image is determined in which the feature appears best focused. The radius values σ of the Gaussian PSF which transform the best focused version of the feature into the defocussed patterns in all other images are determined. Using a previously calibrated relationship between blur σ and depth d, the depth of the feature is determined for each image frame based on the measured blur. Conceptually, this approach is a fusion of the Depth from Focus and Depth from Defocus methods introduced in Section 3. The principal difference is that the Depth from Defocus part of the new method is based on camera or object motion instead of changes to focal distance or aperture, and utilises the defocus information contained in all images.

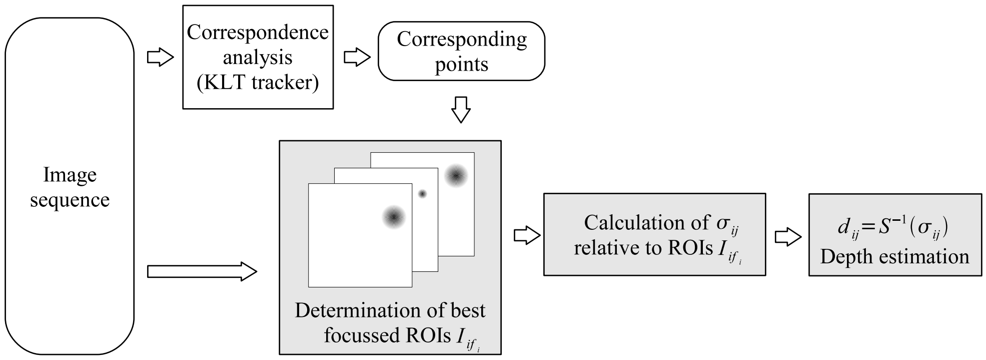


Figure 7.2: Overview of the Depth from Defocus by motion part of the proposed method. The defocus σ_{ij} and the corresponding depth d_{ij} are determined for all features j in all images i

7.1.1 The Depth-Defocus-Function

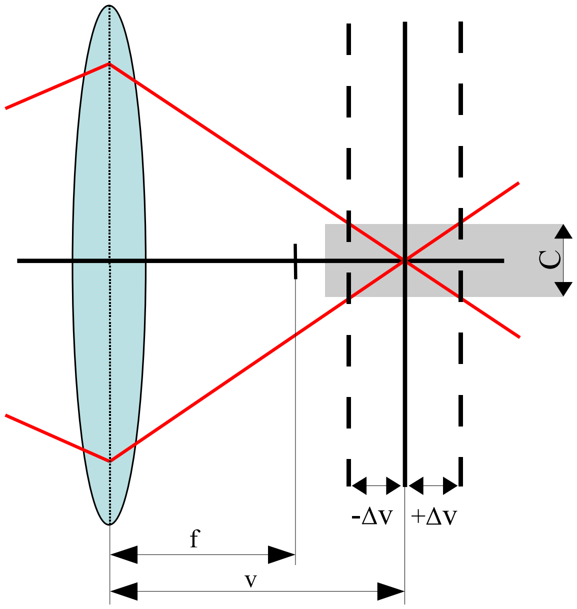
The Depth-Defocus-Function $\mathcal{S}(d) = \sigma$ expresses the radius σ of the Gaussian PSF as a function of depth d, i.e. the distance between the object and the lens plane. It is based upon the lens law (Pedrotti 1993):

$$\frac{1}{v} + \frac{1}{d} = \frac{1}{f}. \tag{7.1}$$

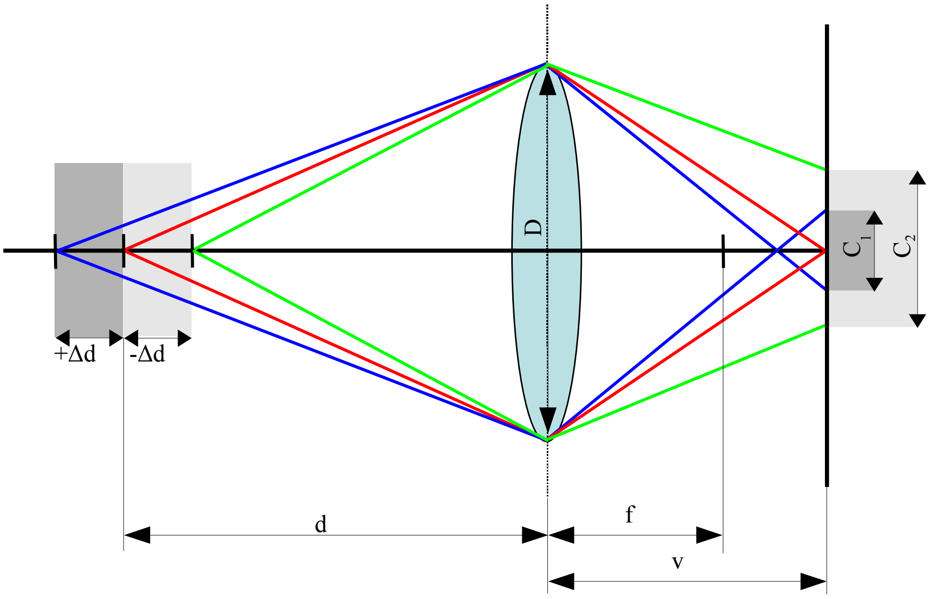
An object at distance d is focused if the distance between lens and image plane is v, with f denoting the focal length of the lens. Varying the image plane distance v by a small amount Δv causes the object to be defocused as the light rays intersect before or behind the image plane. In the geometric optics approximation, a point in the scene is transformed into a so-called circle of confusion of diameter $C = |\Delta v|/\kappa$ in the image plane, where κ is the f-stop number expressing the focal length in terms of the aperture diameter D (Fig. 7.3a). The diameter C of the circle of confusion and the PSF radius σ are related to each other in that σ is a monotonously increasing nonlinear function of C. Most DfD approaches (Chaudhuri et al. 1999) assume a linear relationship between C and σ. This does not hold for the lenses used in our experiments. Empirically, we found that for small shifts of the image plane by $|\Delta v|$, the radius σ of the Gaussian PSF can be modelled well by a zero-mean Gaussian function:

$$F(\Delta v) = \frac{1}{\phi_1} e^{-\frac{1}{\phi_2}\Delta v^2} + \phi_3\ . \tag{7.2}$$

The parameters ϕ_1, ϕ_2 and ϕ_3 are lens and camera specific and need to be determined in a separate calibration step. Hence, the symmetric behaviour of $C(\Delta v)$ apparent from Fig. 7.3



(a) Symmetric dependence of C on Δv.



(b) Asymmetric dependence of C on ΔD.

Figure 7.3: Dependence of the diameter C of the circle of confusion (a) on the offset Δv in image space and (b) on the offset Δd in object space. The offsets Δv and Δd are measured with respect to the distance v between lens and image plane and the distance d between lens and object for the perfectly focused scenario described by the lens law (7.1). The value of C increases more strongly for motion towards the camera than for motion away from the camera.

implies a symmetric behaviour of $\sigma(\Delta v)$. Depending on the constructional properties of lenses different from those we used in our experiments, analytic peak-shaped functions different from Eq. (7.2) but also symmetric in Δv may better match the observed behaviour of the PSF.

However, since the Depth-Defocus-Function expresses the relation between the depth d of an object and its defocus, the image plane is assumed to be fixed while the distance d of the object varies by the amount Δd, such that $\Delta d = 0$ refers to an object that is well focused. But since neither d nor Δd are known, the functional relation needs to be modelled with respect to Δv:

$$\frac{1}{v + \Delta v} + \frac{1}{d} = \frac{1}{f} \, . \tag{7.3}$$

An object point at distance d will be focused if the image plane is placed at v. A value of $\Delta v \neq 0$ refers to a defocused object point. The circle of confusion C observed for an object point at distance d can be calculated by solving Eq. (7.3) for Δv and inserting Δv into $C = |\Delta v|/\kappa$ (Pentland 1987):

$$C = \frac{1}{\kappa} \left| \frac{f\, d}{d - f} - v \right| \tag{7.4}$$

Since the defocus behaviour of our camera is better described by Eq. (7.2), we insert Δv in Eq. (7.2) and arrive at the Depth-Defocus-Function

$$\sigma = \mathcal{S}(d) = \frac{1}{\phi_1} e^{-\frac{1}{\phi_2}\left(\frac{f d}{d-f} - v\right)^2} + \phi_3 \, , \tag{7.5}$$

which describes the radius σ of the Gaussian PSF applied to an object point located at depth d. Calibrating the Depth-Defocus-Function $\mathcal{S}(d)$ for a given lens corresponds to determining the parameters ϕ_1, ϕ_2, and ϕ_3 in Eq. (7.5). This is achieved by obtaining a large set of measured (σ_{ij}, d_{ij}) data points and perform a least mean squares fit to Eq. (7.5), where d_{ij} is the distance of the feature i in image j from the camera and σ_{ij} the radius of the Gaussian PSF G used to blur the well focused image according to

$$I_{ij} = G(\sigma_{ij}) * I_{if_i}. \tag{7.6}$$

Here, I_{if_i} represents a small region of interest (ROI) around feature i in the image f_i in which this feature is best focused, and I_{ij} a ROI of equal size around feature i in image j. By inverting the Depth-Defocus-Function $\mathcal{S}(d)$ according to Eq. (7.5), the depth d

corresponding to σ_{ij} can be determined up to a two-fold ambiguity (cf. Fig. 7.4)

At this point it is useful to to examine under which conditions Depth from Defocus can obtain a depth measurements with a good accuracy. For this we assume that the object is described by a bounding cube with a lateral extension b. In order to archive an accurate reconstruction we choose the focal length f and the distance between the lens and the image plane v so that the front of the object placed at distance d fills a large image area and is in focus. All other parameters, such as the height h of the image plane and the f-stop number $\kappa = f/r_0$ of the lens are kept constant. Using the law of similar triangles, the required distance v between the lens and the image plane is given by $v = d\,h/b$. Using the lens law, Eq. (7.1), the focal length required for a focussed image at distance d amounts to

$$f = \frac{d\,h}{h+b}. \tag{7.7}$$

By inserting these constraints into Eq. (3.2), the diameter C of the circle of confusion for object points at the back of the bounding box, located at distance $d+b$ is given by

$$C = \frac{d\,h^2}{\kappa\,(d+b)\;(h+b)}. \tag{7.8}$$

Closer inspection reveals that for increasing depth d, C converges to $h^2/(\kappa(h+b))$, whereas C approaches 0 for increasing object size b. Depth from Defocus is thus limited to the reconstruction of small scale objects, which can be placed at an arbitrary distance to the camera, as long as corresponding focal length given by Eq. (7.7) is used. By solving Eq. (7.8) for the object size b, it is possible to estimate the maximum object size for which Depth from Defocus yields depth values with reasonable accuracy. In order to achieve a measurable defocus, let us assume that we require a circle of confusion of $C = 4$ pixels for object points located at a distance of $d+b$. For example, a typical industrial camera with a resolution of 1032 by 776 pixels, a pixel pitch of 4.65μm and a lens with an aperture value of $\kappa = 1.4$ is suitable for the reconstruction of objects with a size of up to 0.5m.

7.1.2 Calibration of the Depth Defocus function

For calibration of the DfD method, an image sequence is acquired while the camera approaches at uniform speed a calibration rig displaying a checker board. The sharp black-and-white corners of the checker board are robustly and precisely detectable (Krüger et al. 2004) even in defocused images. Small ROIs around each corner allow the estimation of

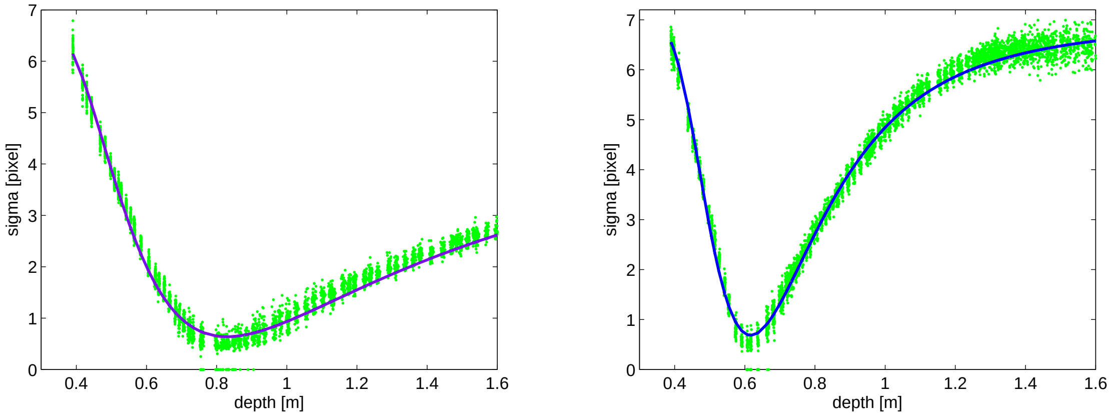


Figure 7.4: Depth-Defocus-Functions of two lenses with $f = 12$ mm (left) and $f = 20$ mm (right), fitted to the measured data points according to Eq. (7.5), respectively.

defocus using their greyvalue variance χ. The better focused the corner, the higher is the variance χ. Instead of choosing the ROI with the highest variance as reference, a fitting procedure considering all ROIs of a corner is applied to introduce robustness with respect to pixel noise. We found experimentally that the parameterised defocus model according to Eq. (7.5) is also a reasonable description of the dependence of χ on the depth d. For our calibration sequence the camera motion is uniform and the image index j is strongly correlated with the object distance d. Hence Eq. (7.5) can also be used to describe the variance χ as a function of j. Eq. (7.5) is fitted to the measured (χ, j) data points for each corner i, such that the location of the maximum of $\mathcal{S}$ yields the index f_i of the image in which the ROI around corner i is best focused. This ROI corresponds to I_{if_i}. The fitting procedure is applied to introduce robustness with respect to pixel noise. For non-uniform camera motion the index f_i can be obtained by a parabolic fit to the values of χ around the maximum or by directly selecting the ROI with maximal χ. The depth d of each corner is reconstructed from the pose of the complete rig according to Bouguet (1997).

For each tracked corner i, we compute for each ROI I_{ij} the amount of defocus, i.e. the σ value relative to the previously determined best focused ROI I_{if_i} according to Eq. (7.6). The magnitude of the gray value variance χ depends on the image content of the tracked ROI and can thus only be used as a relative defocus measure between ROI representing the same image feature. By employing the bisection method, we determine the value of σ for which the root mean square deviation between $G(\sigma) * I_{if_i}$, denoting the best-focused ROI convolved with a Gaussian PSF of radius σ, and I_{ij}, the currently observed ROI, becomes minimal. The Depth-Defocus-Function is then obtained by a least mean squares fit to all determined (σ, d) data points. Two examples are shown in Fig. 7.4 for lenses with focal lengths of 12 mm and 20 mm and f-stop numbers of 1.4 and 2.8, respectively. Objects at a distance of about 0.8 m and 0.6 m, respectively, are in focus, corresponding to the minimum of the curve.

7.2 Integration of Structure from Motion and Defocus algorithms

Structure from Motion analysis involves the extraction of salient features from the image sequence which are tracked using the Kanade Lucas Tomasi (KLT) tracker (Shi and Tomasi 1994). To facilitate the integration of defocus information, a ROI of constant size is extracted around each feature point at each time step. For each tracked feature, the best focused image has to be identified in order to obtain the increase of defocus for the other images. We found that the greyvalue variance as a measure for defocus does not perform well on features other than black-and-white corners. Instead we make use of the amplitude spectrum $|\mathscr{F}_I(\omega)|$ of the ROI extracted around the feature position. High-frequency components of the amplitude spectrum denote sharp details, whereas low-frequency components refer to large-scale features. Hence, the integral over the high-frequency components serves as a measure for the sharpness of a certain tracked feature. However, since the highest-frequency components are considerably affected by pixel noise and defocus has no perceivable effect on the low-frequency components, a frequency band between ω_0 and ω_1 is taken into account according to

$$H = \int_{\omega_0}^{\omega_1} |\mathscr{F}_I(\omega)| d\omega \tag{7.9}$$

with $\omega_0 = \frac{1}{4}\omega_{\mathrm{max}}$ and $\omega_1 = \frac{3}{4}\omega_{\mathrm{max}}$, where ω_{max} is the maximum frequency. The amount of defocus increases with decreasing value of H. The defocus measure H is used to determine the index of the best focused ROI for each tracked feature in the same manner as the greyvalue variance χ in Section 7.1.2. Fig. 7.5 illustrates that the value of H cannot be used for comparing the amount of defocus among different feature points since the maximum value of H depends on the image content. The same is true for the greyvalue variance. Hence, both the integral H of the amplitude spectrum as well as the greyvalue variance are merely used for determining the index of the image in which a certain feature is best focused.

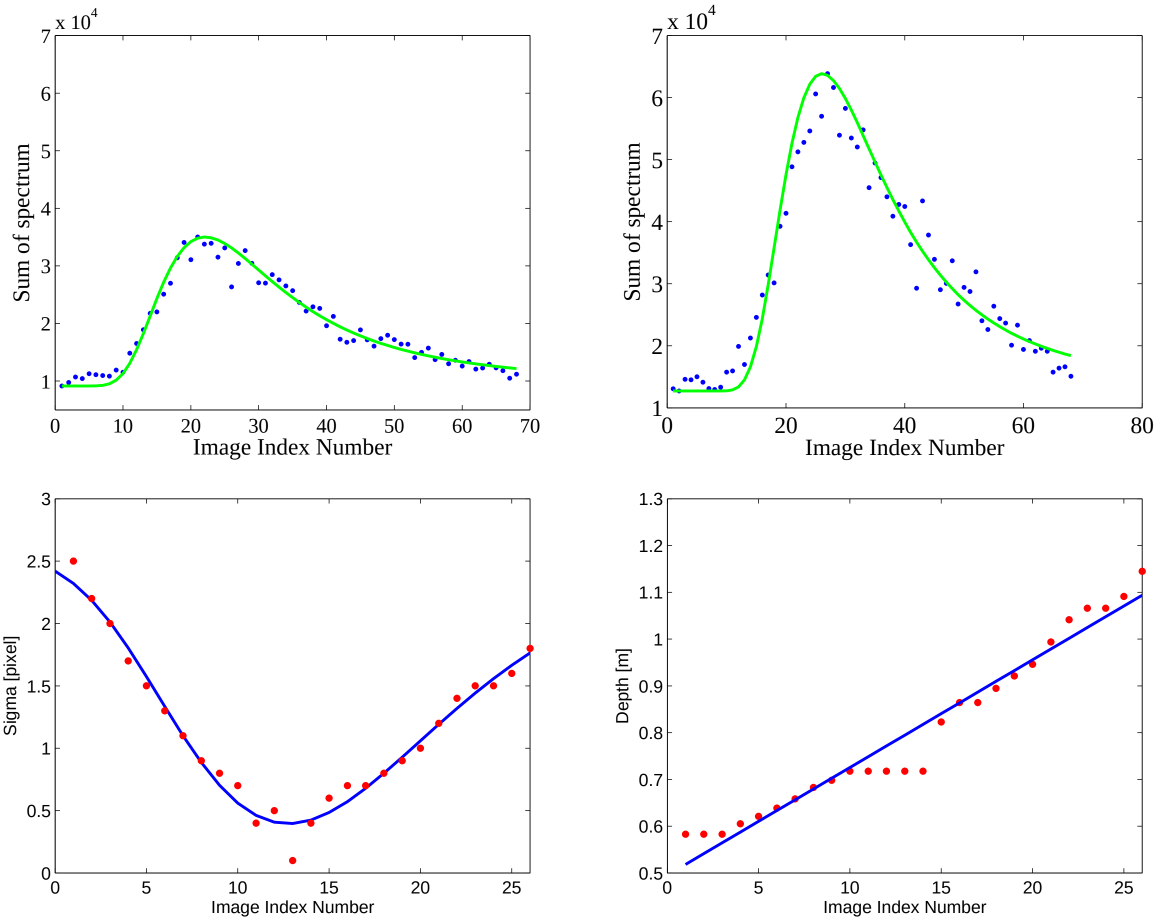


Figure 7.5: Upper left and upper right: Image index vs. defocus measure H for two different tracked image features. Lower left: Image index vs. PSF radius σ. Lower right: Image index vs. inferred depth D.

The defocus, i.e. the radius σ of the Gaussian PSF, is then computed relative to the best focused ROI according to Section 7.1.2. The depth d is obtained by inverting the Depth-Defocus-Function $\mathcal{S}(d)$ according to Eq. (7.5). The encountered two-fold ambiguity is resolved by using information about the direction of camera motion, which is obtained either based on a-priori knowledge or by performing a Structure from Motion analysis according to Eq. (2.4), yielding information about the path of the camera. If the estimated value of σ is smaller than the minimum of $\mathcal{S}(d)$, the depth is set to the value at which $\mathcal{S}(d)$ is minimal. For an example feature, the calculated defocus and the inferred depth values are shown in Fig. 7.5.

We found experimentally that the random scatter of the feature positions extracted by the KLT tracker is largely independent of the image blur for PSF radii smaller than 5 pixels and is always of the order 0.1 pixels. However, more features are detected and less features are lost by the tracker when the tracking procedure is started on a well-focused image. Hence, the tracking procedure is repeated, starting from the "sharpest" image located near the middle of the sequence which displays the largest value of H averaged over all previously detected features, proceeding towards either end of the sequence and using the ROIs extracted from this image as reference patterns. The internal camera parameters, such as focal length and distortion parameters, K_i have been determined by geometrical calibration (Krüger et al. 2004) during the defocus calibration in Sec. 7.1.2. The 3D coordinates X_i of the scene points are then computed by determination of the minimum of the combined reprojection and defocus error term

$$E_{\text{comb}}\left(\{T_j\},\{X_i\}\right)=\sum_{i=1}^{N}\sum_{j=1}^{M}\left[\left(\mathcal{P}\left(T_j,K_j,X_i\right)-x_{ij}\right)^2+\alpha\left(\mathcal{S}\left(\left[T_jX_i\right]_z\right)-\sigma_{ij}\right)^2\right] \tag{7.10}$$

with respect to the M camera transforms T_j and the N scene points X_i. The value of σ_{ij} corresponds to the estimated PSF radius for feature i in image j, α is a weighting factor, $\mathcal{S}$ the Depth-Defocus-Function that calculates the expected defocus of feature i in image j, and $[\cdot]_z$ the z coordinate, i.e. the depth d, of a scene point relative to camera pose T_j. The correspondingly estimated radii σ_{ij} of the Gaussian PSFs define a regularisation term in Eq. (7.10), such that absolutely scaled 3D coordinates X_i of the scene points are obtained. The values of X_i are initialised according to the depth values estimated based on the Depth from Defocus approach. To minimise the error term E_{comb} we employ the

Levenberg-Marquardt algorithm (Levenberg 1944).

7.2.1 Online version of the algorithm

The 3D reconstruction method outlined so far is an offline (or “batch”) algorithm since the error term (7.10) is minimised once for the complete acquired image sequence. In this section we present an online version of the proposed combination of SfM and DfD which processes the acquired images instantaneously, thus generating a new 3D reconstruction result after acquisition of each image of the sequence. This is a desired property e.g. in the context of mobile robot navigation or in-situ exploration.

The online version of our algorithm starts by acquiring the current image. The feature tracker attempts to track the features already present in the previous image and may replace lost features with new ones. Again, the KLT feature tracker (Shi and Tomasi 1994) is used. The sharpness of each feature within the current frame is obtained based on the integral H of the amplitude spectrum of the ROI around the feature position (cf. Eq. (7.9)). The next step is the determination of the best focused frame for each feature. Since fitting the Depth-Defocus-Function (7.5) imposes a considerable computational burden, a second-degree polynomial is fitted to the values of H instead. A threshold rating this fit selects possible candidates that may already have passed their point of maximum sharpness. The Depth-Defocus-Function (7.5) is then fitted to the H values of the pre-selected feature candidates only. After determination of the sharpest frame, the initial depth values for the respective feature are computed by estimating the PSF radius σ as outlined in Section 7.1.2.

The depth values obtained by the DfD method are then used to initialise the Levenberg-Marquardt scheme which determines the camera transforms and 3D feature points that minimise error function (7.10). To reduce the effect of outliers, we employ the M-estimator technique with the “fair” weighting function $w(x) = 1/(1 + |x|/c)$, where $c = 1.4$ is a favourable choice (Rey 1983). For each image, the current optimisation result is used as an initialisation to the subsequent iteration step involving the next acquired image.

8 Shape from Photopolarimetric Reflectance

Methods for surface reconstruction based on reflectance properties of the surfaces, such as Shape from Shading and Shape from Polarisation, have been introduced in Section 4. However, no approach that uses both irradiance and polarisation constraints has been proposed in the literature, even though they are caused by the same physical phenomenon, namely the reflection of light at the object surface. The novel *Shape from Photopolarimetric Reflectance* (SfPR) method developed in this chapter reconstructs a dense depth map of the surface using a generalisation of SfS which additionally considers the linear polarisation state of the light reflected from a surface.

8.1 Basic principles

In our scenario, we will assume that the surface $z(x, y)$ to be reconstructed is illuminated by a point light source and viewed by a camera, both situated at infinite distance in the directions $\vec{s}$ and $\vec{v}$, respectively (cf. Fig. 8.1a). The xy plane is parallel to the image plane. Parallel unpolarised incident light and an orthographic projection model are assumed. For each pixel location (u, v) of the image we intend to derive a depth value $z(u, v)$. The surface normal is given in gradient space by the vector $\vec{n} = (-p, -q, 1)^T$ with $p = \partial z/\partial x$ and $q = \partial z/\partial y$. The incidence angle θ_i is defined as the angle between surface normal $\vec{n}$ and illumination direction $\vec{s}$, the emission angle θ_e as the angle between surface normal $\vec{n}$ and viewing direction $\vec{v}$, and the phase angle α as the angle between illumination direction $\vec{s}$ and viewing direction $\vec{v}$. A measure for the intrinsic reflectivity of the surface is given by the surface albedo $\rho(u, v)$.

In the framework of Shape from Photopolarimetric Reflectance, the light reflected from a surface point located at the world coordinates (x, y, z) with corresponding image coordinates (u, v) is described by the observed pixel intensity $I(u, v)$, the polarisation angle $\Phi(u, v)$ (i. e. the direction in which the light is linearly polarised), and the polarisation degree $D(u, v)$. Measurement of polarisation properties is thus limited to linear polarisation while circular or elliptic polarisation is not taken into account. It is assumed that

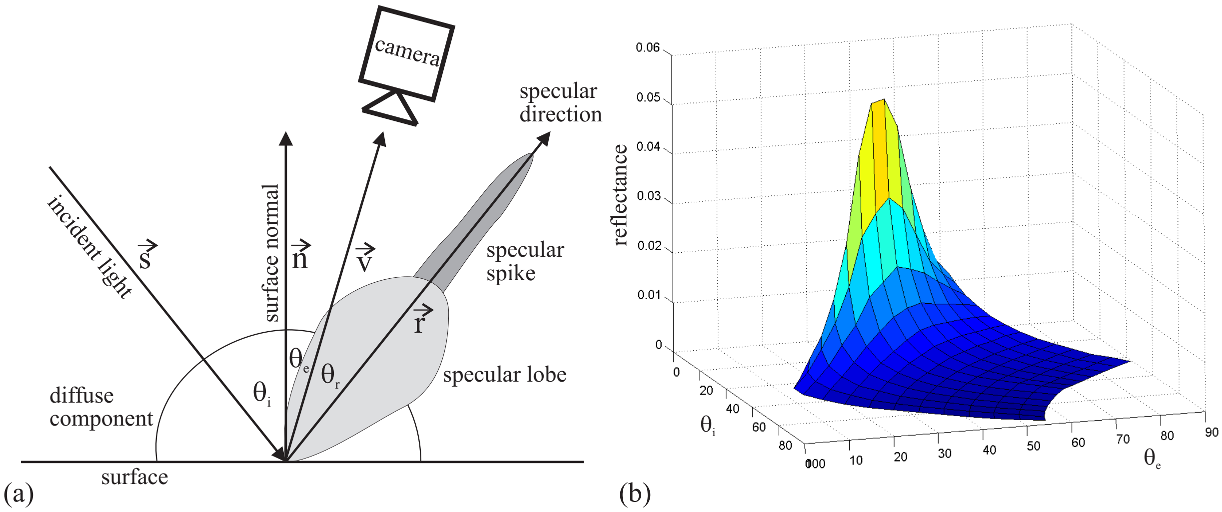


Figure 8.1: (a) Sketch of the three reflectance components. (b) Measured reflectance of a raw forged iron surface for $\alpha = 75°$.

models are available that express these photopolarimetric properties in terms of the surface orientation $\vec{n}$, illumination direction $\vec{s}$, and viewing direction $\vec{v}$. These models may either be physically motivated or empirical (cf. Section 8.2) and are denoted by R_I (intensity reflectance), R_Φ (polarisation angle reflectance), and R_D (polarisation degree reflectance). The aim of surface reconstruction in the presented framework is to determine for each pixel (u, v) the surface gradients $p(u, v)$ and $q(u, v)$, given the illumination direction $\vec{s}$ and the viewing direction $\vec{v}$, such that the modelled photopolarimetric properties of a pixel correspond to the measured values:

$$
\begin{aligned}
I(u,v) &= R_I\left(p(u,v), q(u,v), \vec{s}, \vec{v}\right) && (8.1)\\
\Phi(u,v) &= R_\Phi\left(p(u,v), q(u,v), \vec{s}, \vec{v}\right) && (8.2)\\
D(u,v) &= R_D\left(p(u,v), q(u,v), \vec{s}, \vec{v}\right) && (8.3)
\end{aligned}
$$

The reflectance functions (8.1)–(8.3) may depend on further, e. g. material-specific, parameters which possibly in turn depend on the pixel coordinates (u, v), such as the surface albedo $\rho(u, v)$ which influences the intensity reflectance R_I.

In our scenario, we assume unpolarised incident light, as produced by most common light sources such as light bulbs, light emitting diodes and fluorescent tubes. Most surfaces partially polarise the light upon reflection. The polarisation state of light scattered from smooth dielectic and metallic surfaces can be modeled using the fresnel reflection model, as

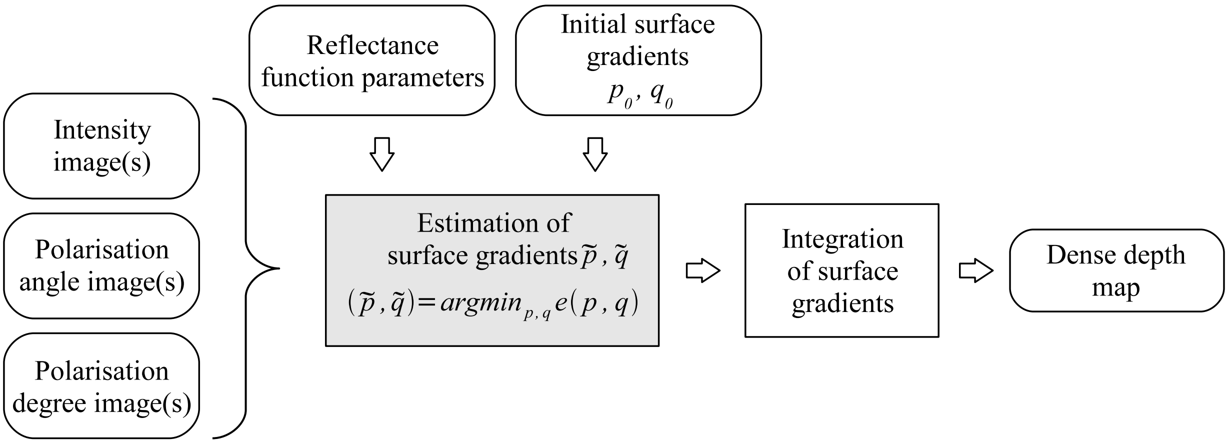


Figure 8.2: Overview of the Shape from Photopolarimetric Reflectance method. The processing steps (rectangles) and their input and output data (rounded rectangles).

described in section 4.3. No sufficiently accurate physical model exists so far which is able to describe the polarisation behaviour of light scattered from a rough metallic surface. We will therefore determine the functions $R_\Phi(\vec{n}, \vec{s}, \vec{v})$ and $R_D(\vec{n}, \vec{s}, \vec{v})$, describing the polarisation angle and degree of the material, respectively, for the phase angle α between the vectors $\vec{s}$ and $\vec{v}$ over a wide range of illumination and viewing configurations. To obtain analytically tractable relations rather than discrete measurements, we fit phenomenological models to the obtained measurements (cf. Section 8.2).

In the following paragraphs we describe a global and a local approach to adapt the surface gradients $p(u, v)$ and $q(u, v)$ to the observed photopolarimetric properties $I(u, v)$, $\Phi(u, v)$, and $D(u, v)$ by solving the (generally nonlinear) system of equations (8.1)–(8.3). An overview of this approach is shown in Fig. 8.2.

The 3D surface profile $z(u, v)$ is then obtained by integration of the surface gradients according to the method proposed by (Simchony et al. 1990). Their approach determines the surface $z(u, v)$ of maximum consistency with the usuable non-integrable gradient field $(\tilde{p}(u, v), \tilde{q}(u, v))$ by minimising the integrability error term

$$e_{\text{int}} = \sum_{u,v} \left[\left(\frac{\partial z}{\partial x} - \tilde{p} \right)^2 + \left(\frac{\partial z}{\partial y} - \tilde{q} \right)^2 \right] \tag{8.4}$$

as defined by Horn (1989b). Compared to the original approach by Horn (1989b), the direct analytical optimisation method by Simchony et al. (1990), which is implemented in

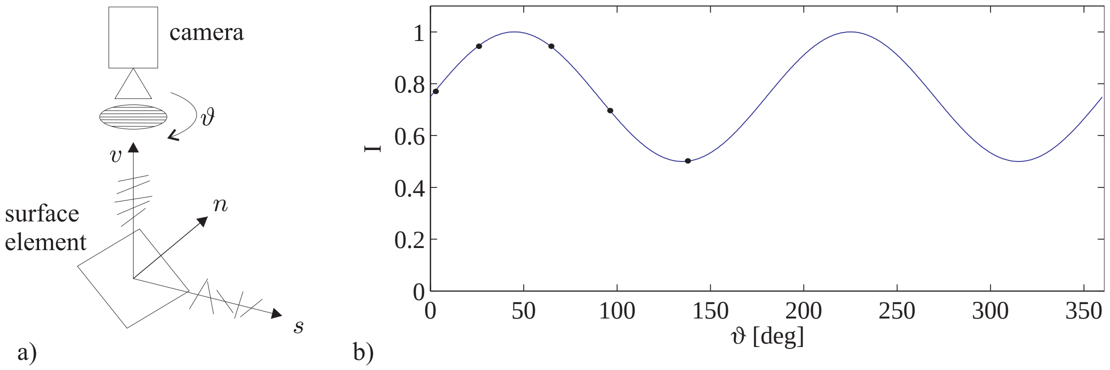


Figure 8.3: Measurement of linear polarisation using a rotating linear polarising filter. a) Unpolarised light incident from direction $\vec{s}$ is partly polarised upon reflection on the surface before passing the linear polarisation filter in front of the camera. The small lines on the $\vec{v}$ and $\vec{s}$ arrows indicate the direction of the electric field vector. b) Intensity captured by the camera over the rotation angle of the linear polarising filter.

frequency space, is computationally more efficient.

8.1.1 Measurement of the linear polarisation state

Most imaging sensors (for example CCD and CMOS sensors and photographic film) are sensitive only to the energy of the incoming light but not to its polarisation state. By mounting a rotating linear polarising filter in front of a camera (cf. Fig. 8.3a) and capturing multiple images with different angles ϑ of the transmission axis of the linear polariser, the measured irradiance oscillates sinusoidally between a maximum $I_{\max}$ and a minimum $I_{\min}$:

$$I(\vartheta) = I_c + I_v cos(2(\vartheta - \Phi)). \tag{8.5}$$

The polarisation angle $\Phi \in [0°, 180°]$ denotes the orientation under which maximum transmitted radiance $I_{\max} = I_c + I_v$ is observed. The parameter I_c corresponds to the intensity of the unpolarised fraction of the reflected light, while I_v is the amplitude of the fitted sinusoid and corresponds to the polarised fraction. The value of I_c is used as the pixel brightness of the intensity image $I(u, v)$. The degree of polarisation is defined by $D = (I_{\max} - I_{\min})/(I_{\max} + I_{\min}) = I_v/I_c \in [0, 1]$. The parameters I_c, I_v and Φ can be determined by fitting three or more intensity measurements captured at various polariser angles ϑ to Eq. (8.5). We use the linear fitting procedure introduced in (Rahmann 1999). In principle, three measurements would be sufficient to determine the three parameters

I_c, I_v, and Φ, but the fit becomes less noise-sensitive and thus more accurate when more measurements are used.

8.2 Empirical determination of photopolarimetric models

This section explains how the reflectance and polarisation properties of the surface material are measured and described in terms of suitable analytical phenomenological models for further processing.

8.2.1 Measurement of reflectance properties

According to Nayar et al. (1991), the reflectance of a typical rough metallic surface consists of three components: a diffuse (Lambertian) component, the specular lobe, and the specular spike. The diffuse component is generated by internal multiple scattering processes. The specular lobe, which is caused by single reflection at the surface, is distributed around the specular direction and may be rather broad. The specular spike is concentrated in a small region around the specular direction and represents mirror-like reflection, which is dominant in the case of smooth surfaces. Fig. 8.1a illustrates the three components of the reflectance function. We define an analytical form for the reflectance for which we perform a least-mean-squares fit to the measured reflectance values, depending on the incidence angle θ_i and the angle θ_r between the specular direction $\vec{r}$ and the viewing direction $\vec{v}$ (cf. Fig. 8.1a):

$$R_I(\theta_i, \theta_r, \alpha) = \rho_0 \left[\cos\theta_i + \sum_{n=1}^{N} \sigma_n \cdot (\cos\theta_r)^{m_n} \right]. \tag{8.6}$$

For $\theta_r > 90^\circ$ only the diffuse component is considered. The reflectance measurement is performed for a small part of the surface, for which the albedo ρ_0 can be assumed to be constant. The shapes of the specular components of the reflectance function are approximated by $N = 2$ terms proportional to powers of $\cos\theta_r$. The coefficients $\{\sigma_n\}$ denote the strength of the specular components relative to the diffuse component, while the exponents $\{m_n\}$ denote their widths. Generally, all introduced phenomenological parameters depend on the phase angle α. The angle θ_r can be expressed in terms of incidence angle, emission angle, and phase angle according to

$$\cos\theta_r = 2\cos\theta_i \cos\theta_e - \cos\alpha, \tag{8.7}$$

such that our phenomenological reflectance model only depends on the incidence angle θ_i, the emission angle θ_e, and the phase angle α. A goniometer is used to adjust the angles θ_i and θ_e. The illumination vector $\vec{s}$, the viewing direction $\vec{v}$, and thus the phase angle α between the vectors $\vec{s}$ and $\vec{v}$ are assumed to be constant across the image.

For our experiments we have used a CCD camera with a linear response curve. As the dynamic range of the CCD camera used for our experiments is not sufficiently high to cover both the diffuse and the specular reflectance components, we acquired a series of images over a wide range of shutter times for each configuration of θ_i, θ_e, and α, combining each series into a single high dynamic range image. Pixels with greyvalues in an intermediate range are selected from each individual frame and normalised by the corresponding exposure time, respectively. The normalised pixel brightnesses are then arranged into a synthetised high dynamic range image, which is used in all further processing steps. The reflectance of the sample surface under the given illumination conditions is then obtained by computing the average greyvalue over an area in the synthesised high dynamic range image that contains a flat part of the sample surface. The reflectance value is obtained up to the overall scale factor ρ_0, which is computed based on the image used for surface reconstruction. A reflectance measurement typical for raw forged or cast iron surfaces is shown in Fig. 8.1b for $\alpha = 75°$. The corresponding parameters of the reflectance map according to Eq. (8.6) are given by $\sigma_1 = 3.85$, $m_1 = 2.61$, $\sigma_2 = 9.61$, and $m_2 = 15.8$, where the specular lobe is described by σ_1 and m_1 and the specular spike by σ_2 and m_2, respectively.

8.2.2 Measurement of polarisation properties

The measurement of the polarisation properties of the light reflected from the surface is similar to the reflectance measurement. For each configuration of goniometer angles, five high dynamic range images are acquired through a linear polarisation filter at orientation angles ω of 0°, 45°, 90°, 135°, and 180°. For each filter orientation ω, an average pixel intensity over an image area containing a flat part of the sample surface is computed as described in Section 8.2.1. The sinusoidal fitting procedure described in Section 8.1.1 yields the linear polarisation state.

According to Fig. 8.1b, the rotation angles of the goniometer define the surface normal $\tilde{\vec{n}} = (-\tilde{p}, -\tilde{q}, 1)$ of the sample surface in a coordinate system with positive x and zero y component of the illumination vector $\vec{s}$, corresponding to $p_s < 0$ and $q_s = 0$. Without loss of generality we will in the following assume a viewing direction $\vec{v} = (0, 0, 1)^T$. The surface normal $\vec{n}$ in the world coordinate system, in which the azimuth angle of the light source is denoted by the angle ψ, is related to $\tilde{\vec{n}}$ by a rotation $R_z(\psi)$ around the z axis with $\vec{n} = R_z(\psi)\tilde{\vec{n}}$, leading to

$$\tilde{p} = p\cos\psi + q\sin\psi, \quad \tilde{q} = -p\sin\psi + q\cos\psi. \tag{8.8}$$

Due to the lack of an accurate physically motivated model for the polarisation properties of rough metallic surfaces, we perform a polynomial fit in terms of the surface gradients $\tilde{p}$ and $\tilde{q}$ to the measured values of the polarisation angle Φ and degree D. In this framework, the modelled polarisation angle R_Φ is represented by an incomplete third-degree polynomial of the form

$$R_\Phi(\tilde{p}, \tilde{q}) = a_\Phi + b_\Phi\tilde{p}\tilde{q} + c_\Phi\tilde{q} + d_\Phi\tilde{p}^2\tilde{q} + e_\Phi\tilde{q}^3. \tag{8.9}$$

The constant offset a_Φ can be made zero by correspondingly defining the zero position of the orientation angle ω of the linear polarisation filter. Eq. (8.9) is antisymmetric in $\tilde{q}$ with respect to a_Φ. At the same time, we have $R_\Phi(\tilde{p}, \tilde{q}) = a_\Phi = \text{const}$ for $\tilde{q} = 0$, corresponding to coplanar vectors $\vec{n}$, $\vec{s}$, and $\vec{v}$. These properties are required for geometric symmetry reasons as long as the interaction between the incident light and the surface material can be assumed to be isotropic, since the polarisation angle is then antisymmetric with respect to the plane containing the camera, the surface point, and the light source.

The observed polarisation degree R_D is represented in an analogous manner by an incomplete second-degree polynomial of the form

$$R_D(\tilde{p}, \tilde{q}) = a_D + b_D\tilde{p} + c_D\tilde{p}^2 + d_D\tilde{q}^2. \tag{8.10}$$

In Eq. (8.10), symmetry in $\tilde{q}$ is imposed for geometric reasons, once more due to the assumed isotropic interaction between the incident light and the surface material. Accordingly, the maximum of the polarisation degree is always observed for surface normals $\vec{n}$ coplanar with the viewing direction $\vec{v}$ and the illumination vector $\vec{s}$, and the behaviour of the polarisation degree is symmetric with respect to that plane. Fig. 8.4 illustrates the polarisation properties of a raw forged iron surface at a phase angle of $\alpha = 75°$ along with the polynomial fits according to Eqs. (8.9) and (8.10). The polynomial model (8.9) of the polarisation angle yields a reasonable representation of the observations. For rough metallic surfaces like the one regarded in Fig. 8.4, however, the behaviour of the polarisation

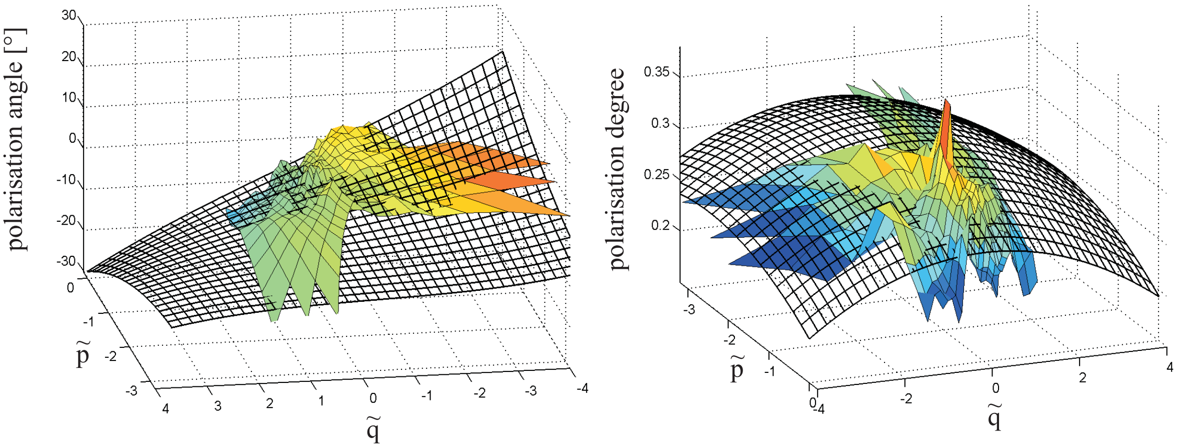


Figure 8.4: Measured and modelled polarisation properties of a raw forged iron surface for $\alpha = 75°$. Left: polarisation angle. Right: polarisation degree.

degree appears rather irregular. We additionally found that it is strongly dependent on local variations of the surface roughness (cf. Section 11.2). We will utilise the fitted model for the polarisation degree in our experiments on synthetic data (cf. Section 11.1), showing that it does not increase the accuracy of 3D reconstruction when used together with the intensity. Hence, in our real-world experiments we will limit the use of polarisation information to the polarisation angle (cf. Section 11.2). In the general case, however, the preferrable choice of polarisation features to be used for a given 3D reconstruction task depends on the polarisation properties of the regarded surface material.

8.3 Global optimisation scheme

The first solving technique is based on the optimisation of a global error function simultaneously involving all image pixels. This approach is described in detail by Horn and Brooks (1989a), Horn (1989b), and Jiang and Bunke (1996). One part of this error function is the intensity error term

$$e_I = \sum_{l=1}^{L} \sum_{u,v} \left[I^{(l)}(u,v) - R_I \left(\rho(u,v), \theta_i^{(l)}(u,v), \theta_e(u,v), \alpha^{(l)} \right) \right]^2 . \tag{8.11}$$

The number of light sources and thus of acquired images is given by L, and both the incidence angle $\theta_i^{(l)}$ and the phase angle $\alpha^{(l)}$ depend on the individual light source l. The angles $\theta_i^{(l)}$ and θ_e depend on the surface orientation at image location (u, v), while the phase angles $\alpha^{(l)}$ are assumed to be constant across the image.

As the pixel intensity information alone is not necessarily sufficient to provide an unambiguous solution for the surface gradients $p(u, v)$ and $q(u, v)$, a regularisation constraint e_s is introduced which requires smoothness of the surface, i. e. for example small absolute values of the directional derivatives of the surface gradients. We will therefore make use of the additional error term

$$e_s = \sum_{u,v} \left[\left(\frac{\partial p}{\partial x} \right)^2 + \left(\frac{\partial p}{\partial y} \right)^2 + \left(\frac{\partial q}{\partial x} \right)^2 + \left(\frac{\partial q}{\partial y} \right)^2 \right] \tag{8.12}$$

(Horn and Brooks 1989a; Jiang and Bunke 1996). In the scenarios regarded in this thesis, the assumption of a macroscopically smooth surface is realistic. Note that the microscopic surface roughness which strongly affects the reflectance function is much smaller than the surface elements considered for shape reconstruction. Thus the smoothness term can be applied to microscopically rough surfaces. Using Eq. (8.12) for the reconstruction of wrinkled surfaces can lead to unsatisfactory results, in this case it can be replaced by the departure from integrability error term (8.4) as discussed in detail by Horn (1989b).

To integrate polarisation angle and degree into the 3D surface reconstruction framework, we define two error terms e_Φ and e_D which denote the deviations between the measured values and those computed using the corresponding phenomenological model, respectively:

$$e_\Phi = \sum_{l=1}^{L} \sum_{u,v} \left[\Phi^{(l)}(u, v) - R_\Phi \left(\theta_i^{(l)}(u, v), \theta_e(u, v), \alpha^{(l)} \right) \right]^2 \tag{8.13}$$

$$e_D = \sum_{l=1}^{L} \sum_{u,v} \left[D^{(l)}(u, v) - R_D \left(\theta_i^{(l)}(u, v), \theta_e(u, v), \alpha^{(l)} \right) \right]^2 . \tag{8.14}$$

Based on the feature-specific error terms e_I, e_Φ, and e_D, a combined error term e is defined which takes into account both reflectance and polarisation properties:

$$e = e_s + \lambda e_I + \mu e_\Phi + \nu e_D. \tag{8.15}$$

Minimising error term (8.15) yields the surface gradients $p(u, v)$ and $q(u, v)$ that optimally correspond to the observed reflectance and polarisation properties, where the parameters λ, μ, and ν denote the relative weights of the individual reflectance-specific and polarisation-specific error terms.

With the discrete approximations $\frac{\partial p}{\partial x}(u,v) = \left[p(u+1,v) - p(u-1,v)\right]/2$ and $\frac{\partial p}{\partial y}(u,v) = \left[p(u,v+1) - p(u,v-1)\right]/2$ for the second derivatives of the surface $z(u,v)$ and $\bar{p}(u,v)$ as the local average over the four nearest neighbours of pixel (u,v) we obtain an iterative update rule for the surface gradients by setting the derivatives of the error term e with respect to p and q to zero:

$$\begin{aligned} p_{n+1} &= \bar{p}_n + \lambda \sum_{l=1}^{L} \left(I - R_I(\bar{p}_n, \bar{q}_n)\right) \frac{\partial R}{\partial p} + \mu \sum_{l=1}^{L} (\Phi - R_\Phi(\bar{p}_n, \bar{q}_n)) \frac{\partial R_\Phi}{\partial p} + \\ &\qquad \nu \sum_{l=1}^{L} (D - R_D(\bar{p}_n, \bar{q}_n)) \frac{\partial R_D}{\partial p}, \end{aligned} \tag{8.16}$$

where n denotes the iteration index. A corresponding expression for q is obtained in an analogous manner. This derivation is described in more detail by Horn (1989b). The initial values $p_0(u,v)$ and $q_0(u,v)$ must be provided based on a-priori knowledge about the surface or on independently obtained depth data (cf. Section 9). The partial derivatives in Eq. (8.16) are evaluated at $(\bar{p}_n, \bar{q}_n)$, respectively, making use of the phenomenological model fitted to the measured reflectance and polarisation data (cf. Section 8.2).

The reconstruction is done in a multi-scale approach to speed up convergence and avoid getting stuck in local minima. Reconstruction of the gradient field starts at a low resolution and is repeated on the next pyramid level, using the gradients estimated at the previous level as initial gradient values. The surface profile $z(u,v)$ is then derived from the resulting gradients $p(u,v)$ and $q(u,v)$ by means of numerical integration of the gradient field according to the method suggested by Simchony et al. (1990).

The reconstruction method presented above requires the knowledge of the surface albedo, which is often unknown in practical applications. For many materials the intensity reflectance function is of the form

$$R_I(\rho(u,v), p(u,v), q(u,v)) = \rho(u,v)\tilde{R}(p(u,v), q(u,v)). \tag{8.17}$$

Many objects exhibit a uniform surface albedo ρ_0 over the whole surface. One possible way to determine ρ_0 is based on measuring the irradiance I_m for surface orientation, camera and light source directions that correspond to the maximum of the reflectance function. For Lambertian surfaces, these are regions where light source direction $\vec{s}$ and the surface normal $\vec{n}$ are equal. In case of a strongly specular reflectance function, the maximum intensity I_m is measured for viewing directions equal to the direction of specular reflection $\vec{r} = \vec{v}$. Solving Eq. (8.17) for ρ and evaluating it at the maximum of $\tilde{R}$ yields:

$$\rho_0 = \frac{1}{L} \sum_{l=1}^{L} \frac{I_m^{(l)}}{\tilde{R}}. \tag{8.18}$$

In principle, a single image is already sufficient to determine the value of ρ_0 as long as the object contains a surface orientation that results in the maximum of the the reflectance function.

An albedo estimation according to Eq. (8.18) is not possible when the maximum intensity in the image is observed at points that do not correspond to the maximum of the reflectance function. In the global optimisation scheme the surface albedo ρ_0 can then be estimated in each iteration step n simultaneously along with the surface gradients. This is achieved by solving Eq. (8.6) for $\rho(u, v)$ individually for each pixel (u, v) based on the values of $p_n(u, v)$ and $q_n(u, v)$ according to

$$\rho_n(u, v) = \frac{I(u, v)}{\tilde{R}\left(p_n(u, v), q_n(u, v)\right)} \tag{8.19}$$

with $\tilde{R}(p, q)$ defined according to Eq. (8.17). The uniform albedo ρ_0 is then obtained by computing an appropriate average of the computed $\rho_n(u, v)$ values. For the strongly specular surfaces regarded in our experiments, we found that the median of $\rho_n(u, v)$ provides a more robust estimate of ρ_0 than the mean, since already a small number of pixels with inaccurately estimated surface gradients (which occur especially at the beginning of the iteration procedure) leads to a significant shift of the mean value while leaving the median value largely unaffected. If no a-priori information about the surface gradients is available, the initial guess of ρ_0, which in turn depends on the initial guess of $p(u, v)$ and $q(u, v)$, has a strong influence on the solution found by the global optimisation scheme. Such a-priori information can be obtained based on independently measured depth data (cf. Section 9).

8.4 Local optimisation scheme

Provided that the model parameters of the reflectance and polarisation functions R_I, R_Φ, and R_D are known and measurements of intensity and polarisation properties are available for each image pixel, the surface gradients $p(u, v)$ and $q(u, v)$ can be obtained by solving the nonlinear system of equations (8.1)–(8.3) individually for each pixel. For this purpose we make use of the Levenberg-Marquardt algorithm (Madsen et al. 2004). In the overdetermined case, the root of Eqs. (8.1)–(8.3) is computed in the least-mean-squares sense. The contributions from the different terms are then weighted according to the measurement errors, respectively, which we have determined to $\sigma_I \approx 10^{-3} I_{\mathrm{spec}}$ with I_{spec} as the intensity of the specular reflections, $\sigma_\Phi \approx 0.1^\circ$, and $\sigma_D \approx 0.02$. The surface profile $z(u, v)$ is again derived from the resulting gradients $p(u, v)$ and $q(u, v)$ by means of numerical integration of the gradient field (Simchony et al. 1990).

It is straightforward to extend this approach to photopolarimetric stereo because each light source provides an additional set of equations. Eq. (8.1) can only be solved, however, when the surface albedo $\rho(u, v)$ is known for each surface point. A constant albedo can be assumed in many applications. If this assumption is not valid, albedo variations will strongly affect the accuracy of surface reconstruction.

For surfaces with a reflectance function of the form (8.17), but non-uniform albedo $\rho(u, v)$ it is possible to utilise two images $I_1(u, v)$ and $I_2(u, v)$ acquired under different illumination conditions, such that Eq. (8.1) can be replaced by

$$\frac{I_1(u, v)}{I_2(u, v)} = \frac{\tilde{R}_1 \left(p(u, v), q(u, v)\right)}{\tilde{R}_2 \left(p(u, v), q(u, v)\right)} \tag{8.20}$$

In Eq. (8.20), the albedo cancels out. The quotient approach has been introduced in the context of photoclinometric analysis of planetary surfaces by McEwen (1985) and has been integrated into the Shape from Shading formalism by Wöhler and Hafezi (2005) and by Lena et al. (2006).

An advantage of the described local approach is that the 3D reconstruction result is not affected by additional constraints such as smoothness of the surface but directly yields the surface gradient vector for each image pixel. A drawback, however, is the fact that due to the inherent nonlinearity of the problem, existence and uniqueness of a solution for $p(u, v)$ and $q(u, v)$ are not guaranteed for both the albedo-dependent and the albedo-independent case. But in the experiments presented in Chapter 11 we show that in practically relevant scenarios reasonable solutions for the surface gradient field and the resulting depth $z(u, v)$ are obtained even in the presence of noise.

9 Shape from Photopolarimetric Reflectance and Depth

Since the solution of Shape from Shading and, depending on the reflectance function, Shape from Photopolarimetric Reflectance may be ambiguous as long as single images are regarded, integrating additional information into the surface reconstruction process leads to improved reconstruction results. For example, a sparse 3D point cloud of the object surface can be reconstructed by stereo vision, laser triangulation, or shadow analysis. Previous approaches either merge the results of Stereo and SfS (Cryer et al. 1995) or embed the SfS algorithm into Stereo (Samaras et al. 2000) or SfM (Lim et al. 2005) algorithms.

In this chapter we will describe how independently acquired additional depth information can be integrated into the SfPR framework outlined in Chapter 8. We found that the fusion between SfPR and depth-measuring algorithms is especially useful if the depth data are dense but display a considerable amount of noise, or if they are accurate but only available for a sparse set of surface points. Hence, in Section 9.1 we will concentrate on dense but noisy depth information, examining as an example the monocular DfD technique. In Section 9.2 we will regard reasonably accurate but sparse depth information, based on the examples of stereo image analysis and SfM. An overview of both approaches is shown in Fig. 9.1.

9.1 Dense but noisy depth information – Depth from Defocus

A well-known monocular technique to recover the depth of scene points from still images is Depth from Defocus (DfD). This method makes the realistic assumption of a finite aperture of the camera lens and relies on the fact that the image of a scene point situated at a distance z from the camera becomes more and more blurred with increasing depth offset $(z - z_0)$ between the scene point at distance z and the plane at distance z_0 on which the camera is focused. A general property of the DfD approach is that the computed depth

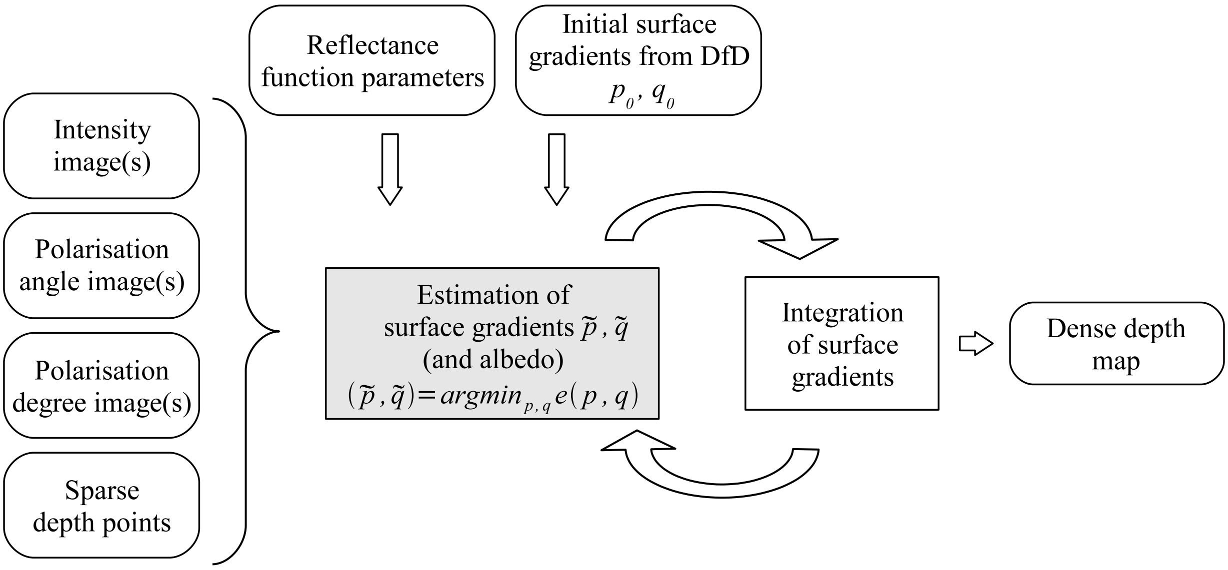


Figure 9.1: Overview of the Shape from Photopolarimetric Reflectance and Depth method. The processing steps (rectangles) and their input and output data (rounded rectangles). Not all input data shown here is required, the method can also work without DfD and polarisation data.

map is dense, i. e. for each image location a depth value is determined, but the resulting depth values tend to display a considerable scatter.

9.1.1 Calibration of the DfD method, data acquisition

As shown in Sections 3 and 7.1.1, the observed image $I(u, v)$ can be modelled by the convolution of the “ideal” image $I_0(u, v)$ with the point spread function (PSF) G according to $I(u, v) = G * I_0(u, v)$. As a measure for the PSF width, we will denote the standard deviation of the Gaussian PSF in frequency space by σ, where the observed image blur decreases with increasing value of σ.

The calibration procedure for estimating DfD then involves the determination of the lens-specific characteristic curve $\sigma\,(z - z_0)$ (Chaudhuri et al. 1999). For this purpose we acquire two pixel-synchronous images of a rough, uniformly textured plane surface consisting of forged iron, inclined by 45° with respect to the optical axis. The image part in which the intensity displays maximum standard deviation (i. e. most pronounced high spatial frequencies) is sharp and thus situated at distance z_0. A given difference in pixel coordinates with respect to that image location directly yields the corresponding depth offset $(z - z_0)$. The first image is taken with small aperture, e. g. $f/8$, resulting in virtually absent image blur, while the second image is taken with the aperture that will be used later on for 3D reconstruction, e. g. $f/2$, resulting in a perceivable image blur that depends on the depth offset $(z - z_0)$.

The images are partitioned into quadratic windows, for each of which the average depth offset $(z - z_0)$ is known. After Tukey windowing, the PSF width parameter σ in frequency space is computed based on the ratio $Q(\omega_u, \omega_v)$ of the amplitude spectra of the corresponding windows of the first and the second image, respectively:

$$\sigma = \langle \tilde{\sigma}(\omega_u, \omega_v) \rangle_{\omega_u, \omega_v} \quad \text{with} \quad \tilde{\sigma}(\omega_u, \omega_v) = -\frac{\omega_u^2 + \omega_v^2}{\ln Q(\omega_u, \omega_v)} \tag{9.1}$$

(Subbarao 1988), where $\langle \ldots \rangle_{\omega_u, \omega_v}$ denotes the average over the coordinates ω_u and ω_v in frequency space. Only the range of intermediate spatial frequencies is regarded in order to reduce the influence of noise on the resulting value for σ. If the amplitude spectrum of the examined image window displays a very low value at (ω_u, ω_v), the corresponding ratio value $Q(\omega_u, \omega_v)$ tends to be inaccurate, which may result in a substantial error of σ. Hence, we first compute σ according to Eq. (9.1), identify all spatial frequencies (ω_u, ω_v) for which $\tilde{\sigma}(\omega_u, \omega_v)$ deviates by more than one standard deviation from σ, and recompute σ after neglecting these outliers.

For a given value of σ, the corresponding value of $(z - z_0)$ is ambiguous since two depth values $z_1 < z_0$ and $z_2 > z_0$ may correspond to the same value of σ. In our experiments we avoided this two-fold ambiguity by placing the complete surface to be reconstructed behind the plane at distance z_0, implying $z > z_0$. One would expect $\sigma \to \infty$ for $z \to z_0$, since ideally the small-aperture image and the large-aperture image are identical for $z = z_0$. We found empirically, however, that due to the imperfections of the optical system, even for $z = z_0$ an image window acquired with larger aperture is slightly more blurred than the corresponding image window acquired with smaller aperture. This remains true as long as the small aperture is sufficiently large for diffraction effects to be small. As a consequence, σ obtains a finite maximum value at $z = z_0$ and decreases continuously for increasing z.

Pentland (1987) utilises the approximation of geometric optics and assumes that the image of a point light source corresponds to a point for a sharp image (obtained with the camera sensor located in the focal plane) and to a circle of uniform brightness, the so-called circle of confusion, for a blurred image (obtained with the camera sensor displaced by a distance Δv from the focal plane). The geometric optics approach implies that the diameter of the circle of confusion is proportional to the value of Δv (Subbarao 1988). In turn, the

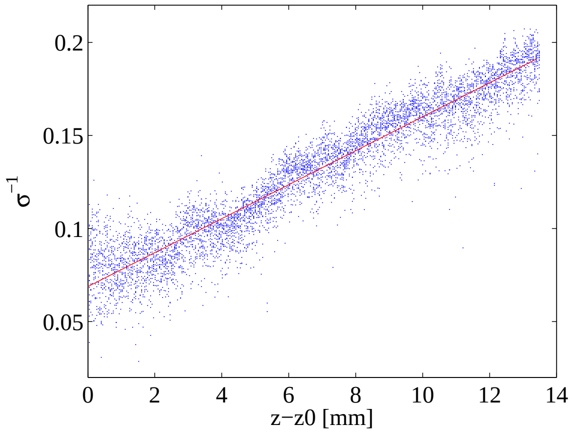


Figure 9.2: Calibration of the DfD algorithm: measurements $(\sigma^{-1}, (z - z_0))$ and fitted characteristic curve.

PSF width in *image* space (being proportional to σ^{-1}) is assumed to be proportional to the diameter of the circle of confusion, implying $\sigma^{-1} \rightarrow 0$ for $z \rightarrow z_0$. For the lenses, CCD sensors, and object distances regarded in our experiments (cf. Section 11.2.1), it follows from the models by Pentland (1987) and Subbarao (1988) that σ^{-1} is proportional to $|z - z_0|$ for small values of $|z - z_0|$ of some millimetres and for small diameters of the circle of confusion of a few pixels. In the general case, the geometric optics approximation may introduce systematic errors (cf. Section 7.1.1).

In our experimental setting, we found that the measured $(\sigma^{-1}, (z - z_0))$ data points can be represented fairly well by a linear function, displaying a non-zero offset due to the aforementioned imperfections of the optical system. This result is shown in Fig. 9.2, which will serve as our DfD calibration curve.

9.1.2 Fusion of SfPR with DfD

Once the characteristic curve $\sigma(z - z_0)$ is known, it is possible to extract a dense depth map from a pixel-synchronous pair of images of a surface of unknown shape, provided that the images are acquired at the same focus position and with the same apertures as the calibration images. The resulting depth map $z_{\mathrm{DfD}}(u, v)$, however, tends to be very noisy. It is therefore favourable to fit a plane $z_{\mathrm{DfD}}(u, v)$ to the computed depth points, since higher-order information about the surface is usually not contained in the noisy DfD data. This procedure reveals information about the large-scale properties of the surface.

Approximate surface gradients can then be obtained by computing the partial derivatives $p_{\mathrm{DfD}}(u,v) = \partial z_{\mathrm{DfD}}(u,v)/\partial x$ and $q_{\mathrm{DfD}}(u,v) = \partial z_{\mathrm{DfD}}(u,v)/\partial y$.

In many cases there exists no unique solution for the surface gradients $p(u,v)$ and $q(u,v)$ within the SfPR framework, especially for highly specular reflectance functions. This applies both to the global (Section 8.3) and to the local (Section 8.4) optimisation scheme. Therefore, the obtained solution tends to depend strongly on the initial values $p_0(u,v)$ and $q_0(u,v)$. As we assume that no a-priori information about the surface is available, we initialise the optimisation scheme with $p_0(u,v) = p_{\mathrm{DfD}}(u,v)$ and $q_0(u,v) = q_{\mathrm{DfD}}(u,v)$, thus making use of the large-scale surface gradients obtained by DfD analysis. The ambiguity of the solution of the global optimisation scheme is even more pronounced when no a-priori knowledge about both the surface gradients and the albedo is available. In such cases, which are often encountered in practically relevant scenarios, an initial albedo value is computed according to Eq. (8.19) based on the initial surface gradients $p_{\mathrm{DfD}}(u,v)$ and $q_{\mathrm{DfD}}(u,v)$ (cf. Section 8.2.1). We found experimentally that it is advantageous to keep this albedo value constant during the iteration process as long as no additional constraints can be imposed on the surface, since treating the albedo as a further free parameter in the iteration process increases the manifold of local minima of the error function.

The DfD data are derived from two images acquired with large and small aperture, respectively. In practice, it is desirable but often unfeasible to use the well-focused image acquired with small aperture for 3D reconstruction – the image brightness then tends to become too low for obtaining reasonably accurate polarisation data. The surface reconstruction algorithm thus has to take into account the position-dependent PSF. Correspondingly, the observed image $I(u,v)$ is described as a convolution

$$I(u,v) = G(u,v) * I_0(u,v) \tag{9.2}$$

of the "true" image $I_0(u,v)$ with the spatially varying PSF $G(u,v)$.

We incorporate the DfD information into the global optimisation scheme since it is not possible to introduce PSF information (which applies to a local neighbourhood of a pixel) into an approach based on the separate evaluation of each individual pixel, like the local method described in Section 8.4. The error terms (8.11), (8.13), and (8.14) of the SfPR scheme described in Section 8.3 are modified according to

$$e_I^{\mathrm{PSF}} = \sum_{u,v} \left[I(u,v) - G * R_I \left(\rho(u,v), p(u,v), q(u,v) \right) \right]^2 \tag{9.3}$$

$$e_\Phi^{\mathrm{PSF}} = \sum_{u,v} \left[\Phi(u,v) - G * R_\Phi(p(u,v), q(u,v), \alpha) \right]^2 \tag{9.4}$$

$$e_D^{\mathrm{PSF}} = \sum_{u,v} \left[D(u,v) - G * R_D(p(u,v), q(u,v), \alpha) \right]^2 \tag{9.5}$$

describing the mean square deviation between the observed intensity and polarisation values and the modelled reflectances convolved with the PSF G extracted from the image as described in Section 9.1. This approach is related to the Shape from Shading scheme for blurred images introduced by Joshi and Chaudhuri (2004). In that work, however, the PSF width is estimated simultaneously with the surface gradients, while we independently determine the PSF width in dependence of the location in the image during the DfD analysis. The iterative update rule (8.16) then becomes

$$\begin{aligned} p_{n+1} = {} & \bar{p}_n + \lambda (I - G * R_I) G * \frac{\partial R}{\partial p} + \mu (\Phi - G * R_\Phi) G * \frac{\partial R_\Phi}{\partial p} + \\ & \nu (D - G * R_D) G * \frac{\partial R_D}{\partial p}. \end{aligned} \tag{9.6}$$

An analogous expression is readily obtained for q.

9.2 Accurate but sparse depth information

Many methods to recover accurate 3D information exist (cf. Chapter I). One of the most widespread method to recover 3D information about a scene is stereo analysis. Another well-known geometric approach to 3D scene reconstruction is Structure from Motion, where point features are tracked across an image sequence acquired by a single moving camera (Faugeras 1993). The SfM approach recovers both the 3D coordinates of the tracked scene points and the path of the camera up to an unknown but uniform scaling factor, which we determine based on a-priori knowledge about the scene, or by using the Structure from Motion and Defocus method presented in Chapter 7.

9.2.1 Description of the employed Stereo algorithm

In several real-world experiments we use a block-matching Stereo algorithm to generate sparse depth data. We assume that the images are rectified to standard stereo geometry with epipolar lines parallel to the horizontal image axis (Krüger et al. 2004). For each pixel i at position (u, v) in the left image, a corresponding point is searched along the epipolar line in the right image. We use the normalized cross-correlation coefficient $\mathcal{C}$ as similarity measure. A square region of 7×7 pixels of the left image (L) is correlated with regions on the corresponding epipolar line in the right image (R) for all candidate disparities d, resulting in an array of correlation coefficients $c_i(d) = \mathcal{C}(L_{u,v}, R_{u-d,v})$. The disparity with the maximum correlation coefficient $d_i = \operatorname{argmax}_d (c_i(d))$ is determined, and a parabola $P(d) = ad^2 + bd + e$ is fitted to the local neighbourhood of the maxima. By locating the position of the maximum of $P(d)$, $d_i = -b/(2a)$ the disparity can be estimated with subpixel accuracy. To ensures that well localised correspondences are considered for further processing, only fits with $c_i(d_i) > 0.8$ and $a_i < -0.1$ are used. The coordinates of a point (u_i, v_i) in the left stereo camera coordinate system are then given by $Z_i = bf/d_i$, $X_i = u_i b/d_i$, and $Y_i = v_i b/d_i$ (Faugeras 1993). The focal length f and base distance b between the cameras are determined in a separate off-line binocular camera calibration step according to the method by Krüger et al. (2004). Our block-matching algorithm performs a left-right consistency check prior to establishing a point correspondence. In both regarded examples, a disparity value could be determined for a fraction of only 0.3 percent of all pixels, due to limited texture, repeating patterns, or different appearance of corresponding surface parts in the stereo images as a consequence of the strongly specular reflectance behaviour.

Apart from block-matching techniques, a variety of dense Stereo algorithms have been proposed, which compute a depth value for each image pixel independent of the presence of texture. An intensity-based variational dense Stereo approach is described by Horn (1986). Scharstein and Szeliski (2002) provide an extensive survey about dense Stereo methods and their performance. However, parts of the surface may show no surface texture at all, or corresponding parts of the stereo image pair no not display a similar structure. The latter behaviour e. g. occurs due as a consequence of specular reflectance properties leading to a different appearance of the respective surface part in the stereo images. In such cases of missing or contradictory texture information, dense Stereo algorithms usually interpolate the surface across the ambiguous image parts, leading to an inaccurate 3D reconstruction result for the corresponding region. This problem is addressed by Hirschmüller (2006), who proposes an intensity consistent disparity selection for dealing with untextured surface parts and suggests a discontinuity preserving interpolation approach for filling holes in the disparity map. In this framework, however, interpolation is performed based on a-priori assumptions of the surface rather than reflectance properties. Hence, we prefer to compute depth points only in places where point correspondences can be established unambiguously and accurately and to compute dense depth data in a subsequent step based on an integration of the available photometric or photopolarimetric information.

9.2.2 Fusion of SfPR with sparse depth data

Horovitz and Kiryati (2004) have shown that methods directly enforcing sparse depth constraints during the surface gradient integration suffer from the fact that the sparse depth values only have a local influence and lead to spikes in the reconstructed surface. Hence, they propose a local approach, assigning a radial weighting function to each depth point, and a global approach which consists of interpolating a bias correction surface to the depth differences between gradient reconstruction and depth points. But even in their framework, the locality of the influence of the depth points on the gradient field is only partially removed.

Our approach to incorporate sparse depth information into the global optimisation scheme presented in Section 8.3 consists of defining a depth error term based on the surface gradient field and depth differences between sparse 3D points. The depth difference between two 3D points i and j is given by

$$(\Delta z)^{ij} = Z^j - Z^i. \tag{9.7}$$

The corresponding depth difference of the reconstructed surface gradient field is calculated by integration along a path C^{ij} between the coordinates (u^j, v^j) and (u^i, v^i):

$$(\Delta z)^{ij}_{\text{surf}} = \int_{C^{ij}} (p dx + q dy) \, . \tag{9.8}$$

In our implementation the path C^{ij} is approximated by a list of K discrete pixel positions (u_k, v_k) with $k = 1, \ldots, K$. While in principle any path C^{ij} between the points i and j is possible, the shortest integration path, a straight line between i and j, is used here. Longer paths tend to produce larger depth difference errors because the gradient field is not guaranteed to be integrable.

Using these depth differences, it is possible to extend the global optimisation scheme introduced in Section 8.3 by adding an error term which minimises the squared distance between all N depth points:

$$e_z = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \frac{\left((\Delta z)^{ij} - (\Delta z)^{ij}_{\mathrm{surf}}\right)^2}{\|(u_i, v_i) - (u_j, v_j)\|_2} \tag{9.9}$$

The iterative update rule Eq. (8.16) then becomes

$$\begin{aligned} p_{n+1}(u,v) = \bar{p}_n(u,v) + \lambda \frac{\partial e_I}{\partial p} + \mu \frac{\partial e_\Phi}{\partial p} + \nu \frac{\partial e_D}{\partial p} \\ + 2\chi \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\frac{(\Delta z)^{ij} - (\Delta z)^{ij}_{\mathrm{surf}}}{\|(u_i, v_i) - (u_j, v_j)\|_2} \right] \left. \frac{\partial (\Delta z)^{ij}_{\mathrm{surf}}}{\partial p} \right|_{u,v} . \end{aligned} \tag{9.10}$$

An analogous expression is obtained for q. The derivatives of $(\Delta z)^{ij}_{\mathrm{surf}}$ with respect to p and q may only be nonzero if the pixel (u_k, v_k) belongs to the path C^{ij} and are zero otherwise. They are computed based on the discrete gradient field. The derivative depends on the direction (d_u, d_v) of the integration path at pixel location (u_k, v_k) with $d_u = u_{k+1} - u_k$ and $d_v = v_{k+1} - v_k$:

$$\begin{aligned} \left. \frac{\partial (\Delta z)^{ij}_{\mathrm{surf}}}{\partial p} \right|_{u_k, v_k} &= d_u p(u_k, v_k) \\ \left. \frac{\partial (\Delta z)^{ij}_{\mathrm{surf}}}{\partial q} \right|_{u_k, v_k} &= d_v q(u_k, v_k) \end{aligned} \tag{9.11}$$

The update of the surface gradient at location (u, v) is then normalised with the number of paths to which the corresponding pixel belongs. Error term (9.9) will lead to the evaluation of $N(N-1)/2$ lines at each update step and becomes prohibitively expensive for a large number of depth measurements. Therefore only a limited number of randomly chosen lines is used during each update step. Due to the discrete pixel grid, the width of each line can be assumed to be one. It is desirable that a large proportion of the image area is covered by the lines. For randomly distributed points and square images of size $w \times w$ pixels, we found that about 70 percent of all image pixels are covered by the lines when the number of lines corresponds to $10\,w$.

An earlier approach by Wöhler and Hafezi (2005) combines SfS and shadow analysis using a similar depth difference error term, which is, however, restricted to depth differences along the light source direction. The method proposed by Fassold et al. (2004) directly imposes depth constraints selectively on the sparse set of surface locations with known depth. As a consequence, in their framework the influence of the depth point on the reconstructed surface is restricted to its immediate local neighbourhood. Horovitz and Kiryati (2004) reduce this effect by applying a weighted least squares extension of depth from gradient-field computation and by adding an interpolation surface to the reconstructed shape. In their framework, the influence of the 3D points on the reconstructed surface is more well-behaved but still decreases considerably with increasing distance. In contrast, our method effectively transforms sparse depth data into dense depth difference data as long as a sufficiently large number of paths C^{ij} is taken into account. The influence of the depth error term is thus extended across a large number of pixels by establishing large-scale surface gradients based on depth differences between 3D points.

Beyond 3D points derived from image-based methods like stereo analysis or SfM, the framework described in this section can be applied to sparse depth data obtained from arbitrary sources, such as tactile measurement, laser scanners, or photogrammetric devices based on the projection of structured light.

Part III

Experimental investigations and evaluation

10 Structure from Motion and Defocus

10.1 Offline algorithm

In all described experiments we used a Baumer 1032×776 pixels industrial CCD camera with Cosmicar-Pentax video lenses. In order to validate our approach we first reconstructed a planar object with known ground truth. A chequerboard as shown in Fig. 10.1 with 10×8 squares of size 15×15 mm^2, respectively, was used. The 99 corners serve as features and are extracted in every image using the method described by Krüger et al. (2004) to assure sub-pixel accuracy. The true pose of the chequerboard is obtained according to Bouguet (1997) based on the given size of the squares. Note that Bouguet (1997) determines the true pose of the chequerboard by applying a least mean squares fit on a single image, whereas the proposed algorithm estimates the 3D structure of a scene by means of a least mean squares fit applied to the whole image sequence. Comparing the obtained results with the determined true pose of the object is actually a comparison between two methods conducting different least mean squares fits.

The deviation E_{rec} of the reconstructed 3D scene point coordinates X_i from the ground truth values X_i^{true} is given by

$$E_{\mathrm{rec}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_i - X_i^{\mathrm{true}}\right)^2}. \tag{10.1}$$

To determine an appropriate weight parameter α in Eq. (7.10) we computed E_{rec} for different α values in the range between 0 and 1. For $\alpha = 0$ the global minimisation is equivalent to Structure from Motion initialised with the calculated Depth from Defocus values. One must keep in mind, however, that the absolute scaling factor is then part of the gauge freedom of the bundle adjustment method, resulting in a corresponding “flatness” of the error function. Small α values lead to an instable convergence. The value of E_{rec} levels off to 16 mm for $\alpha \approx 0.3$ and obtains its minimum value of 7 mm for $\alpha = 0.42$. The root mean square deviation of the reconstructed size of the squares from the true value of 15 mm

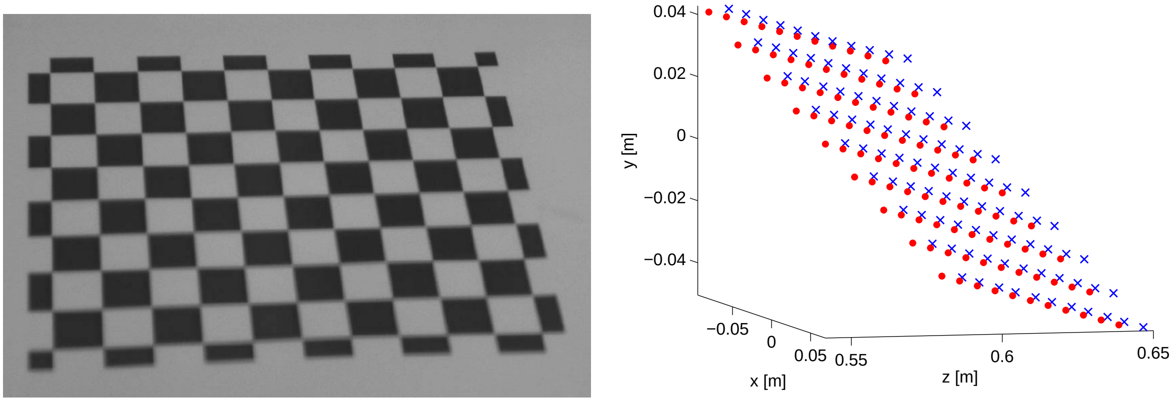


Figure 10.1: True (dots) and reconstructed (crosses) 3D pose of the chequerboard ($\alpha = 0.42$).

then amounts to 0.2 mm or 1.3%. The most accurate scene reconstruction results are obtained with α between 0.3 and 0.5. The reconstructed 3D scene points X_i for $\alpha = 0.42$ are illustrated in Fig. 10.1, the dependence of E_{rec} on α in Fig. 10.2 (top).

In addition to the reconstruction error E_{rec}, a further important error measure is the reprojection error

$$E_{\mathrm{repr}} = \sqrt{\frac{1}{MN} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\mathcal{P}\left(T_j X_i\right) - x_{ij}\right)^2} \tag{10.2}$$

denoting the root-mean-square deviation between the measured 2D feature positions x_{ij} and the reconstructed 3D scene points X_i reprojected into the images using the reconstructed camera transforms T_j.

The defocus error denotes the root-mean-square deviation between measured and expected radii σ_{ij} of the Gaussian PSFs according to

$$E_{\mathrm{def}} = \sqrt{\frac{1}{NM} \sum_{i=1}^{N} \sum_{j=1}^{M} \left(\mathcal{S}([T_j X_i]_z) - \sigma_{ij}\right)^2}. \tag{10.3}$$

Fig. 10.2 (bottom) shows the relation between the weight parameter α, the reprojection error E_{repr}, and the defocus error E_{def}. For $\alpha > 0.3$ the defocus error stabilises to 0.58 pixels per feature. Larger α values lead to a stronger influence of the Depth from Defocus values on the optimisation result, leading to an increasing reprojection error E_{repr} due to the inaccuracy of the estimated σ_{ij} values. Although the depth values derived by Depth

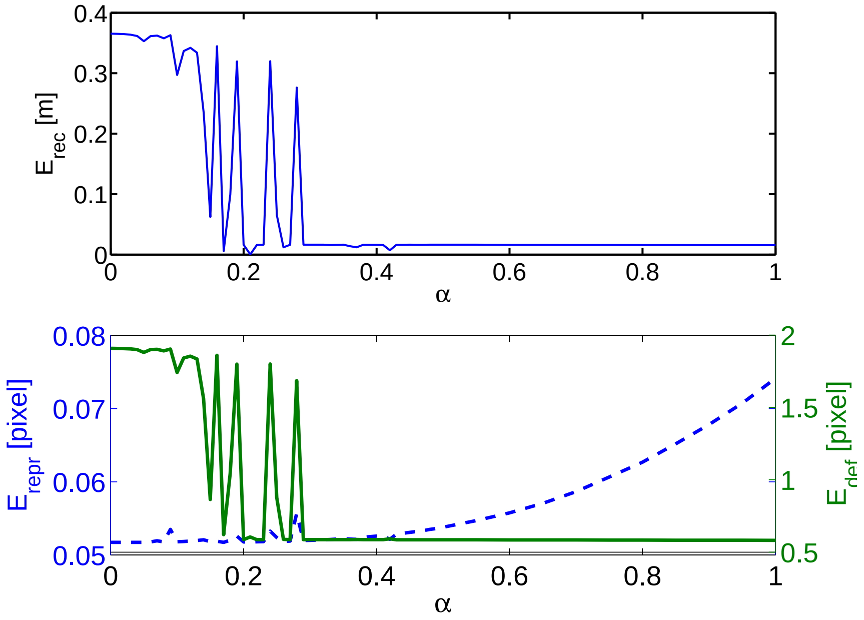


Figure 10.2: Dependence of E_{rec} (upper diagram), E_{repr} (lower diagram, dashed curve, left axis), and E_{def} (lower diagram, solid curve, right axis) on the weight parameter α.

from Defocus are noisy, they are sufficient to establish a reasonably accurate absolute scale. Hence, this first evaluation shows that the combined approach is able to reconstruct scenes at absolute scale without prior knowledge. Our approach is favourably used for small scale objects (bounding cube width ~ 0.5 m) as long as standard video cameras and lenses (f below ~ 20 mm, pixel size $\sim 5\mu$m, image size $\sim 10^6$ pixels) are used. The focal length required to obtain a comparable relative accuracy with respect to the object size is proportional to d, implying a narrow field of view for increasing depth d. Depth from Defocus is thus favourably applied in the close range domain.

Further experiments performed on real-world objects are described in the following paragraphs. Images from the beginning, the middle, and the end of the corresponding sequences, respectively, are shown in Fig. 10.3. In order to separate random fluctuations from systematic deviations, we computed the averages and standard deviations of the error measures over 100 runs for each example, respectively. For this purpose, we determined the pixel noise level of the employed CCD camera. The noise leads to a standard error of the feature positions x_{ij} obtained by the KLT tracker of 0.1 pixels. For each of the 100 runs, we added a corresponding amount of Gaussian noise to the images of the sequence.

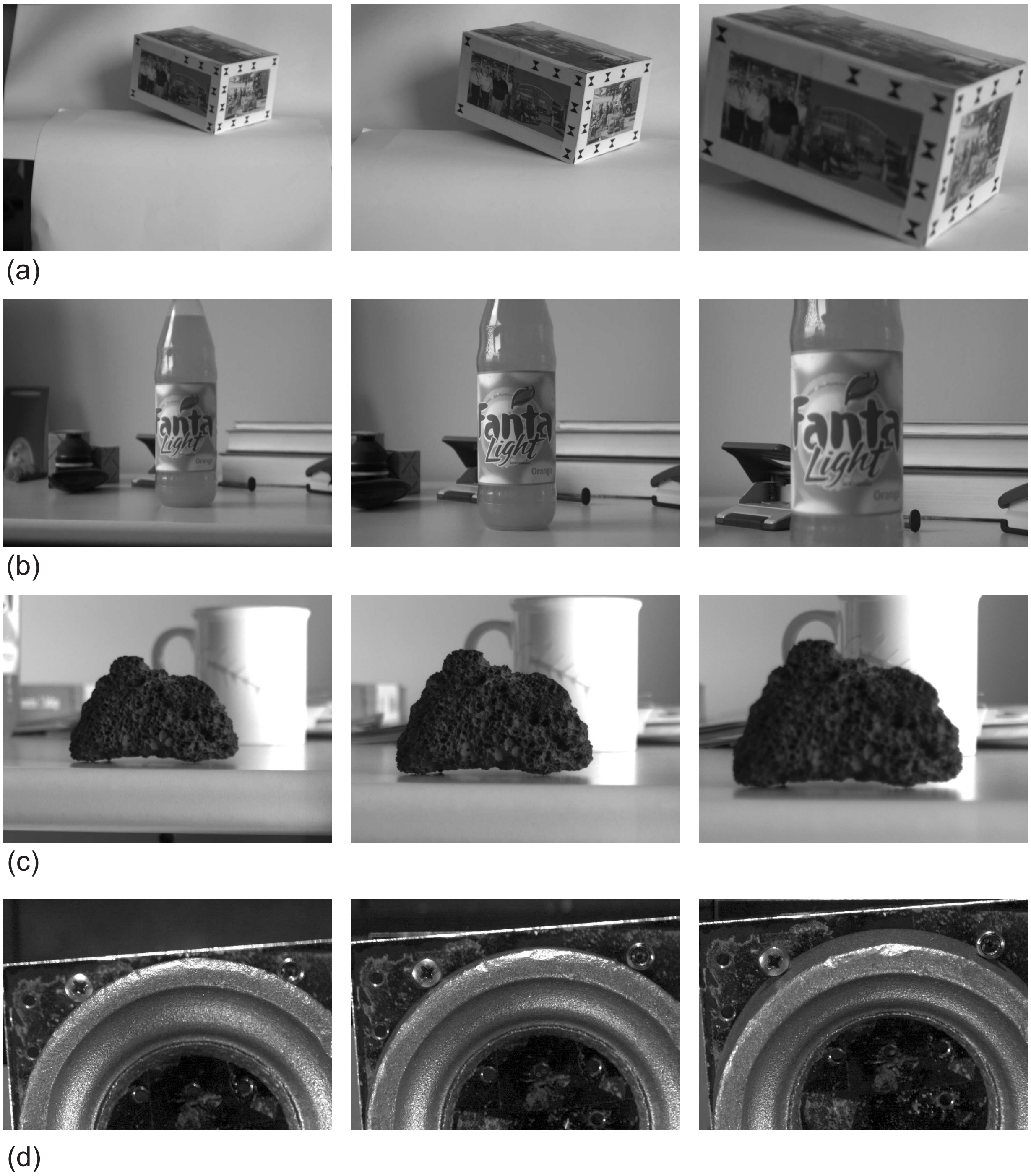


Figure 10.3: Images from the beginning, the middle, and the end of (a) the cuboid sequence, (b) the bottle sequence, (c) the lava stone sequence, and (d) the flange sequence.

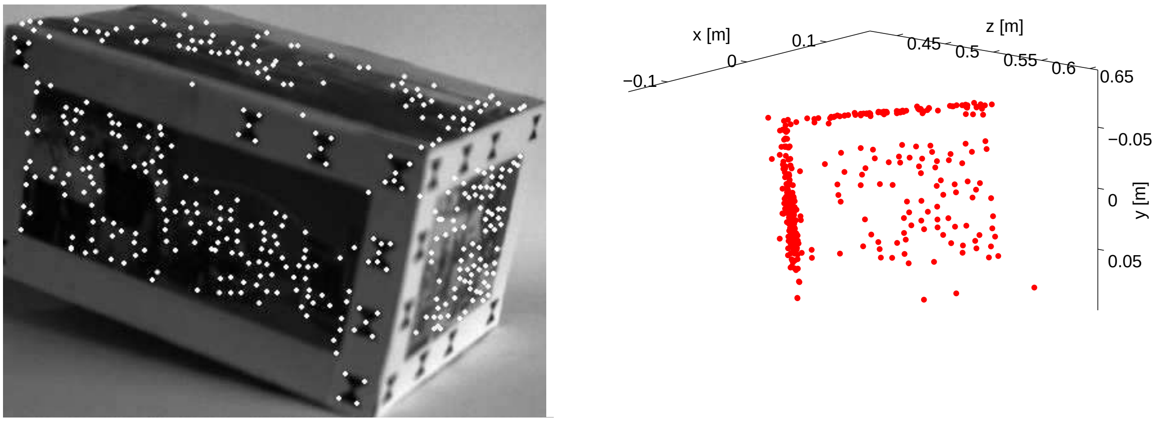


Figure 10.4: 3D reconstruction of the cuboid.

10.1.1 Cuboid sequence

To demonstrate the performance of our approach on a non-planar test object of known dimensions we applied our method to the cuboid-shaped object shown in Fig. 10.3a. This object displays a sufficient amount of texture to generate "good features to track" (Shi and Tomasi 1994). In addition, black markers on white background with known mutual distances are placed near the edges of the cuboid. As described in Section 7.2, feature points are extracted and tracked using the KLT algorithm, and the 3D coordinates of the scene points are obtained by minimising the error term E_{comb} according to Eq. (7.10) with $\alpha = 0.5$ as the weight parameter. This value of α will be used in all subsequent experiments. Tracking outliers are removed by determining the features with associated very large reprojection errors of more than $3E_{\mathrm{repr}}$ and neglecting them in a subsequent second bundle adjustment step.

The 3D reconstruction result for the cuboid sequence is shown in Fig. 10.4. We obtain for the reprojection error $E_{\mathrm{repr}} = 0.642 \pm 0.002$ pixels and for the defocus error $E_{\mathrm{def}} = 0.64 \pm 0.07$ pixels. In order to verify the absolute scale, we compared the reconstructed pairwise distances between the black markers on the object (as seen e.g. in the top right corner of the front side) to the corresponding true distances. For this comparison we utilised a set of six pairs of markers with an average true distance of 32.0 mm. The corresponding reconstructed average distance amounts to 34.1 ± 1.6 mm (cf. Table 10.1.4).

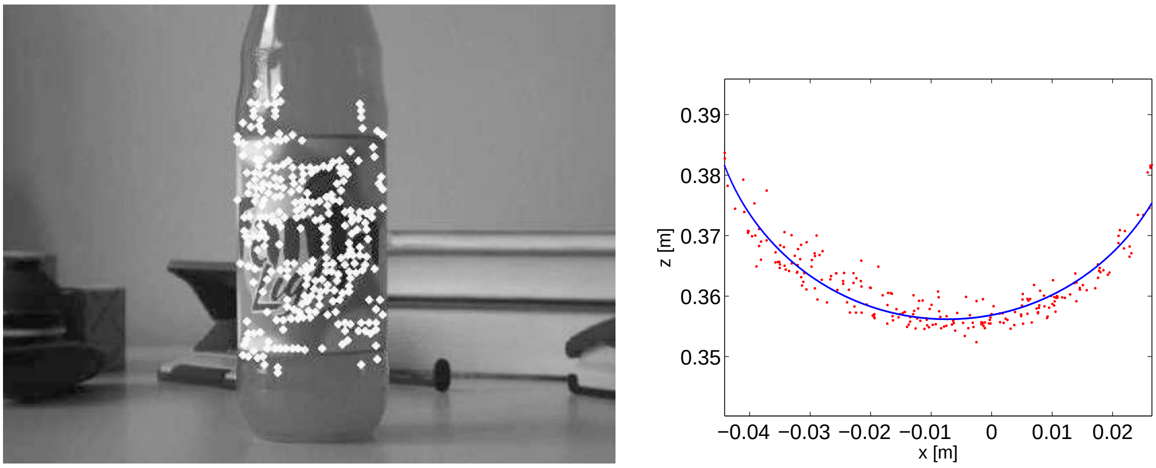


Figure 10.5: 3D reconstruction of the cylindrical surface of the bottle.

10.1.2 Bottle sequence

In order to demonstrate the performance of our approach on a real-world object, we applied it to a bottle as shown in Fig. 10.3b. No background features are selected by the algorithm since none of these feature obtains its maximum sharpness in the acquired sequence. The 3D reconstruction result is shown in Fig. 10.5. We obtained for the reprojection error $E_{\mathrm{repr}} = 0.75 \pm 0.12$ pixels and $E_{\mathrm{def}} = 0.39 \pm 0.02$ pixels. To quantify the accuracy of the determined absolute scale, we compared the diameter of the reconstructed object with that of the real bottle. For this purpose we projected the reconstructed points into the xz plane of the camera coordinate system, in which the x axis is parallel to the image rows, the y axis is parallel to the image columns (and thus to the central axis of the cylinder), and the z axis is parallel to the optical axis. The circle fit to the projected 3D points as shown in Fig. 10.5 yields a diameter of 82.8 ± 1.4 mm (cf. Table 10.1.4).

10.1.3 Lava stone sequence

As a further real-world object, we examined the lava stone shown in Fig. 10.3c. The 3D reconstruction result is shown in Fig. 10.6. The shaded view of the object was obtained by triangulation of the reconstructed set of 3D points. The cusp visible in the left part of the reconstructed surface is due to three outlier 3D points generated by inaccurately determined feature positions. For this scene we have $E_{\mathrm{repr}} = 0.357 \pm 0.002$ pixels and $E_{\mathrm{def}} = 0.174 \pm 0.005$ pixels. The very low standard deviations indicate a high consistency of the DfD measurements due to the strong texture of the object surface. Two points on

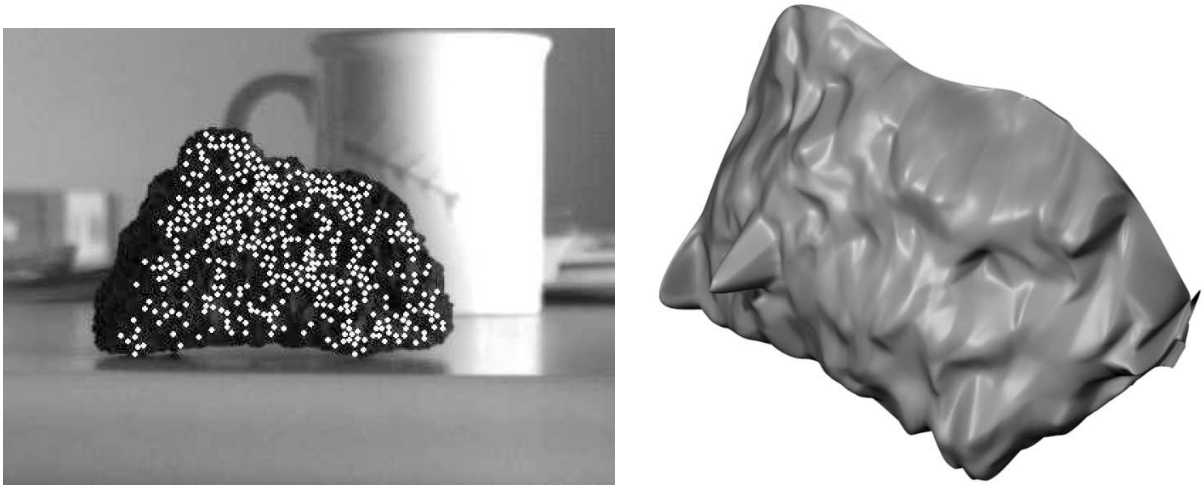


Figure 10.6: 3D reconstruction of the lava stone. The cusp visible in the left part of the reconstructed surface is produced by three outlier points.

the object with a true mutual distance of 60.0 mm were chosen as reference locations for estimation of the accuracy of the determined absolute scale. The reconstructed distance of the reference points amounts to 58.3 ± 0.78 mm (cf. Table 10.1.4).

10.1.4 Flange sequence: Raw cast iron surface

Another experimental evaluation of the offline algorithm addresses an industrial quality inspection scenario. We regard the 3D reconstruction of the raw cast iron surface of a flange. In this setting, the metal part is attached to a goniometer and can thus be rotated around two axes, while the camera is fixed. In an industrial inspection system, it would probably be more favourable to mount the camera on an industrial robot such that it can be moved with respect to a fixed part to be inspected. Three images of the acquired sequence are shown in Fig. 10.3d. Although the surface is rough, the extracted set of 3D points is rather sparse, which is due to specular reflections changing across time, leading to premature termination of tracks by the KLT tracker. The extracted set of depth points shows that the reconstructed surface region is essentially flat and inclined with respect to the image plane (Fig. 10.7a). We fitted a plane to the set of 3D points and determined a RMS distance of the 3D points from this reference plane of 1.46 mm. To examine the absolute scale of the reconstructed scene, we regarded two feature points situated at well-defined locations on the edge of small deformations of the surface (marked as 1 and 2 in Fig. 10.7b). According to our 3D reconstruction result, the mutual distance of the

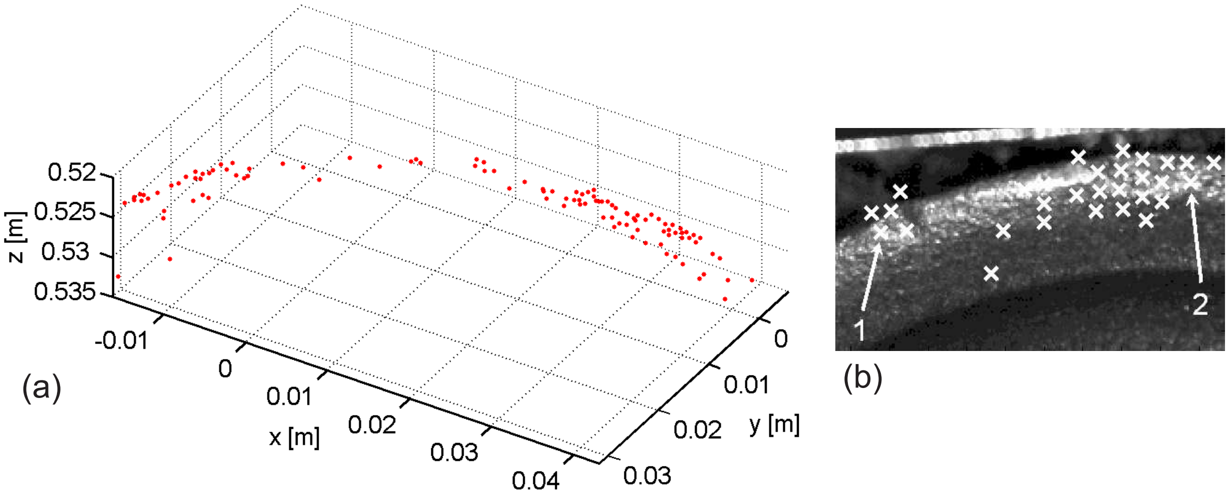


Figure 10.7: 3D reconstruction of the raw cast iron surface of a flange. (a) Reconstructed 3D points. (b) Location of the reference points used to determine the accuracy of the estimated absolute scale.

corresponding two 3D points amounts to 15.45 ± 0.01 mm. The true distance, determined by tactile measurement, is 15.2 mm. Being too large by the small amount of 1.6%, the estimated absolute scale is in good agreement with the true value.

In contrast to the previous examples, the ROIs around the extracted feature positions show strong small-scale intensity variations, such that the noise added to the ROIs has a negligible effect on the DfD result. On the other hand, the orientation of the surface with respect to the light source changes across the sequence, leading to an appearance of the ROIs that changes systematically over time. These variations are not of geometrical nature but due to the strongly non-Lambertian reflectance behaviour of the surface. Hence, they cannot be taken into account by the KLT tracker, which therefore tends to produce a slight drift of the feature positions relative to the object across the image sequence. At the same time the variations have a systematic influence on the amount H of defocus determined according to Eq. (7.9) and are presumably the main reason for the observed small discrepancy of 1.6% between the estimated and the true absolute scale of the scene.

The extracted set of 3D points is too sparse to reveal the small-scale deformations visible in the images. Hence, we used our sparse set of depth points shown in Fig. 10.7a as an input to the Shape from Photopolarimetric Reflectance and Depth method introduced in chapter 9. These results are presented in chapter 11.

Table 10.1: Summary of the evaluation results for the offline algorithm.

Sequence	Length	E_{repr} [pixels]	E_{def} [pixels]	Reference length [mm]	
				Ground truth	3D reconstruction
Cuboid	46	0.642 ± 0.002	0.636 ± 0.067	32.0	34.1 ± 1.6
Bottle	26	0.747 ± 0.121	0.387 ± 0.020	80.0	82.8 ± 1.4
Lava stone	15	0.357 ± 0.002	0.174 ± 0.005	60.0	58.3 ± 0.8
Flange	36	1.06 ± 0.01	1.96 ± 0.01	15.2	15.45 ± 0.01

10.2 Online algorithm

A systematic evaluation of the online algorithm was performed for the cuboid, bottle, and lava stone sequence. In the surface inspection scenario (flange sequence) we only employed the offline algorithm since here no knowledge about the object structure is necessary before the final dense reconstruction step performed after termination of image acquisition. The online algorithm generally starts with a very noisy set of 3D points, due to the small number of features already having reached their maximum sharpness at the beginning of the image sequence. After processing more and more images of the sequence, the 3D reconstruction result resembles the result of the offline algorithm. The results are not identical because generally a similar but not identical index f_i (cf. Eq. (7.6)) is determined for the sharpest ROI of each track by the offline and the online algorithm, respectively. The DfD results then differ correspondingly.

In Figs. 10.9–10.11, the standard deviations across the 100 online runs are indicated by error bars.

In the cuboid example shown in Fig. 10.8, the first 3D reconstruction result can be obtained after processing 21 images. After 5 iterations, still only 40 features are available, having passed their point of maximum sharpness, respectively. In this example, the same six pairs of reference points as those regarded in Section 10.1 were used for evaluating the accuracy of the determined absolute scale. Fig. 10.9 shows the behaviour of the reconstruction accuracy with increasing number of iterations, averaged over 100 online runs carried out after adding Gaussian noise to the images as described in Section 10.1. After 43 processed images, the measured average mutual distance of the six pairs of reference points differs by less than 1.8% from the true value, and this difference is smaller than 1.2 standard deviations. For less than 38 processed images, not all reference points have passed their point of maximum sharpness, such that their 3D positions have not yet been computed. The average relative scale error shown in Fig. 10.9 is derived from those pairs of reference

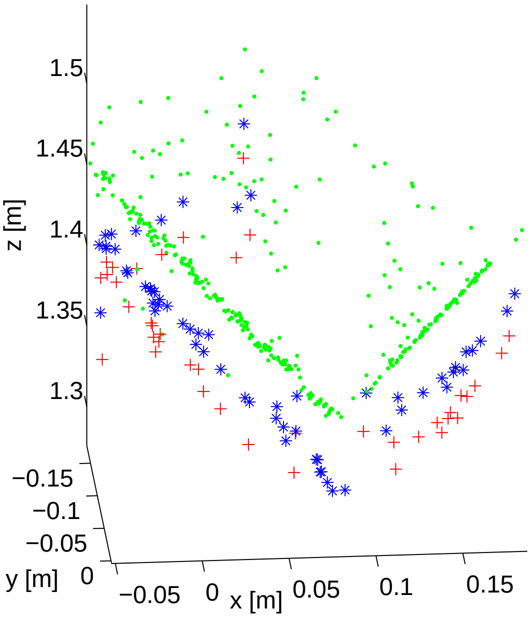


Figure 10.8: 3D reconstruction of the cuboid, obtained with the online algorithm. The 3D reconstruction result is displayed after 5 (crosses), 9 (stars), and 25 (dots) iterations. The first iteration is performed after the 21st image.

points that already have passed their point of maximum sharpness.

Fig. 10.8 suggests that with increasing number of available features the reconstructed size and shape of the cuboid become more accurate. However, Fig. 10.9 shows that the maximum accuracy of the inferred absolute scale is already obtained after 44 or 45 processed images. The reason is that the very last images of the sequence are strongly blurred. This leads to large inaccuracies of the depth values derived from the estimated PSF radii σ_{ij}, since far away from its minimum, the Depth-Defocus-Function $\mathcal{S}(d)$ according to Eq. (7.5) displays a low slope. Furthermore, systematic deviations tend to arise since the observed behaviour $\sigma(d)$ of the PSF radius is best represented for small and intermediate values of σ by the analytic form chosen in Eq. (7.5) for the Depth-Defocus-Function $\mathcal{S}(d)$ (cf. Section 10.3). Analogous experimental evaluations were conducted for the bottle sequence and the lava stone sequence. The results are shown in Fig. 10.10. For the bottle sequence, the average accuracy of the determined absolute scale (represented by the inferred diameter of the bottle as outlined in Section 10.1) is better than 3.0% already after 12 processed images. However, at the beginning of the sequence the random scatter across the 100 runs is about two times larger than at the end (more than 23 processed images). The final

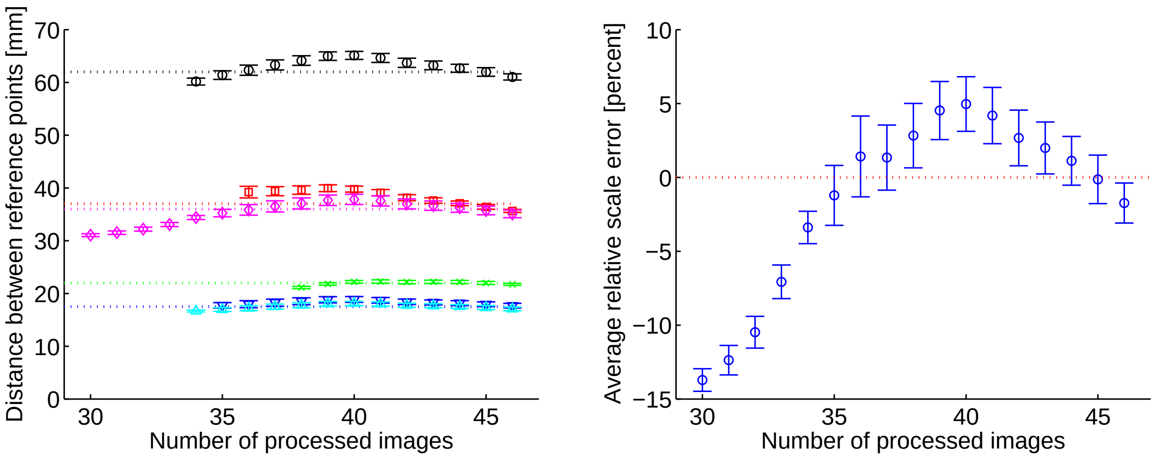


Figure 10.9: Left: Behaviour of the distances between six pairs of reference points on the cuboid surface, obtained with the online algorithm, for increasing number of iterations, compared to the corresponding true values. The standard deviations across the 100 online runs are indicated by error bars. Right: Relative accuracy of the inferred absolute scale, given by the average relative deviation between the measured and true absolute distances between those pairs of reference points already having passed their point of maximum sharpness, for increasing number of iterations.

difference between measured and true absolute scale corresponds to 1.9 standard deviations. For the lava stone sequence, the determined absolute scale is about 1.2% (corresponding to one standard deviation) too small after 12 processed images. The deviation becomes larger and appears to be of systematic nature when 13 and more images are processed. The last three images of the lava stone sequence are strongly blurred, which we assume to be the main reason for this behaviour (cf. Section 10.3). For this sequence it is favourable to adopt the 3D reconstruction result obtained after processing 12 images and to avoid utilising the last three, strongly blurred, images.

10.3 Analysis of random errors and systematic deviations

The main source of random errors is the pixel noise of the CCD sensor, which influences the estimation of the PSF radius according to the Depth-Defocus-Function (7.5) and furthermore leads to a random scatter of the extracted feature positions of the order 0.1 pixels. The 3D reconstruction results obtained with the offline algorithm for 100 runs over the cuboid, bottle, and lava stone sequences, respectively, show that the relative differences

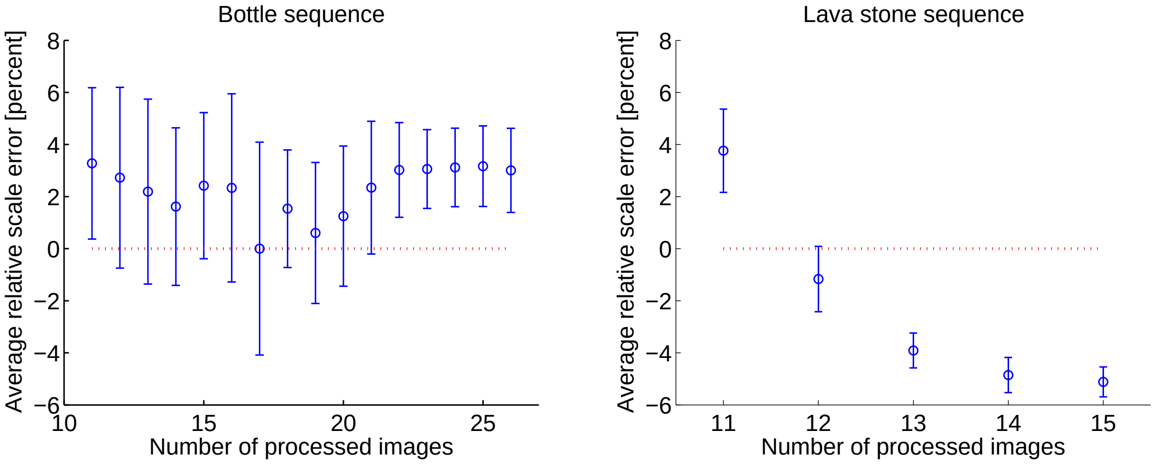


Figure 10.10: Relative deviation between measured and true absolute scale for increasing number of processed images for the bottle sequence (left) and the lava stone sequence (right).

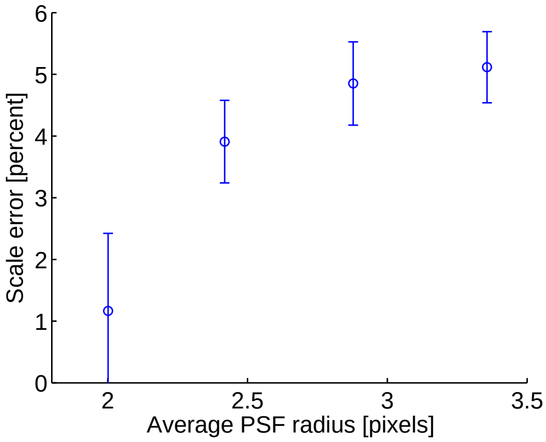


Figure 10.11: Correlation between scale error and average PSF radius for the last 4 images of the lava stone sequence.

between the ground truth and the reconstructed absolute scale of the scene amount to a few percent and always correspond to between 1 and 2 standard deviations (cf. Table 1). Hence, the observed deviations are presumably due to a combination of random fluctuations and systematic errors.

Systematic errors may be introduced at the end of the sequence, where the images tend to be strongly blurred. For PSF radii smaller than 5 pixels we found that the random scatter of the feature positions extracted by the KLT tracker are of the order 0.1 pixels and independent of the PSF radius, such that the extracted feature positions do not introduce systematic errors. However, the observed relation $\sigma(d)$ between PSF radius and depth is accurately represented only for small and intermediate values of σ by the Depth-Defocus-Function $\mathcal{S}(d)$ according to Eq. (7.5), while the fit becomes less accurate for larger values of σ. Fig. 7.4 illustrates that both examined lenses show this effect for values of σ between 2 and 3 pixels on both sides of the minimum of $\mathcal{S}(d)$. For the lava stone sequence processed with the online algorithm, Fig. 10.11 illustrates the correlation between scale error and average PSF radius of the last processed image. The systematic effect is especially pronounced for this sequence since it comprises only 15 images (cf. Table 1). Measurement errors obtained while processing the last three images, which are strongly blurred with $\sigma > 2$ pixels, thus have a substantial effect on the 3D reconstruction result. These findings suggest that images with an average PSF radius larger than about 2 pixels should be excluded from the 3D reconstruction process.

A further important source of systematic errors is the thermal expansion of the optical system. The body of the lens used for our experiments consists of Aluminium, having a relative thermal expansion coefficient of $\nu = 2.3 \times 10^{-5}\ \mathrm{K}^{-1}$. We assume that with a lens of focal length f at calibration temperature T_0 an image of maximum sharpness is observed at depth d_0. In our approximate treatment, the focal length f is assumed to be constant. The corresponding image distance v_0 is obtained according to the lens law (7.1). Assuming that the measurement is performed at temperature T, the thermal expansion of the lens body yields an image distance $v(T) = [1 + \nu(T - T_0)]v_0$, and the corresponding depth $d(T)$ for which an image of maximum sharpness is observed at temperature T is computed according to the lens law (7.1). As a result, the Depth-Defocus-Function (7.5) is shifted by the amount $d(T) - d_0$ along the d axis (cf. Fig. 7.4), which introduces a corresponding systematic error. We find that for a given temperature difference $|T - T_0| \ll T_0$, the relative systematic deviation $[d(T) - d_0]/d_0$ of the DfD measurement is largely proportional to d_0, and for a given value of $d_0 \gg f$, it is largely proportional to $|T - T_0|$. As an example, for a focal length $f = 20$ mm, a depth $d_0 = 1000$ mm, and a temperature difference $|T - T_0| = 10$ K, we obtain a relative systematic deviation of the DfD measurements of 1.1%.

Further possible sources of systematic deviations are vibrations and shocks occurring after calibration (which we avoided during our experiments) and systematic variations of the appearance of the extracted ROIs across the image sequence especially for specular surfaces (cf. Section 10.1.4). In general, such influences are difficult to quantify, but they may lead to systematic errors of at least the same order of magnitude as those inferred for thermal expansion.

11 Shape from Photopolarimetric Reflectance

11.1 Synthetic examples

To examine the accuracy of the 3D surface reconstruction methods described in Sections 8 and 9 in comparison to ground truth data and to reveal possible systematic errors, we test our algorithms on synthetically generated surfaces. To examine the behaviour of the local and global optimisation schemes described in Sections 8.3 and 8.4 and their combination with sparse depth data as described in Section 9.2.2, dependent on how many images based on which reflectance and polarisation features are used, we apply the developed algorithms to the synthetically generated surface shown in Fig. 11.1a. We assume a perpendicular view on the surface along the z axis, corresponding to $\vec{v} = (0, 0, 1)^T$. The scene is illuminated by $L = 2$ light sources (one after the other) under an angle of 15° with respect to the horizontal plane at azimuth angles of $\psi^{(1)} = -30°$ and $\psi^{(2)} = +30°$, respectively. This setting results in identical phase angles $\alpha^{(1)} = \alpha^{(2)} = 75°$ for the two light sources. We extracted a set of 500 random points from the ground truth surface, which were used as sparse depth data for 3D reconstruction.

The reflectance functions of the rough metallic surface measured according to Section 8.2 were used to render the synthetic images shown in Fig. 11.1c. The reconstruction was performed with synthetic noisy data, where we applied Gaussian noise with a standard deviation of 5×10^{-4} for the intensity I, where the maximum grey value is $\sim 6 \times 10^{-2}$, 1° for the polarisation angle Φ, and 0.4 pixels for the depth values (Z between 0 and 6 pixels). The weights for the error terms according to Eq. (9.10) were set to $\lambda = 450$, $\mu = 40$, $\nu = 100$, and $\chi = 1$. The surface gradients were initialised with zero values.

Fig. 11.1 shows the reconstruction results on noisy synthetic images, where the plots (d)–(f), (j), and (k) were obtained by applying SfPR alone, while the plots (g)–(i) depict the results obtained based on integration of sparse depth data into the SfPR framework. The respective reconstruction errors are given in Table 11.1. It is apparent that the Shape from Shading reconstruction using a single light source fails to reconstruct the surface (Fig. 11.1d), while the surface shape can be reconstructed approximately using a single

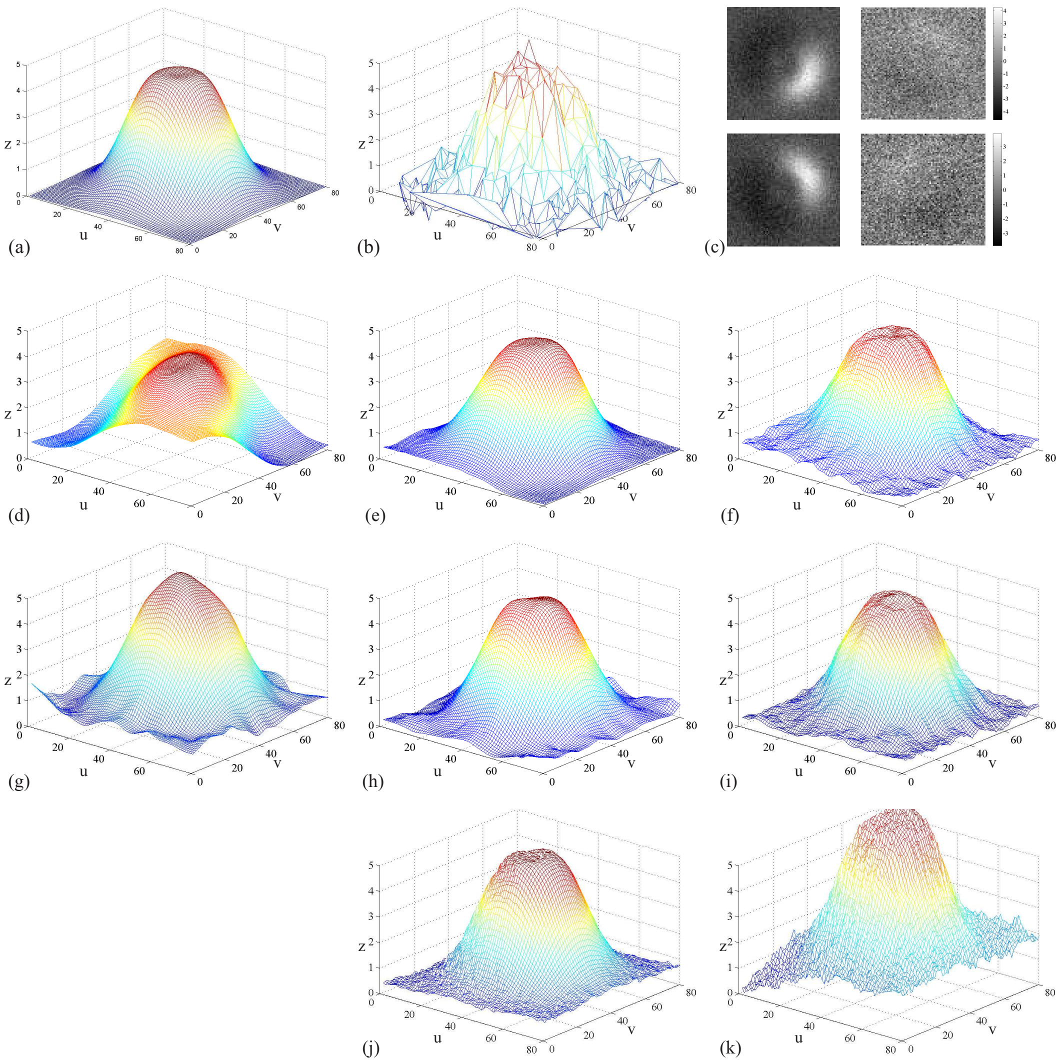


Figure 11.1: 3D reconstruction of a synthetically generated surface. (a) Ground truth. (b) Noisy depth data. (c) From the left: noisy intensity and polarisation angle images, based on measured reflectance functions of a raw forged iron surface. The reconstruction result for noisy images of a surface with uniform albedo, obtained by SfPR with global optimisation without integration of sparse depth data, is shown in (d) using a single intensity image, in (e) using both intensity images, and in (f) using one intensity and one polarisation angle image. (g) Reconstructed surface obtained based on noisy sparse depth data alone. (h) Reconstruction result using sparse depth data and intensity. (i) Reconstruction result using sparse depth data, intensity, and polarisation angle. For comparison, the reconstruction result obtained based on SfPR with local optimisation and without sparse depth data is shown in (j) using both intensity images and in (k) using one intensity and one polarisation angle image.

intensity and polarisation angle image (Fig. 11.1f). To reach a similar reconstruction accuracy without polarisation information, illumination from two different directions is required (Fig. 11.1e). Table 1 illustrates that using intensity and polarisation degree in the 3D reconstruction process leads to poor accuracy both for the global and the local approach, while using intensity and polarisation angle yields a high accuracy which does not further increase when the polarisation degree is additionally used. The reason for this behaviour is the fact that intensity and polarisation degree contain somewhat redundant information, since both display a maximum in the specular direction ($\theta_r = 0°$) and decrease in a qualitatively similar lobe-shaped manner for increasing value of θ_r. The dependence on surface orientation, however, is much stronger for the intensity than for the polarisation degree, while the measurement error tends to be significantly lower for the reflectance. The local optimisation approach according to Section 8.4 provides a very accurate reconstruction for the noise-free case, but performs worse than the global approach on noisy data (Figs. 11.1j and 11.1k). This property can be observed clearly by comparing the corresponding reconstruction errors of p and q given in Table 11.1. With both reflectance and polarisation angle images, the reconstruction result becomes very accurate. Similarly accurate reconstruction results, however, are already obtained based on a single intensity and polarisation image.

Fig. 11.1g shows the reconstruction result using only the sparse depth values, effectively smoothing and interpolating the sparse depth values shown in Fig. 11.1b. The overall shape is correct, but smaller details like the flattened top of the object are missing in the reconstructed 3D profile. Adding intensity and polarisation terms improves the results and captures the finer details which are not visible in the sparse depth data (Figs. 11.1h and 11.1i).

11.1.1 Choosing the weight parameters

The values for the weight parameters of the error terms according to Eqs. (4.8) and (9.10) are related to the magnitudes of the intensity and polarisation features and their measurement uncertainties. We have evaluated the influence of the weight parameters on the reconstruction accuracy using the previously described synthetic data. As a typical example, Fig. 11.2 shows the root mean square depth error of the reconstructed surface profile obtained from one intensity and one polarisation angle image for different weight parameters λ and μ. For noise-free image data, the reconstruction error decreases with increasing λ and μ until the algorithm starts to diverge at fairly well-defined critical values. For noisy

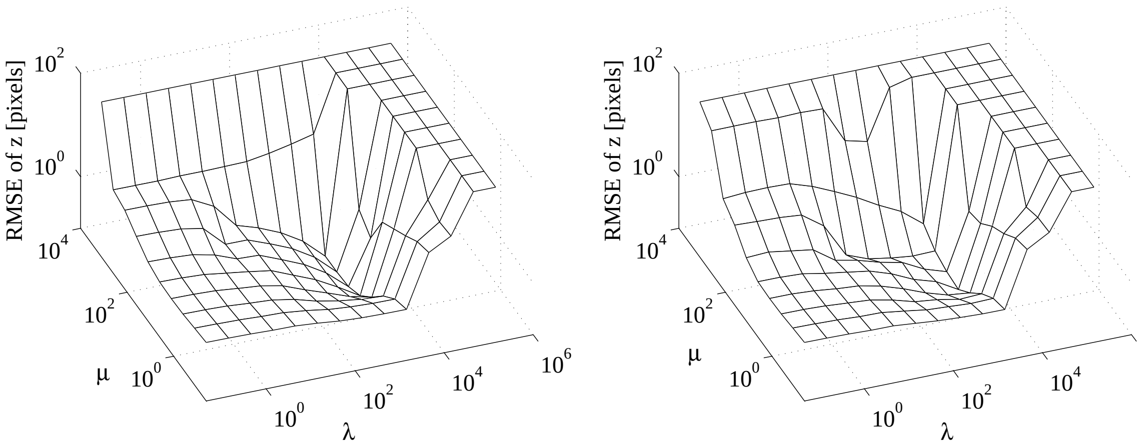


Figure 11.2: Reconstruction error of SfPR with over weight parameters. Left: noise free case, Right: noisy images.

input images (cf. Fig. 11.1c) the reconstruction error displays a weaker dependence on λ and μ and a less pronounced minimum. This is a favourable property since small changes in the weight parameters do not lead to large differences in the reconstruction accuracy as long as the values chosen for λ and μ are well below their critical values for which the algorithm begins to diverge.

11.2 Real-world examples

As a first real-world example, we apply the surface reconstruction algorithms described in Section 9 to the raw forged iron surface of an engine part. In this section we describe the obtained 3D reconstruction results and compare them to a ground truth cross-section of the same surface, measured with a scanning laser focus profilometer.

The second presented real-world example is a slightly damaged section of the raw surface of a flange also consisting of forged iron. The surface shows several small deformations. We compare the depth of these deformations inferred from our 3D reconstruction to ground truth values obtained by tactile measurement.

For the first two examples, we utilise a convergent stereo setup consisting of two CCD cameras of 1032 × 776 pixels image resolution, equipped with lenses of 25 mm focal length. The base distance of the cameras amounts to 320 mm and the average distance to the object is 450 mm. The resulting field of view corresponds to 10°. The size of the image sections

Table 11.1: 3D reconstruction results for the synthetic ground truth surface shown in Fig. 11.1a. If not identified otherwise, the results were obtained with the global optimisation approach. The error values for z are given in pixels.

Method, utilised information	Figure	RMS error (without noise)			RMS error (with noise)		
		z	p	q	z	p	q
I_1	11.1d	1.11	0.046	0.077	1.11	0.047	0.077
ϕ_1	–	2.13	0.102	0.059	3.92	0.163	0.117
I_1, I_2	11.1e	0.22	0.012	0.018	0.21	0.014	0.019
I_1, I_2 (local)	11.1j	0.00	0.000	0.000	0.19	0.046	0.088
I_1, ϕ_1	11.1f	0.17	0.012	0.007	0.19	0.040	0.065
I_1, ϕ_1 (local)	11.1k	0.00	0.000	0.000	0.28	0.134	0.210
I_1, D_1	–	1.11	0.044	0.077	1.13	0.098	0.102
I_1, D_1 (local)	–	2.12	0.088	0.178	7.17	0.837	0.928
I_1, ϕ_1, D_1	–	0.01	0.001	0.001	0.52	0.103	0.088
I_1, ϕ_1, D_1 (local)	–	0.00	0.000	0.000	0.31	0.149	0.214
I_1, I_2, ϕ_1	–	0.11	0.009	0.005	0.20	0.034	0.066
I_1, I_2, ϕ_1 (local)	–	0.00	0.000	0.000	0.35	0.074	0.099
I_1, I_2, ϕ_1, ϕ_2	–	0.01	0.001	0.001	0.21	0.056	0.078
I_1, I_2, ϕ_1, ϕ_2 (local)	–	0.00	0.000	0.000	0.24	0.057	0.079
Z	11.1g	0.14	0.012	0.013	0.20	0.034	0.036
I_1, Z	11.1h	0.11	0.008	0.009	0.15	0.023	0.028
I_1, ϕ_1, Z	11.1i	0.09	0.006	0.005	0.13	0.042	0.065
I_1, I_2, ϕ_1, Z	–	0.09	0.006	0.005	0.15	0.036	0.062
I_1, I_2, ϕ_1, ϕ_2, Z	–	0.07	0.004	0.003	0.11	0.052	0.078

used for 3D reconstruction is 240 × 240 pixels. The surface is illuminated by one or two LED point light sources. Only I, Φ, and Z are used for 3D reconstruction since in addition to the fact that intensity and polarisation degree essentially provide redundant information as pointed out in Section 11.1, for the regarded rough metallic surfaces the behaviour of D is strongly affected by small-scale variations of the surface roughness. Accordingly, the value of D for specular reflections varies across the surface by up to 20 percent. Hence, the polarisation degree does not represent a useful feature for 3D reconstruction in this application context.

This unfavourable behaviour of the polarisation degree is known from previous work in the domain of photopolarimetry. A modified Fresnel equation for the polarisation degree as a function of incidence angle is derived by Morel et al. (2005) for smooth, specularly reflecting metallic surfaces based on the assumption that the absolute value of the complex diffraction index of the surface material is much larger than 1. However, Germer et al. (2000) demonstrate that the polarisation degree strongly depends on the microscopic surface roughness even for smooth, polished and etched steel surfaces. Their measurements cannot be explained by a simple physical model, but it is necessary to take into account microroughness and subsurface scattering effects. The experimental results by Germer et al. (2000) give an impression of the difficulties encountered when attempting to compute the polarisation degree of light reflected from a metallic surface based on physical models. Based on our experiments regarding raw forged iron materials, however, we found that in contrast to the polarisation degree, the polarisation angle is not perceivably influenced by slight variations of the surface roughness. As a consequence, the polarisation degree is a feature which is useful for determination of the surface orientation only for smooth dielectric surfaces, which can be accurately described in terms of the Fresnel equations (Atkinson and Hancock 2005).

Our third real-world example deals with the 3D reconstruction of a section of the lunar surface based on a short sequence of images acquired by the SMART-1 spacecraft. Since no polarisation data but only intensity images are available, it is not possible in this example to apply the full photopolarimetrc framework. However, in contrast to the fully controlled industrial quality inspection scenarios we have a semi-controlled setting since the position and motion of the camera relative to the surface are only very approximately known, while the direction of illumination and the surface reflectance behaviour are known. Based on this example we demonstrate the usefulness of our approach to combine SfS information with sparse depth data obtained by means of SfM to compute a digital elevation map of the surface section. This technique results in a dense depth map of the surface, representing a self-consistent solution to the available point correspondences between the images of the sequence and the photometric (pixel brightness) information.

11.2.1 Application of the SfPR technique

11.2.1.1 Forged iron surface

For 3D surface reconstruction of the raw forged iron surface of the engine part with the local SfPR approach according to Section 8.4 we employed two intensity images and one polarisation angle image. Fig. 11.3a shows a flawless part and a part that displays a surface deformation. We utilised the quotient-based and thus albedo-independent intensity error term according to Eq. (8.20). The deviation between the flawless and the deformed surface becomes evident in Figs. 11.3b and 11.3c. The comparison between the ground truth and the corresponding cross-section extracted from the reconstructed profile yields a root mean square error (RMSE) of 220 μm.

We performed two experiments concerning the application of the global SfPR approach to the engine part surface. In the first experiment, we initialised the surface gradients by zero values and determined the uniform surface albedo ρ_0 according to Eq. (8.18), relying on specular reflections. Cross-sections extracted from the corresponding reconstructed surface profiles and their comparison to ground truth are shown in Fig. 11.4e. The RMSE values are 46 μm for the SfPR approach and 287 μm for the SfS approach which neglects polarisation information. While the SfPR approach yields a very accurate reconstruction of the surface, a uniform value of the surface gradient perpendicular to the direction of incident light is estimated by the SfS approach due the minor influence of this gradient on the error function.

In the second experiment, we initialised the global optimisation scheme with the surface gradients $p_{\mathrm{DfD}}(u,v)$ and $q_{\mathrm{DfD}}(u,v)$ inferred from the DfD result as described in Section 9.1.2. To calibrate the DfD algorithm, we fitted a linear function to the $(\sigma^{-1},(z-z_0))$ data points measured over a range of depth values of about 13 mm (cf. Section 9.1.1). The calibration curve is shown in Fig. 9.2. For illustration purposes, the very noisy raw DfD measurements are shown in Fig. 11.5b. The standard error of the depth values obtained by DfD corresponds to 1.2 mm, with an average object distance of 450 mm. Although this corresponds to an absolute depth error of only 0.3%, due to the relatively small depth

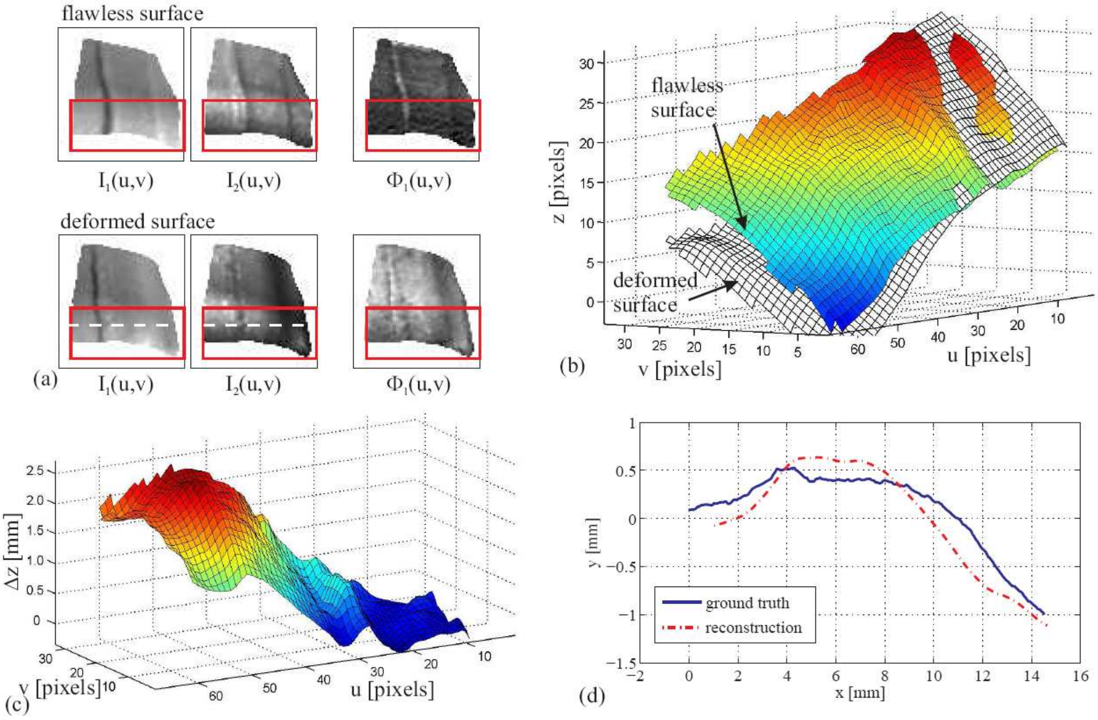


Figure 11.3: Application of the local SfPR algorithm to the raw forged iron surface of the engine part. (a) Images of a flawless and of a deformed surface. (b) Comparison of the 3D surface profiles obtained with the local optimisation scheme. (c) Difference between the two profiles. (d) Comparison of the 3D reconstruction result of a cross-section of the deformed surface with the ground truth acquired by a laser focus profilometer.

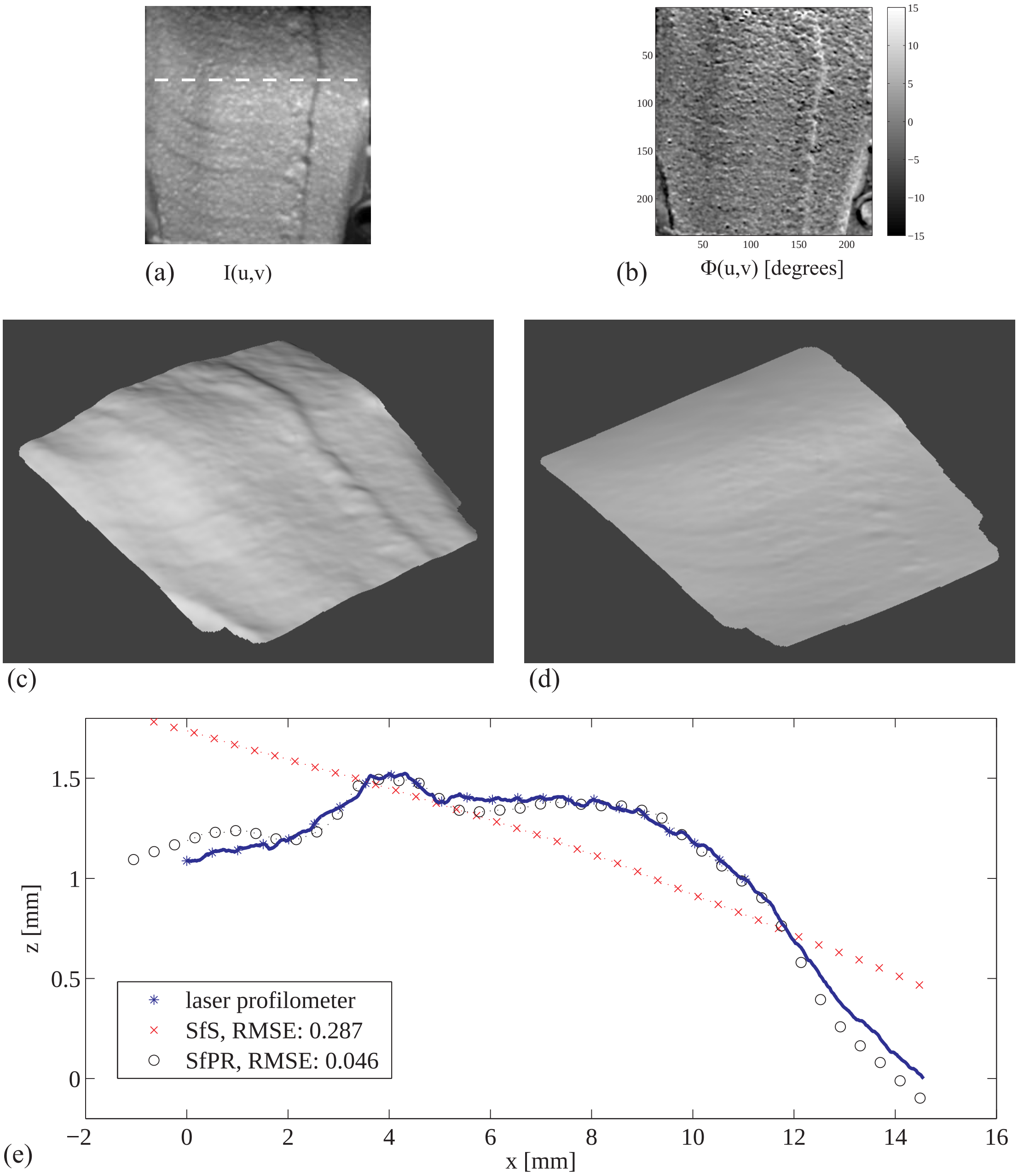


Figure 11.4: (a) and (b): Reflectance and polarisation angle images of the raw forged iron surface of the engine part. (c) Reconstruction obtained with the SfPR approach, initialisation with zero surface gradients, albedo estimated based on specular reflections according to Eq. (8.18). (d) Shape from Shading result (e) Cross-sections, compared to ground truth.

extension of the scene of about 4 mm the extracted set of 3D points contains reasonably reliable information only about the average surface gradient but not about higher-order surface properties. 3D reconstruction was performed based on a combination of intensity and polarisation angle (Fig. 11.4a). The albedo ρ_0 was estimated based on all image pixels according to Eq. (8.19) with the surface gradients set to $p_{\mathrm{DfD}}(u,v)$ and $q_{\mathrm{DfD}}(u,v)$ and was kept constant during the iteration process. Cross-sections extracted from the corresponding reconstructed surface profiles and their comparison to ground truth are shown in Fig. 11.5e. The RMSE values are 61 μm for the SfPR approach and 719 μm for the SfS approach. The SfS approach again does not estimate correctly the surface gradients perpendicular to the direction of incident light, which results in a large RMSE value. Including polarisation information yields largely the same result as obtained with the albedo estimated from specular reflections, but without requiring the presence of specular reflections in the image.

11.2.1.2 Flange

For the flange as shown in Fig. 11.6a we initialised the global SfPR approach with zero surface gradients and determined the uniform surface albedo ρ_0 according to Eq. (8.18), relying on specular reflections. Although the small-scale deformations of the surface are clearly apparent in the SfPR result (Fig. 11.6c) and to a lesser extent also in the SfS result (Fig. 11.6d), large-scale deviations from the essentially flat true surface shape are apparent.

11.2.2 Fusion of SfPR and sparse depth data

We calibrated the stereo setup with the automatic camera calibration system described by Krüger et al. (2004) and generated images in standard epipolar geometry in a subsequent rectification step. Effectively, this results in typical disparity values of around 4000 pixels at the object distance in the rectified stereo image pairs. Experiments with synthetic data have shown that the standard deviation of the disparity amounts to 0.3 pixels, resulting in an estimated standard deviation of 30 μm of the resulting depth points. One of the stereo cameras is equipped with a rotating linear polarisation filter and is used to acquire the images required for SfPR according to Section 8. Due to the highly specular reflectance of the metallic surfaces, only a sparse set of depth points can be reliably extracted using the block-matching Stereo algorithm.

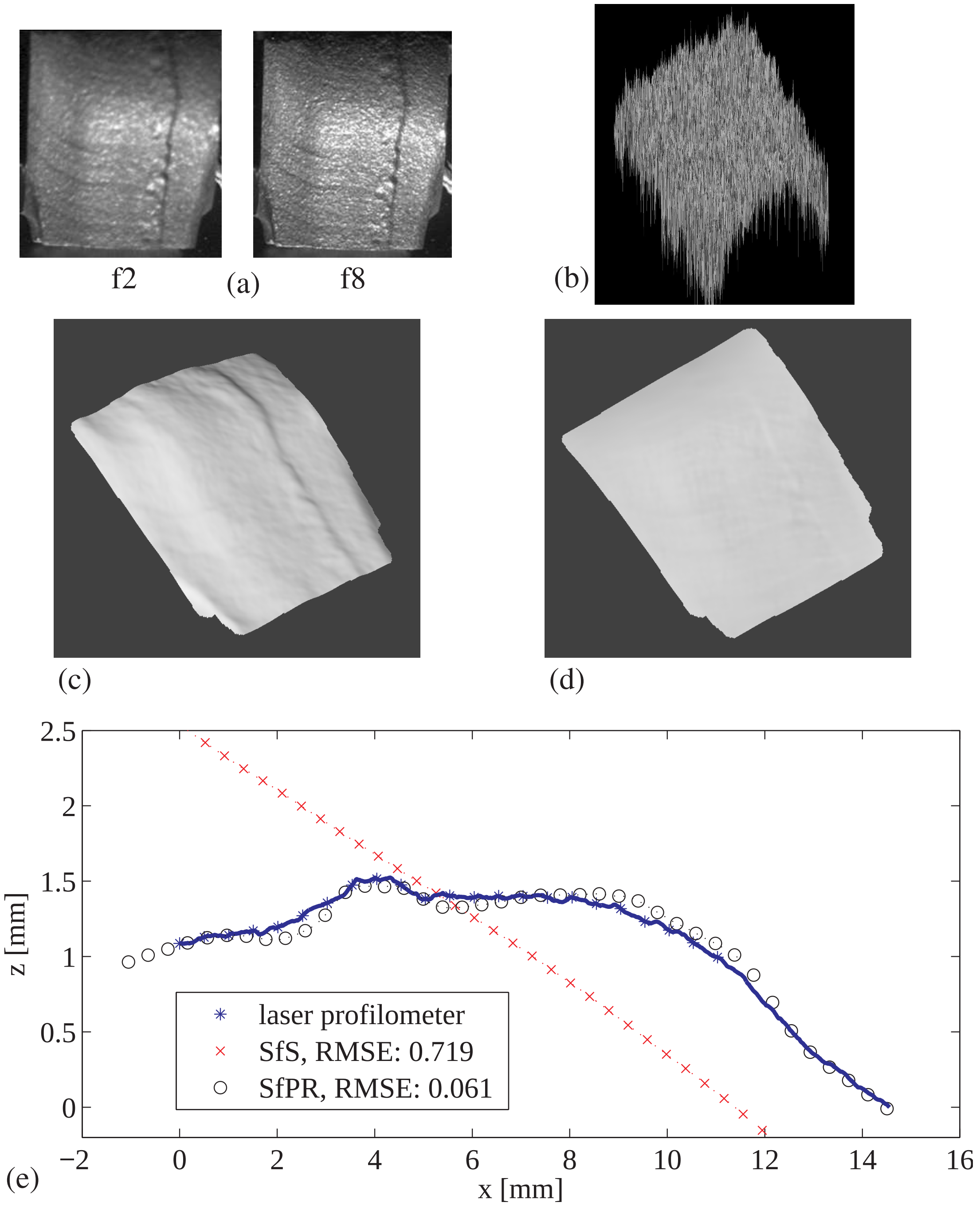
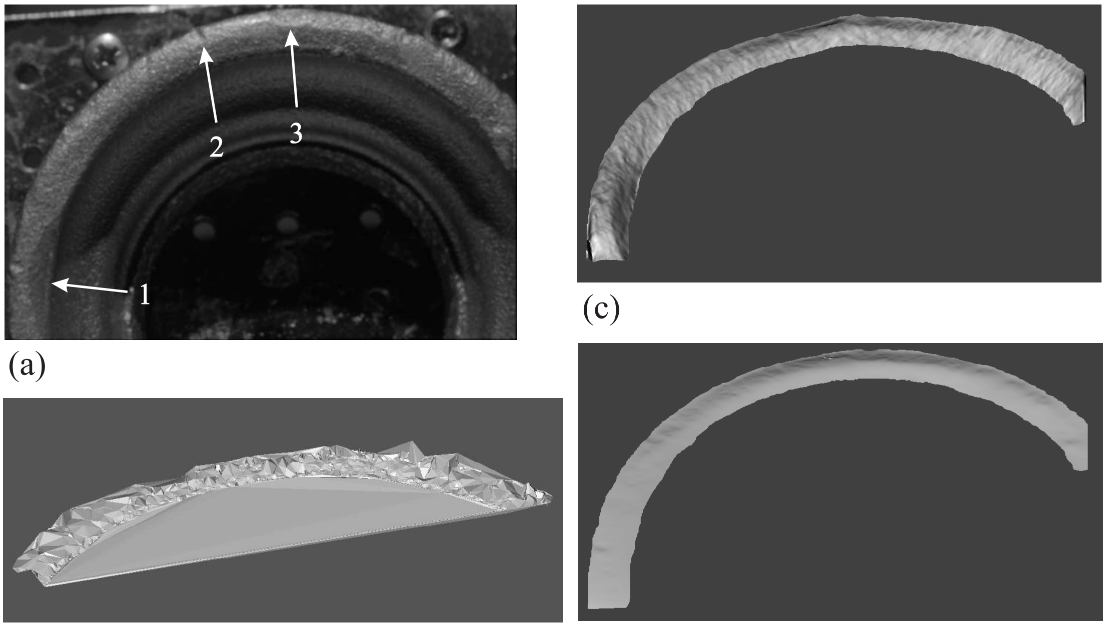


Figure 11.5: (a) Images captured with different aperture setting. (b) Depth map reconstruced by DfD. Reconstruction results for (c) SfPR and (d) SfS approach, initialisation with surface gradients obtained by DfD, initial albedo estimated based on all image pixels and DfD initialisation according to Eq. (8.19). (e) Cross-sections of the raw forged iron surface of the engine part, compared to ground truth.



(a)

(c)

(b)

(d)

Figure 11.6: (a) Image of the flange. 3D reconstruction is performed for the ring-shaped surface part. The depths of the indicated dents were measured on the reconstructed surface profile and compared to ground truth data. (b) Triangulated Stereo reconstruction result. All results are shown from a viewpoint that was not included in the input images. (c) SfPR, no Stereo information. (d) SfS, no Stereo information. Albedo estimated based on specular reflections according to Eq. (8.18).

For the raw forged iron surface of the engine part, Fig. 11.4a-b shows the rectified stereo pair, Fig. 11.7c the triangulated Stereo reconstruction result. The surface albedo was estimated based on Eq. (8.19) during each step of the iteration process. We found that the RMSE between the corresponding cross-section extracted from our reconstructed 3D profile and the ground truth amounts to 39 μm (Figs. 11.7d and 11.7f). If the SfS approach is used such that polarisation information is not taken into account, the RMSE is 51 μm (cf. Fig. 11.7e). For the examined strongly specular surface, Figs. 11.7d and 11.7e illustrate that in contrast to the SfS approach, the SfPR method reveals a large amount of small scale surface detail. The results of the comparison to ground truth data are summarised in Table 11.2.

Fig. 11.8 shows the 3D reconstruction of the flange surface calculated using one intensity and one polarisation angle image along with Stereo depth information. The 3D reconstruction is performed for the ring-shaped part only as the surface normals of the neighbouring parts are nearly orthogonal to the viewing direction. Our goniometer setup for measuring the intensity and polarisation reflectance functions does not cope with such an extreme viewing geometry, such that in this range the reflectance function values are unknown. What is more, photometric surface reconstruction techniques are most favourably applied when the view on the surface is large perpendicular (McEwen 1991). The triangulated set of stereo depth points is shown in Fig. 11.6b. As in the previous example, the surface albedo was estimated during the iteration process according to Eq. (8.19). The depth of the three dents indicated in Fig. 11.6a were obtained by tactile measurement and compared to the reconstructed depth differences. Due to the small size of the surface defects the accuracy of the tactile depth measurement only amounts to 0.1 mm. The true depth of dent 1 is 1.2 mm, the reconstructed depth 1.3 mm. Dents 2 and 3 each have a true depth of 0.25 mm, while the corresponding depth on the reconstructed surface profile amounts to 0.30 mm and 0.26 mm, respectively. On large scales, our 3D reconstruction correctly displays a flat surface. These comparisons indicate a reasonable correspondence between the true surface and our reconstruction results. Figure 11.9 displays the reconstruction using the depth data recovered by the Structure from Motion and Depth from Defocus method described in Section 10.1.4. The dense depth map shown in Fig. 11.9b has been created using the SfPR and depth algorithm. Compared to the surface profile based on depth data from Stereo (cf. Fig 11.8a), several large dents are visible in the central and right reconstruction size. These depth of these dents is overestimated by the Structure from Motion and Defocus analysis, since the KLT tracker tends to produce a slight drift

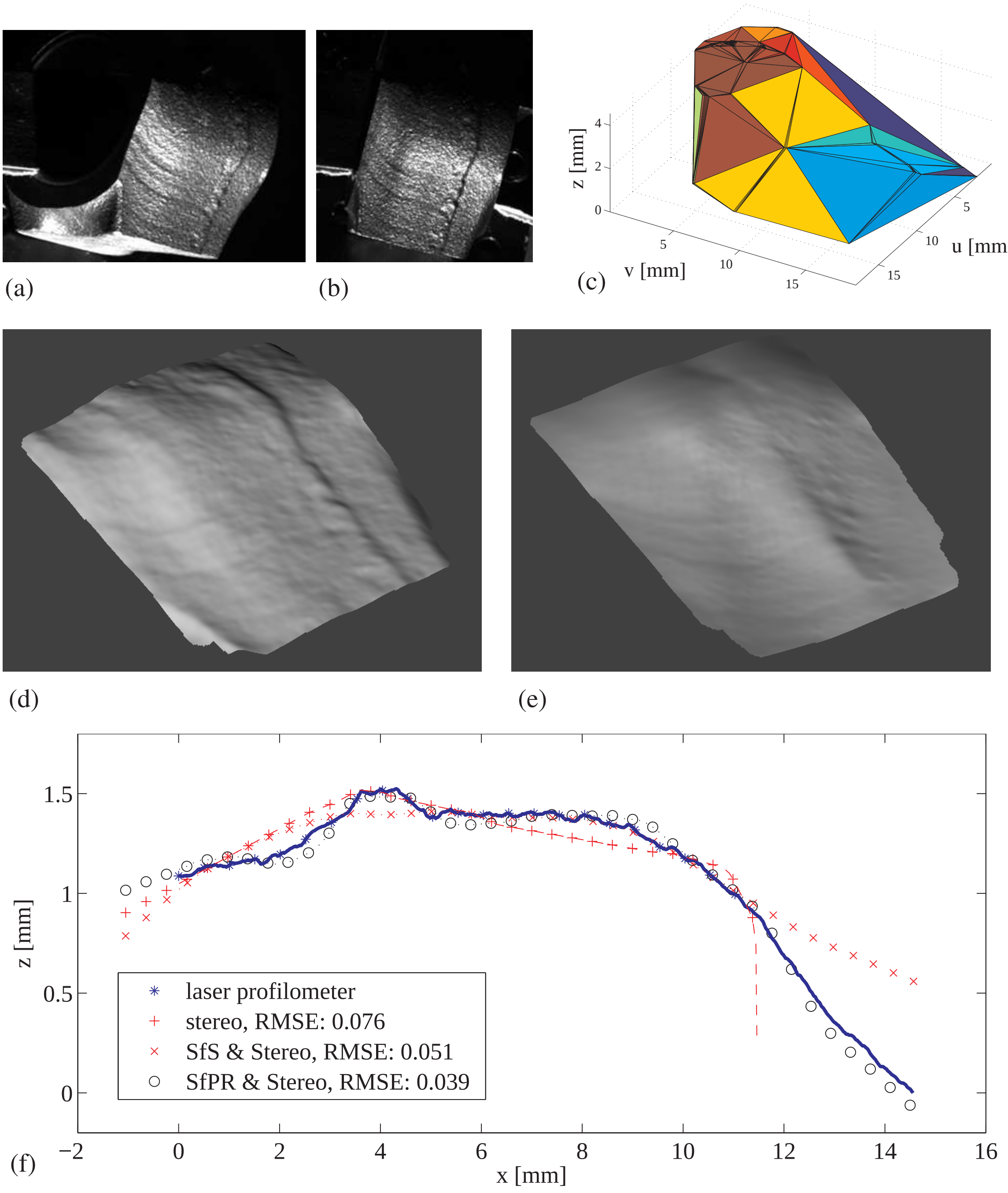
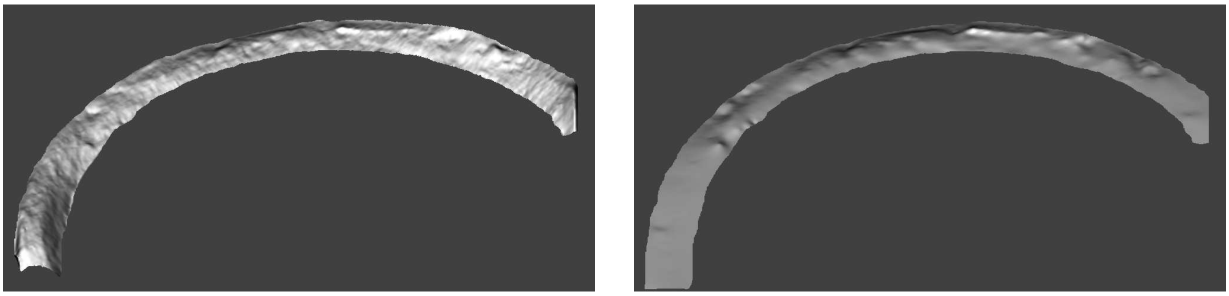
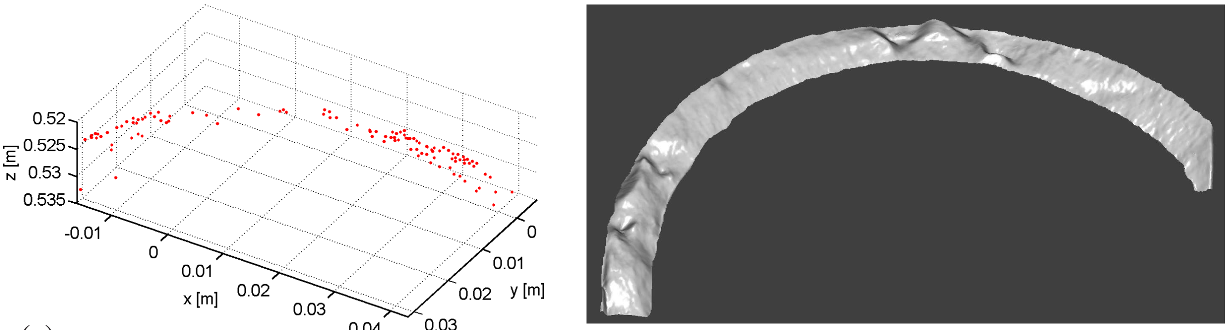


Figure 11.7: Reconstructed surface profile of the engine part using Stereo depth information. (a) and (b) rectified stereo pair. (c) Triangulated Stereo reconstruction result. (d) SfPR and (e) SfS approach combined with Stereo information, albedo estimated during the iteration process based on all image pixels according to Eq. (8.19). (f) Cross-sections of the raw forged iron surface, compared to ground truth.



(a) (b)

Figure 11.8: Reconstructed surface profile of the flange. (a) SfPR and (b) SfS approach combined with Stereo information, albedo estimated during the iteration process based on all image pixels according to Eq. (8.19).



(a) (b)

Figure 11.9: 3D of the raw cast iron surface using depth data obtained using Structure from Motion and Defocus. (a) Sparse 3D Reconstruction, cf. section 10.1.4 for details. (b) Dense 3D reconstruction using the combination of Structure from Motion and Defocus with SfPR.

Table 11.2: 3D reconstruction results for the raw forged iron surface of the engine part, obtained based on comparison of the cross-section shown in Fig. 11.4a. Albedo determination marked as "initial" denotes that the albedo was estimated prior to the iteration process either based on specular reflections or based on DfD data and was not changed afterwards, while "adapted" denotes an estimation of the albedo during the iteration process.

Utilised information	Albedo determination	Figure	Depth RMSE [μm]
SfS	Eq. (8.18), initial	11.4	287
SfPR	Eq. (8.18), initial	11.4	46
SfS, DfD	Eq. (8.19), initial	11.5	719
SfPR, DfD	Eq. (8.19), initial	11.5	61
Stereo	–	11.7	76
SfS, Stereo	Eq. (8.19), adapted	11.7	51
SfPR, Stereo	Eq. (8.19), adapted	11.7	39

of the feature positions relative to the object across the image sequence, due to the highly specular surface reflectance.

As a third application example, we have analysed a sequence of five images of the lunar crater Kepler acquired by the SMART-1 spacecraft on January 13, 2006, from heights above the lunar surface between 1613 and 1702 km (ESA 2006). The crater diameter amounts to 32 km. During image acquisition the spacecraft flew over the crater and at the same time rotated around its axis, such that the crater remained in the field of view over a considerable period of time. The first and the last image of the sequence are shown in Fig. 11.10a. Image size is 512 × 512 pixels. Fig. 11.10b shows the reconstructed part of the surface, which is smaller than the complete field of view as the surface albedo becomes non-uniform at larger distances from the crater. The image is rotated such that north is to the top and west to the left. After tracking corresponding points using the KLT tracker Shi and Tomasi (1994), we used bundle adjustment to reconstruct a sparse set of 3D point from the image sequence, which is shown in Fig. 11.10c after triangulation. Since no lens calibration data was available we had to assume that the lens can be described by the pinhole model with the principal point in the image centre. The image scale amounts to 146 m per pixel (ESA 2006), such that the scaling constant could be readily determined for the SfM result.

Since no polarisation information was available, we combined the SfS method with the result of SfM (cf. Section 9.2.2). For this purpose we employed the Lunar-Lambert reflectance

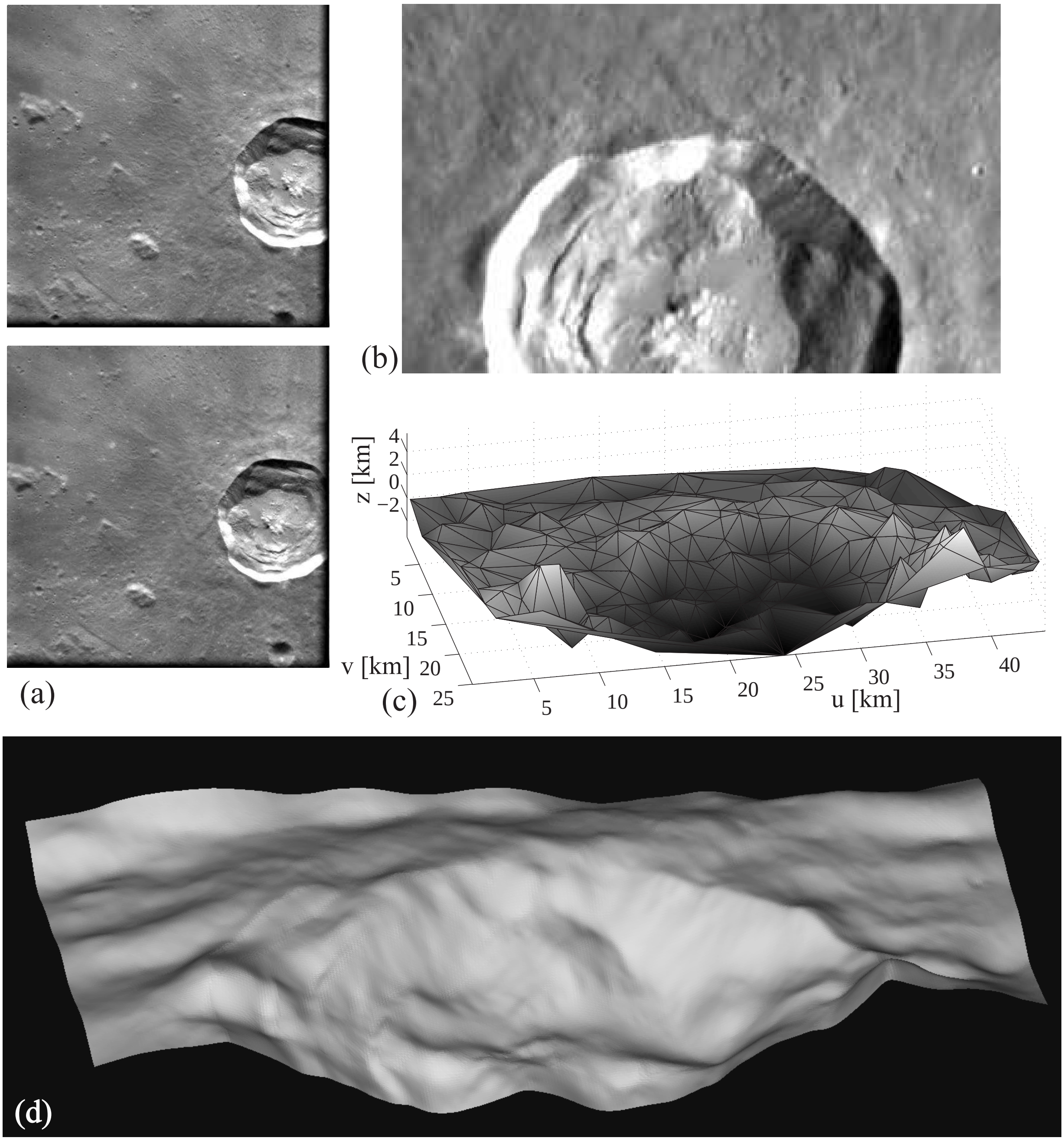


Figure 11.10: 3D reconstruction of the lunar crater Kepler. (a) First and last image of the five-image sequence acquired by the SMART-1 spacecraft (images courtesy of ESA). (b) Region selected for 3D reconstruction, oriented such that north is to the top and west to the left. (c) 3D Reconstruction obtained by SfM (triangulated set of 3D points). (d) Dense 3D reconstruction using the combination of SfM with SfS.

function

$$R_I(\theta_i, \theta_e, \alpha) = \rho_0 \left[2L(\alpha) \frac{\cos\theta_i}{\cos\theta_i + \cos\theta_e} + (1 - L(\alpha)) \cos\theta_i \right] \tag{11.1}$$

(McEwen 1991), where $L(\alpha)$ is an empirical phase angle dependent parameter. At the time of image acquisition, the phase angle corresponded to $\alpha = 51^\circ$ for the spacecraft, corresponding to $L(\alpha) = 0.8$ according to McEwen (1991), and the solar elevation angle was 37°. The viewing direction was determined according to the normal vector of a plane fitted to the set of 3D points extracted by SfM analysis. For this non-specular surface, the albedo ρ_0 was estimated based on all image pixels in the course of the iteration process according to Eq. (8.19) as explained in Section 8.2.1. Saturated (white) pixels were excluded from the SfS analysis.

The 3D reconstruction result shown in Fig. 11.10d distinctly reveals the uneven crater floor of Kepler as well as the material that has slumped down the inner crater wall at several places, especially at the northern rim. The reconstructed surface obtained with the combined SfM and SfS approach reveals much finer detail than the SfM data alone. The typical depth difference between crater floor and rim amounts to about 2850 m. No ground truth is available for this crater since it is not covered by the existing lunar topographic maps. A crater depth of 2750 m is reported in the lunar atlas by Rükl (1999). This is an average value since most crater depths given in lunar atlases were determined by shadow length measurements based on telescopic or spacecraft observations. The crater depth extracted from our 3D reconstruction result is in reasonable agreement with the value given by Rükl (1999).

This example demonstrates the usefulness of the combination of intensity data and sparse depth data obtained from a camera moving in an uncontrolled manner, regarding a surface with well-defined reflectance properties under accurately known illumination conditions. The self-consistent solution for the 3D surface profile obtained according to Section 9.2.2 yields a crater rim of largely uniform height, indicating that the estimated surface gradients in the horizontal and in the vertical image direction are essentially correct. In contrast, surface reconstruction by SfS alone based on images acquired under identical illumination conditions is not able to simultaneously estimate both surface gradients for each pixel as long as no boundary values are known for the surface to be reconstructed. What is more, in contrast to previous methods (Samaras et al. 2000; Fassold et al. 2004; Horovitz and Kiryati 2004), the sparse depth points do not introduce spurious artifacts into the reconstructed surface profile despite the considerable noise in the 3D point cloud extracted by SfM (cf.

Fig. 11.10c).

11.3 Discussion

In the previous paragraphs we have shown that the proposed framework for 3D surface reconstruction can be favourably applied to the application scenarios of industrial quality inspection as well as planetary exploration. In the quality inspection application, the accuracy of the measured depth values is about twice as high as the lateral pixel resolution of the utilised images. The duration of image acquisition for the combined photopolarimetric and stereoscopic approach amounts to a few seconds. About one second is necessary to compute the 3D reconstruction result on a standard industrial PC. In the planetary exploration scenario, the accuracy of the obtained depth values is comparable to the lateral image resolution. We will now discuss the possible advantages and drawbacks of the framework proposed in this study in comparison to active triangulation-based scanning devices in the context of industrial quality inspection and to time-of-flight sensors in the domain of planetary exploration.

Active triangulation-based laser scanning devices may be an alternative approach to the close-range problems regarded in Section 11.2.2. Simple and inexpensive sensors of this kind only measure a single profile across the surface at a time, based on a laser line projected onto the surface and an image acquired by a camera calibrated with respect to the projector. Hence, either the sensor or the object has to be moved synchronously with image acquisition when an area measurement is performed, which may in turn introduce an intricate and expensive mechanical setup such as a laser probe combined with translation-rotation motors, articulated arms, or coordinate measurement machines (Beraldin 2004). What is more, a large number of profiles are necessary to obtain a high lateral resolution, which results in long measurement cycles. For example, a 3D reconstruction of the two metallic surfaces regarded in Section 11.2.2 would require the acquisition of about 200 and 1000 line profiles, respectively, given the lateral resolution of 0.1 mm per pixel in these experiments. Such 3D reconstruction methods may well be suitable for the inspection of parts randomly selected from the production line but not for in-line inspection scenarios with a few seconds cycle time.

Area measurements can also be carried out by image-based evaluation of fringe or coded patterns projected on the surface (Batlle et al. 1998; Beraldin 2004). However, as soon as it is desired to obtain a lateral and vertical resolution comparable to the one achieved in Section 11.2.2, the costs of such a measurement system are currently at least about an order of magnitude higher that those of the instrumental setup utilised in our experiments, i. e. a pair of standard industrial cameras, a rotating polarisation filter, and one or several LED illumination devices.

Furthermore, it is well known that projection-based systems relying on single lines or coded patterns suffer from strong difficulties in the presence of highly specular surfaces like those regarded in Section 11.2.2. The reason for this behaviour is the fact that the intensity variations in the image of the projected pattern may be considerable because for such surfaces the amount of diffuse reflection is small while mirror-like reflection is dominant. As a consequence, for some surface parts the projected light is reflected directly into the camera, leading to pixel saturation or blooming, while for other parts it is reflected past the camera, resulting in invisible parts of the pattern – the reconstructed profile then shows significant gaps for which no data are available.

In long-range applications such as planetary exploration, it is common practice to employ time-of-flight sensors to determine elevation maps. The obtained surface profiles usually have a fairly low lateral resolution, compared to images obtained from the same position in space. In principle, time-of-flight methods are able to provide very accurate measurements of the distance between the spacecraft and the surface, but they require accurate spacecraft tracking data (which are not always available) to tie the measured depth points to a common coordinate system. As an example, the Clementine lidar data have a typical lateral resolution of about 7.5 km (Neumann 2001; Bussey and Spudis 2004) and a vertical accuracy of not better than 100 m, which is suitable for extracting large-scale features such as impact basins but insufficient for the analysis of small-scale surface features. As a consequence, the recently established Unified Lunar Control Network (Archinal et al. 2006), representing the currently most accurate network of control points with known absolute spatial positions, has been generated based on stereophotogrammetry of Clementine images (despite the fact that they are actually difficult to exploit due to the high illumination) rather than lidar data. Similarly, the MOLA lidar data of the surface of Mars (Neumann 2001) have a typical lateral resolution of 230 m while the orbital images acquired during the same mission are better resolved by 1–2 orders of magnitude.

According to the above discussion, the framework proposed in this study is an accurate, cost-efficient, and fast method for 3D surface reconstruction in largely controlled environments. In uncontrolled settings, active scanning devices are presumably more suitable. In many application scenarios, it may thus be favourable to combine depth data obtained with active scanning devices (triangulation or time-of-flight, depending on the range) with image data according to the framework outlined in Section 9.2.2.

12 Summary and conclusion

In this thesis we have explored how 3D reconstructions based on different principles, such as triangulation, defocus and shading, can be integrated into a unified reconstruction framework. The key idea is to combine algorithms based on principles which provide complementary constraints on the reconstructed surface.

We have described a method for combining geometric and real-aperture methods for monocular 3D reconstruction of static scenes at absolute scale. The proposed algorithm is based on a sequence of images of the object acquired by a monocular camera of fixed focal setting from different viewpoints. Feature points are tracked over a range of distances from the camera, resulting in a varying degree of defocus for each tracked feature point. After determining the best focused image of the sequence, we obtain information about absolute depth by a Depth from Defocus approach. The inferred PSF radii for the corresponding scene points are utilised to compute a regularisation term for an extended bundle adjustment algorithm that simultaneously optimises the reprojection error and the absolute depth error for all feature points tracked across the image sequence. The proposed method yields absolutely scaled 3D coordinates of the object feature points without any prior knowledge about scene structure and camera motion. To our best knowledge, the presented algorithm is the first integration of the principles of defocus and triangulation into a self-consistent framework for 3D reconstruction. We have described the implementation of the proposed method as an offline and as an online algorithm.

Based on experiments with real-world objects, we have demonstrated that the offline version of the proposed algorithm yields absolutely scaled 3D coordinates of the feature points with typical relative errors of a few percent. For the online algorithm, the accuracy of 3D reconstruction increases with increasing number of processed images as long as the images do not become strongly blurred. At the end of the sequence, the reconstruction results of the online and the offline versions of the proposed algorithm are of comparable accuracy.

We have shown that the 3D reconstruction inaccuracies observed in our experiments can be explained by a combination of the random scatter of the extracted feature positions and the estimated PSF radii, which are both due to the noise of the pixel greyvalues, and systematic deviations of the order 1% due to thermal expansion of the optical system. Further systematic errors may be introduced if the image sequence contains strongly blurred images with an average PSF radius larger than about 2 pixels, due to deviations of the observed depth dependence of the PSF radius from the analytic model used for the Depth-Defocus-Function. Since the PSF radius is continuously computed in the course of the 3D reconstruction process, it is possible and favourable to reject such strongly blurred images or image parts accordingly. For specular surfaces, the changing appearance of the ROIs tracked across the sequence may introduce further systematic effects.

Possible application scenarios of the Structure from Motion and Defocus 3D reconstruction approach are in the domains of self-localisation and mapping for mobile robotic systems as well as 3D reconstruction of objects and surfaces in the context of industrial machine vision tasks.

Another type of 3D reconstruction considered in this thesis is reconstruction of dense depth maps based on photometric principles. We have presented an image-based 3D surface reconstruction method relying on simultaneous evaluation of intensity and polarisation features (Shape from Photopolarimetric Reflectance) and its combination with absolute depth data. The proposed technique is based on the analysis of single or multiple intensity and polarisation images. To compute the surface gradients, we have presented a global optimisation method based on a variational framework and a local optimisation method based on solving a set of nonlinear equations individually for each image pixel. These approaches are suitable for strongly non-Lambertian surfaces and those of diffuse reflectance behaviour. We have demonstrated that including polarisation information into the 3D reconstruction scheme significantly increases the accuracy of the reconstructed surface for the single light source case.

Furthermore, we have described how independently measured absolute depth data are integrated into the Shape from Photopolarimetric Reflectance framework in order to increase the accuracy of the 3D reconstruction result. In this context we concentrated on dense but noisy depth data obtained by Depth from Defocus and on sparse but more accurate depth data obtained by stereo analysis or Structure from Motion. These image-based approaches are well-known methods to derive depth points, but our framework is open for independently measured 3D data obtained from other sources such as laser triangulation.

We have shown that DfD information can be used for determining the large-scale properties of the surface. It is preferentially integrated into the SfPR approach by appropriately initialising the surface gradients in the optimisation schemes, at the same time providing an estimate of the surface albedo. For integration of sparse depth information, we have suggested an optimisation scheme that simultaneously adapts the surface gradients to the measured intensity and polarisation data and to the surface slopes implied by depth differences between pairs of depth points.

In the single image case without integration of sparse depth data, the estimation of the surface albedo during the iterative reconstruction process strongly increases the manifold of local minima of the error function. Hence, we found experimentally that it is favourable to keep the initial albedo value, obtained by specular reflections or DfD information, constant during the iteration process. On the other hand, if reasonably accurate sparse depth information is available, the surface albedo can be treated as a free parameter in the optimisation process.

Our experiments on synthetic ground truth data have shown that integration of sparse depth data obtained by Stereo and Structure from Motion analysis significantly increases the 3D reconstruction accuracy. These findings are confirmed by experiments on real-world data. These findings are confirmed by experiments on real-world data, indicating the broad applicability of the proposed methods in a variety of controlled and semi-controlled scenarios.

We have shown that our method is especially suited for the 3D reconstruction of rough metallic surfaces, which are of high relevance in many industrial domains. The proposed framework integrating photometric and/or polarimetric information with sparse depth data derived from Stereo or Structure from Motion is a good choice when the material-specific reflectance properties of the surface are known, a high lateral and vertical resolution are required, and a fast and low-cost system is envisioned. Especially the latter issue tends to be of high relevance in the context of in-line industrial quality inspection.

Furthermore, we have demonstrated by applying our framework to a sequence of spacecraft images of a lunar crater that our method can also be applied to non-specular surfaces and that it still yields reasonable results when no polarisation information is available. Hence, our experimental results indicate the broad applicability of the proposed methods.

12.1 Outlook

The methods introduced in this thesis can be extended in several directions. One issue concerning the Structure from Motion and Depth from Defocus approach is the requirement of an ideally focussed image. This is not a problem for sequences where the object of primary interest moves through the plane of maximum sharpness, for example if the camera is performing a forward motion. If this is not the case, further investigations are necessary to examine whether this restriction can be removed by replacing the current absolute blur measurements by relative blur differences between defocussed frames.

Further work on SfPR will include replacing the sparse depth data obtained from stereoscopic image analysis by depth data measured with a triangulation sensor based on the projection of single lines or coded patterns. Since we have shown that our proposed framework is able to cope with fairly sparse depth data, we do not expect a degradation of the measurement accuracy even if these depth data contain considerable gaps or are acquired at low lateral resolution in order to achieve short measurement cycles.

The runtime of the reconstruction algorithm could be further reduced by replacing the iterative multi-resolution surface gradient estimation procedure with a multigrid solver (Ascher and Haber 2003). A multigrid solver considers all resolution steps at the same time, resulting in faster convergence.

With respect to the integration of Stereo into the SfPR framework, further work will concentrate on the use of both images during evaluation of the reflectance constraint. Currently only one image of the stereo pair is used. If Lambertian scenes are considered, the second image does not provide additional constraints on the shape of the surface. However, for surfaces with non-Lambertian reflectance, the additional image provide further constraints on the surface shape, especially if a wide baseline is used.

Bibliography

B. A. Archinal, M. R. Rosiek, R. L. Kirk, and B. L. Redding. The unified lunar control network 2005. USGS Open-File Report 2006-1367, 2006. URL `http://pubs.usgs.gov/of/2006/1367`.

U. Ascher and E. Haber. A multigrid method for distributed parameter estimation problems, 2003.

K. Åström. Algebraic varieties in multiple view geometry. In *ECCV '96: Proceedings of the 4th European Conference on Computer Vision*, volume II, pages 671–682, 1996. ISBN 3-540-61123-1.

G. A. Atkinson and E. R. Hancock. Multi-view surface reconstruction using polarization. *Proc. IEEE Int. Conf. Comp. Vis.*, 1:309–316, 2005. ISSN 1550-5499.

N. Ayache. *Artificial Vision for Mobile Robots*. The MIT Press, 1991.

S. T. Barnard and M. A. Fischler. Computational stereo. *ACM Comput. Surv.*, 14(4): 553–572, 1982. ISSN 0360-0300. doi: http://doi.acm.org/10.1145/356893.356896.

J. Batlle, E. Mouaddib, and J. Salvi. Recent progress in coded structured light as a technique to solve the correspondence problem: a survey. *Pattern Recognition*, 31(7): 963–982, 1998.

P. Beckmann and A. Spizzichino. *The scattering of electromagnetic waves from rough surfaces*. Pergamon, New York, 1963.

P. N. Belhumeur, D. J. Kriegman, and A. L. Yuille. The bas-relief ambiguity. *Int. J. Comput. Vision*, 35(1):33–44, 1999.

J. A. Beraldin. Integration of laser scanning and close-range photogrammetry – the last decade and beyond. In *Proc. 20th Int. Soc. for Photogrammetry and Remote Sensing Congress, Commission VII*, pages 972–983, 2004.

S. Birchfield. An introduction to projective geometry (for computer vision). http://www.ces.clemson.edu/ stb/projective/, 1998.

M. Born and E. Wolf. *Principles of Optics*. Pergamon, 1965.

J. Bouguet. Camera calibration toolbox for MATLAB, 1997. URL `www.vision.caltech.edu/bouguetj/calib_doc`.

Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. Pattern Anal. Machine Intell.*, 26(9):1124–1137, 2004. ISSN 0162-8828.

D. C. Brown. A solution to the general problem of multiple station analytical stereotriangulation. Technical report rca-mtp data reduction technical report no. 43 (or afmtc tr 58-8), Patrick Airforce Base, Florida, 1958.

M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(8):993–1008, 2003. ISSN 0162-8828. doi: http://doi.ieeecomputersociety.org/10.1109/TPAMI.2003.1217603.

C. Bussey and P. Spudis. *The Clementine Atlas of the Moon*. Cambridge University Press, Cambridge, UK, 2004.

S. Chaudhuri, A. Rajagopalan, and A. Pentland. *Depth from Defocus: A Real Aperture Imaging Approach*. Springer, 1999. ISBN 0387986359.

T. Cormen, C. Leiserson, and R. Rives. *Introduction to Algorithms*. McGraw-Hill, New York, 1990.

I. J. Cox, S. L. Hingorani, S. B. Rao, and B. M. Maggs. A maximum likelihood stereo algorithm. *Comput. Vis. Image Underst.*, 63(3):542–567, 1996. ISSN 1077-3142. doi: http://dx.doi.org/10.1006/cviu.1996.0040.

J. J. Craig. *Introduction to Robotics: Mechanics and Control*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1989. ISBN 0201095289.

J. Cryer, P. Tsai, and M. Shah. Integration of shape from shading and stereo. *Pattern Recognition*, 28(7):1033–1043, July 1995.

U. Dhond and J. Aggarwal. Structure from stereo-a review. *IEEE Trans. Systems, Man, and Cybernetics*, 19:1489–1510, 1989.

ESA. Kepler crater as seen by smart-1. Technical report, European Space Agency, 2006. URL `http://www.esa.int/SPECIALS/SMART-1/SEMBGLVTOPE_2.html`.

H. Fassold, R. Danzl, K. Schindler, and H. Bischof. Reconstruction of archaeological finds using shape from stereo and shape from shading. In *9th Computer Vision Winter Workshop*, pages 21–30, Piran, Slovenia, 2004.

O. Faugeras. *Three-Dimensional Computer Vision (Artificial Intelligence)*. The MIT Press, Cambridge, Massachusetts, 1993. ISBN 0262061589.

S. Finsterwalder. *Die geometischen Grundlagen der Photogrammetrie*. Jahresbericht d. Deutschen Mathem. Vereinigung. Teubner Press, Leipzig, Germany, 1899.

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

J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, and R. L. Phillips. *Introduction to Computer Graphics*. Addison-Wesley Professional, 1993.

T. A. Germer, T. Rinder, and H. Rothe. Polarized light scattering measurements of polished and etched steel surfaces. In *Proc. SPIE Scattering and Surface Roughness III*, volume 4100, pages 148–155, 2000.

D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying brdfs from photometric stereo. *Proc. IEEE Int. Conf. Comp. Vis.*, 1:341–348, 2005.

R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.

A. Hertzmann and S. M. Seitz. Example-based photometric stereo: Shape reconstruction with general, varying brdfs. *IEEE Trans. Pattern Anal. Machine Intell.*, 27(8):1254–1264, 2005.

H. Hirschmüller. Stereo vision in structured environments by consistent semi-global matching. In *Proc. IEEE Int. Conf. Comp. Vis. Pat. Recog.*, volume 2, pages 2386–2393, 2006.

B. Horn. Obtaining shape from shading information. In B. Horn and M. Brooks, editors, *(1989) Shape from Shading*, pages 123–171. MIT Press, 1975.

B. K. P. Horn. *Robot vision.* MIT Press, Cambridge, MA, USA, 1986. ISBN 0-262-08159-8.

B. K. P. Horn. Height and Gradient from Shading. Technical Memo AIM-1105, Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts, USA, May 1989b.

B. K. P. Horn and M. J. Brooks. *Shape from shading.* MIT Press, Cambridge, MA, USA, 1989a. ISBN 0-262-08183-0.

I. Horovitz and N. Kiryati. Depth from gradient fields and control points: Bias correction in photometric stereo. *Image and Vision Computing*, 22:681–694, 2004.

S. Intille and A. Bobick. Incorporating intensity edges in the recovery of occlusion regions. In *Proc. Int. Conf. Pattern Recognition*, volume 1, pages 674–677, 1994.

X. Jiang and H. Bunke. *Dreidimensionales Computersehen: Gewinnung und Analyse von Tiefenbildern.* Springer, Berlin, 1996.

M. V. Joshi and S. Chaudhuri. Photometric stereo under blurred observations. In *Proc. 17th Int. Conf. on Pattern Recognition*, volume 3, pages 169–172, Cambridge, UK, 2004.

K. Koshikawa. A polarimetric approach to shape understanding. In *Proc. IJCAI*, pages 493–495, Tokyo, 1979.

E. P. Krotkov. Focusing. *The Int. J. of Comp. Vision*, 1(3):223–237, Oct 1988.

L. Krüger, C. Wöler, A. Würz-Wessel, and F. Stein. In-factory calibration of multiocular camera systems. In *Photonics Europe*, 2004.

E. Kruppa. Zur Ermittlung eines Objektes aus zwei Perspektiven mit innerer Orientierung. *Sitzungsberichte der Mathematisch Naturwissenschaftlichen Kaiserlichen Akademie der Wissenschaften*, 122:1939–1948, 1913.

C. Lee and A. Rosenfeld. Improved methods of estimating shape from shading using the light source coordinate system. *Artificial Intelligence*, 26:125–143, 1985.

R. Lena, C. Wöhler, M. T. Bregante, and C. Fattinnanzi. A combined morphometric and spectrophotometric study of the complex lunar volcanic region in the south of petavius. *Journal of the Royal Astronomical Society of Canada*, 100(1):14–25, 2006.

K. Levenberg. Method for the solution of certain problems in least-squares. *Quarterly Applied Math.*, 2:164–168, 1944.

J. Lim, H. Jeffrey, M. Yang, and D. Kriegman. Passive photometric stereo from motion. In *IEEE Int. Conf. on Computer Vision*, volume II, pages 1635–1642, Beijing, China, 2005.

M. Lourakis and A. Argyros. The design and implementation of a generic sparse bundle adjustment software package based on the levenberg–marquardt algorithm. Technical Report 340, Institute of Computer Science—FORTH, Heraklion, Crete, Greece, 2004. URL `http://citeseer.ist.psu.edu/lourakis04design.html`.

T. Luhmann. *Nahbereichsphotogrammetrie. Grundlagen, Methoden und Anwendungen.* Wichmann, Heidelberg, 2003.

K. Madsen, H. B. Nielsen, and O. Tingleff. Methods for non-linear least squares problems, 2nd edition, 2004.

A. McEwen. Topography and albedo of ius chasma, mars. In *Proc. 16th Conf. on Lunar and Planetary Science*, pages 528–529, 1985.

A. S. McEwen. Photometric functions for photoclinometry and other applications. *Icarus*, 92:298–311, 1991.

D. Miyazaki, M. Kagesawa, and K. Ikeuchi. Transparent surface modeling from a pair of polarization images. *IEEE Trans. Pattern Anal. Machine Intell.*, 26(1):73–82, 2004.

D. Miyazaki, R. T. Tan, K. Hara, and K. Ikeuchi. Polarization-based inverse rendering from a single view. In *IEEE Int. Conf. on Computer Vision*, volume II, pages 982–987, Nice, France, 2003.

P. Moreels and P. Perona. Evaluation of features detectors and descriptors based on 3d objects. In *Proc. IEEE Int. Conf. Comp. Vis.*, volume 1, pages 800–807, 2005. ISBN 0-7695-2334-X-01.

O. Morel, F. Meriaudeau, C. Stolz, and P. Gorria. Polarization imaging applied to 3d inspection of specular metallic surfaces. In *SPIE Machine Vision Applications in Industrial Inspection XIII*, volume 5679, pages 178–186, 2005.

O. Morel, C. Stolz, F. Meriaudeau, and P. Gorria. Active lighting applied to three-dimensional reconstruction of specular metallic surfaces by polarization imaging. *Appl. Opt.*, 45(17):4062–4068, 2006.

Z. Myles and N. da Vitoria Lobo. Recovering affine motion and defocus blur simultaneously. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(6):652–658, 1998. ISSN 0162-8828. doi: http://dx.doi.org/10.1109/34.683782.

S. Nayar and Y. Nakagawa. Shape from Focus. *IEEE Trans. Pattern Anal. Machine Intell.*, 16(8):824–831, Aug 1994.

S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: Physical and geometrical perspectives. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(7):611–634, 1991.

G. A. Neumann. Some aspects of processing extraterrestrial lidar data: Clementine near mola. In *Int. Arch. Photogrammetry Remote Sensing 34(3/W4)*, pages 73–80, 2001.

D. Nister. An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–777, 2004.

J. Oliensis and P. Dupuis. An optimal control formulation and related numerical methods for a problem in shape reconstruction. *The Annals of Applied Probability*, 4(2):287–346, 1994.

F. L. Pedrotti. *Introduction to Optics, 2nd Edition.* Prentice Hall, 1993.

A. P. Pentland. Local shading analysis. *IEEE Trans. Pattern Anal. Machine Intell.*, 6: 170–187, 1984.

A. P. Pentland. A new sense for depth of field. *IEEE Trans. Pattern Anal. Mach. Intell.*, 9(4):523–531, 1987.

E. Prados. *Application of the theory of the viscosity solutions to the Shape From Shading problem.* PhD thesis, University of Nice Sophia-Antipolis, Oct. 2004.

E. Prados and O. Faugeras. Shape from shading: A well-posed problem? In *Proc. IEEE Int. Conf. Comp. Vis. Pat. Recog.*, volume 2, pages 870–877, 2005.

S. Rahmann. Inferring 3D scene structure from a single polarization image. In *Conf. on Polarization and Color Techniques in Industrial Inspection*, volume 3826, pages 22–33, Munich, Germany, 1999.

S. Rahmann and N. Canterakis. Reconstruction of specular surfaces using polarization imaging. In *Proc. IEEE Int. Conf. Comp. Vis. Pat. Recog.*, volume 1, pages 149–155, 2001.

W. J. J. Rey. *Introduction to Robust and Quasi-Robust Statistical Methods.* Springer Verlag, Berlin, Heidelberg, 1983.

E. Rouy and A. Tourin. A viscosity solutions approach to shape-from-shading. *SIAM J. Numer. Anal.*, 29(3):867–884, 1992.

A. Rükl. *Mondatlas.* Verlag Werner Dausien, Hanau, Germany, 1999.

D. Samaras, D. Metaxas, P. Fua, and Y. Leclerc. Variable albedo surface reconstruction from stereo and shape from shading. In *Proc. CVPR*, volume I, pages 480–487, 2000.

D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *IJCV*, 47:7–42, 2002.

J. Shi and C. Tomasi. Good features to track. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, 1994.

T. Simchony, R. Chellappa, and M. Shao. Direct analytic methods for solving poisson equations in computer vision problems. *IEEE Trans. Pattern Anal. Machine Intell.*, 12 (5):435–556, 1990.

S. Squyres. The topography of ganymede's grooved terrain. *Icarus*, 46:156–168, 1981.

H. Stewénius, C. Engels, and D. Nistér. Recent developments on direct relative orientation. *ISPRS Journal of Photogrammetry and Remote Sensing*, 60:284–294, June 2006.

M. Subbarao. Parallel depth recovery by changing camera parameters. In *Proc. IEEE Int. Conf. Comp. Vis.*, pages 149–155, 1988.

M. Subbarao and T. Choi. Accurate recovery of three-dimensional shape from image focus. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17(3):266–274, 1995.

C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *Int. J. Comput. Vision*, 9(2):137–154, 1992.

K. Torrance and E. Sparrow. Theory for off-specular reflection from roughened surfaces. *J. Opt. Soc. Am.*, 57, 1967.

C. Wöhler and K. Hafezi. A general framework for three-dimensional surface reconstruction by self-consistent fusion of shading and shadow features. *Pattern Recognition*, 38(7):965–983, 2005.

C. Wöhler and L. Krüger. A contour based stereo vision algorithm for video surveillance applications. In *Proc. SPIE Visual Communication and Image Processing*, Lugano, 2003.

M. Wolf and E. Born. *Principles of Optics : Electromagnetic Theory of Propagation, Interference and Diffraction of Light*. Pergamon Press, 1959.

L. Wolff. *Polarization Methods in Computer Vision*. PhD thesis, Columbia University, 1991.

L. Wolff and T. Boult. Constraining object features using a polarization reflectance model. *IEEE Trans. Pattern Anal. Machine Intell.*, 13(7):635–657, 1991.

R. J. Woodham. Photometric method for determining surface orientation from multiple images. *J. of Optical Engineering*, 19(1):138–144, 1980.

R. J. Woodham. Gradient and curvature from the photometric-stereo method, including local confidence estimation. *Journal of the Optical Society of America A*, 11:3050–3068, Nov. 1994.

C. Zhang. A survey on stereo vision for mobile robots. Technical report, Dept. of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, 2003.

R. Zhang, P.-S. Tsai, J. E. Cryer, and M. Shah. Shape from shading: A survey. *IEEE Trans. Pattern Anal. Machine Intell.*, 21(8):690–706, 1999. ISSN 0162-8828. doi: http://dx.doi.org/10.1109/34.784284.