



Graz University of Technology
Institute for Computer Vision and Graphics

Master's Thesis

DYNAMIC ILLUMINATION FOR ROBUST
MICROSCOPIC 3D METROLOGY

David M. Ferstl

Graz, Austria, May 2011

Thesis supervisors

Univ.-Prof. Dipl.-Ing. Dr.techn. Horst Bischof

Dipl.-Ing. Dr.techn. Matthias R uther

Deutsche Fassung:
Beschluss der Curricula-Kommission für Bachelor-, Master- und Diplomstudien vom 10.11.2008
Genehmigung des Senates am 1.12.2008

EIDESSTÄTTLICHE ERKLÄRUNG

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

Graz, am

.....
(Unterschrift)

Englische Fassung:

STATUTORY DECLARATION

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

.....
date

.....
(signature)

Abstract

Traditional microscopic shape from focus reconstruction is often limited by the surface dynamic and the texture of the analysed specimen. In many real-world applications, surfaces have a strong varying reflectance leading to saturated image parts, or lack in detectable texture. In such cases, shape from focus generates incorrect and sparse depth maps.

In this thesis, we present a novel method to eliminate these vulnerabilities without additional reconstruction time. Beyond that, we propose a novel method to further reduce the computational costs of traditional shape from focus to a minimum. To overcome the problems of high reflectance differences and lacks in texture we use a projector-camera system to compensate the reflectance variations and additionally project measurable texture. The surface reflection is compensated by a local adaption of the illumination for every acquisition. To reduce measurement time, the compensation pattern is tracked through the image stack and is updated in a prediction-correction step. The exact projector pattern to create additional texture is determined through a detailed analysis of the focus measure operator and the optical effects during the projection.

The additional reduction in measurement time is achieved with a novel focus measure which calculates the focus through a comparison of an estimated all-in-focus image and the stack images by normalized cross correlation. Therewith, the depth estimation of each surface point in the shape from focus algorithm stops if a local focus maximum beyond a predefined threshold is found.

The experiments show, that our method outperforms the traditional shape from focus algorithm and is also a performance enhancement to comparable methods like high dynamic range imaging in terms of speed and accuracy.

Keywords. shape from focus, radiometric compensation, projector-camera system, light transport, reflectance map, active illumination, integral point spread function, normalized cross correlation, 3D reconstruction, optical microscopy

Kurzfassung

Die herkömmliche Shape From Focus (SFF) Rekonstruktion im mikroskopischen Bereich ist beschränkt durch die Oberflächendynamik und durch die Textur der untersuchten Probe. Vor allem im industriellen Bereich treten häufig großen Unterschieden in der Lichtreflexion innerhalb der gemessenen Oberfläche auf, was durch die limitierte Grauwertaufösung der Kamera zu Unter- oder Überbelichtungen führen kann. Bei sehr glatten Objekten kommt es auch zu einem generellen Fehlen von messbarer Textur. In beiden Fällen führt die SFF Rekonstruktion zu einem fehlerhaften Ergebnis.

In dieser Arbeit wird eine neue Methode vorgestellt, mit der man diese Probleme mit Hilfe eines Projektor-Kamera Systems verhindern kann. Darüber hinaus wird eine neuartige Fokussmessung präsentiert, mit der es möglich wird die Rechenzeit des SFF Algorithmus zu verringern.

Durch ein Projektor-Kamera System ist kann die Beleuchtung lokal adaptieren werden. Dadurch können die Unterschiede in der Objektreflektion in jeder Aufnahme kompensiert werden. Bei sehr glatten Oberflächen, die keine messbare Textur besitzen, kann mit diesem System zusätzliche messbare Textur aufgetragen werden.

Eine weitere Beschleunigung des SFF Algorithmus kann erreicht werden indem der Fokus durch Vergleichen der aufgenommenen Bilder mit einem komplett fokussierten Bild mit Hilfe einer normalisierten Kreuzkorrelation berechnet wird. Mit diesem Vergleichsmaß kann die Tiefensuche für jeden Bildpunkt beendet werden, wenn ein lokales Maximum über einem definierten Schwellwert gefunden wird.

In den Experimenten wird gezeigt, dass dadurch die Genauigkeit und Robustheit der Rekonstruktion im Vergleich zu herkömmlichen SFF deutlich verbessert wird, wobei die Aufnahmezeit im Unterschied zu vergleichbaren Methoden wie High Dynamic Range Imaging gleich bleibt.

Acknowledgments

First of all I am heartily thankful to my supervisors, Prof. Dr. Horst Bischof and Dr. Matthias R  ther, for giving me the possibility to work on this interesting project. Their support and guidance from the preliminary to the concluding level enabled me to develop an understanding of the subject and its challenges.

Furthermore I want to thank the Robot Vision Team at the Institute for Computer Vision and Graphics (ICG), especially Martin Lenz, whose extraordinary support was decisive for the success of this project, Katrin Pirker, Markus Heber, Christian Reinbacher and Matthias Straka for the pleasant collaboration, helpful advices and inspiring discussions.

Last but not least, I want to thank my family for giving me the mental and financial support that helped me during my Bachelor and Master studies. They always believed in me and encouraged me in every phase of my work.

Contents

1	Introduction	1
1.1	Problem Statement	1
1.2	Motivation	3
1.3	Overview	5
2	Related Work	7
2.1	Microscopic Systems for 3D Reconstruction	8
2.2	Reconstruction based on Small Depth of Field	9
2.3	Projector-Camera Systems	11
3	Theoretical Background	13
3.1	Shape From Focus	13
3.1.1	Point Spread Function	14
3.1.2	Shape From Focus	16
3.1.3	Focus Measures	17
3.1.4	Depth Estimation - Interpolation Methods	20
3.2	High Dynamic Range Photography	24
3.2.1	Recovering the Camera Response Function	25
3.2.2	HDR irradiance Reconstruction	28
3.2.3	Disadvantages of HDR Imaging	29
3.3	Radiometric Compensation	30
3.3.1	The Projector-Camera Radiometric Chain	30
3.3.2	Calibration of the Projector Response Function	32
3.3.3	Compensation Algorithm through Geometric Correspondence	32
3.3.4	Compensation through Light Transport Matrix	32
3.4	Summary	34
4	Microscopic Projector-Camera System	35
4.1	Hardware System	35
4.2	Optical Path	39
4.3	Calibration	41
4.3.1	Radiometric Calibration	42

4.3.2	Geometric Calibration	45
4.3.3	Calibration of the Point Spread Function	48
4.3.4	Construction of the Scatter Matrix	49
4.4	Summary	53
5	Dynamic Illumination for Robust Microscopic 3D Metrology	55
5.1	Adaptive Illumination	56
5.1.1	Image Formation Process	58
5.1.2	Compensation of the Reflection	60
5.1.3	Predicted Adaptive Illumination	66
5.2	High Frequency Texture Projection	67
5.2.1	Calibration of the Optimal Pattern	68
5.2.2	Application of a HF Pattern for Illumination	72
5.3	Combination of Adaptive Illumination with High Frequency Pattern	75
5.3.1	Adaption with Added HF pattern	76
5.3.2	Adaption of a HF pattern	78
5.4	Integral Shape From Focus	79
5.4.1	Calibration of the IPSF	80
5.4.2	Application of the IPSF for Correlation Based SFF Reconstruction	82
5.5	Summary	86
6	Evaluation and Experiments	89
6.1	Adaptive Illumination	90
6.1.1	Reference Specimen	90
6.1.2	Predictive Adaption	92
6.1.3	Reconstruction with Adaptive Illumination	94
6.2	Projected Texture	100
6.2.1	Reference Specimen	100
6.2.2	Optimal Pattern Calibration	102
6.2.3	Reconstruction with HF pattern projection	102
6.3	Combination of HF Pattern with Adapted Illumination	106
6.3.1	Reference Specimen	106
6.3.2	Robustness Evaluation	106
6.4	Integral Shape From Focus	110
6.4.1	Reference Specimen	110
6.4.2	Noise Influence on the Focused Image Reconstruction	111
6.4.3	ISFF Depth Reconstructions	111
6.4.4	Speed Improvements of ISFF	117
6.5	Summary	118
7	Discussion and Outlook	121

A Acronyms and Symbols

125

Bibliography

127

List of Figures

1.1	SFF Reconstruction with adapted illumination	2
1.2	SFF Reconstruction with projected texture	4
2.1	Optical Geometry Model	9
3.1	Optical lens-model	14
3.2	Bivariate Gaussian distribution model	16
3.3	SFF system setup	16
3.4	Maximum method	21
3.5	Linefit interpolation method	22
3.6	Gaussian interpolation method	23
3.7	Scene with illumination artefacts	25
3.8	Image acquisition pipeline	26
3.9	HDR grayscale photographs	28
3.10	HDR response function	28
3.11	Image reconstruction with HDR imaging	29
3.12	Projector-camera radiometry model	31
4.1	Lacks of focus information	36
4.2	Hardware setup	36
4.3	Beamsplitter model	37
4.4	LCoS display	37
4.5	spatial light modulator	38
4.6	Optical path	39
4.7	Lens aberration	40
4.8	Coplanar focus planes of camera and projector	41
4.9	Camera response function	43
4.10	Projector response function	44
4.11	Geometric pixel correspondence	45
4.12	Geometric projector-camera position	46
4.13	Projector pixel spread on the specimen	46
4.14	Projection grid	47

4.15	Geometric calibration error	48
4.16	Cross-section through PSF approximations	49
4.17	3D PSF calibrations	50
4.18	Cross-section of PSF and OTF	50
4.19	Influences of concurring pixels on the specimen	51
4.20	Scatter matrix sketch	52
4.21	Camera image estimation through the scatter matrix	53
5.1	Sample radiometric compensation	57
5.2	Image formation pipeline	59
5.3	Iterative pattern adaption	61
5.4	Camera-projector radiometry	62
5.5	Iterative adaption process on a 1D model	64
5.6	Iterative adaption process	65
5.7	3D reconstruction with adaptive illumination	67
5.8	Sample texture projection	68
5.9	Influence radius of the projector pixel spread	69
5.10	Focus measure sphere	70
5.11	Reconstruction with different texture pattern	71
5.12	Reconstruction with different pattern frequencies	72
5.13	Focus vector with projected texture	73
5.14	3D reconstruction using texture projection	74
5.15	Combination of adaption and texture projection with histogram	75
5.16	Combination of adapted and HF pattern	76
5.17	Sample combined pattern illumination	77
5.18	Sample illumination adaption with added texture pattern	78
5.19	Sample adaption to a desired texture pattern	79
5.20	3D reconstruction with a combination of adaption and texture pattern	79
5.21	IPSF optimizations	82
5.22	3D IPSF reconstruction	82
5.23	Correlation patch measurement with NCC	83
5.24	NCC calculation for focus measurement	84
5.25	NCC focus measurement stack compared to tenengrad	85
5.26	Correlation based SFF reconstruction	86
6.1	Reference cutter blade specimen	91
6.2	Reference cross-section scans for geometric groundtruth	92
6.3	Evaluate predictive adaption	93
6.4	3D SFF reconstruction using adaptive illumination I	96
6.5	3D SFF reconstruction using adaptive illumination II	97
6.6	All-in-focus texture images for adaptive illumination I	98

6.7	All-in-focus texture images for adaptive illumination II	99
6.8	Reference specimen for texture projection	101
6.9	3D SFF reconstruction using texture projection with different frequencies .	103
6.10	3D SFF reconstruction results using texture projection I	104
6.11	3D SFF reconstruction results using texture projection II	105
6.12	Reference specimen for combination of texture projection and adaption . .	107
6.13	3D SFF reconstruction results using combination of texture projection and adaption I	108
6.14	3D SFF reconstruction results using combination of texture projection and adaption II	109
6.15	ISFF real-world reference specimen	110
6.16	ISFF synthetic reference specimen	112
6.17	Influence of noise on IPSF convolution	113
6.18	Restored synthetic all-in-focus images	114
6.19	3D ISFF reconstruction results	115
6.20	IPSF reconstructed all-in-focus images	116
6.21	Evaluation of Speed Improvements for ISFF	118
6.22	Evaluation of speed improvements in ISFF	119

List of Tables

6.1	Reference profile scans	92
6.2	HDR exposure times for SFF reconstruction	94
6.3	Reconstruction error for SFF using adaptive illumination	95
6.4	Evaluate texture pattern frequency	102
6.5	Reconstruction error of SFF using texture projection	103
6.6	Robustness evaluation of the combination of pattern projection with adaption	107
6.7	Reconstruction error of the combination of texture projection and adaption	108
6.8	ISFF reconstruction results	113
6.9	Evaluation of speed improvements in ISFF	117

Chapter 1

Introduction

Contents

1.1 Problem Statement	1
1.2 Motivation	3
1.3 Overview	5

The three dimensional (3D) reconstruction of microscopic objects is of great industrial importance to get a fast and accurate way for a robust quality inspection. In this work we present novel methods to increase the accuracy, speed and robustness of 3D microscopic object reconstruction using shape from focus (SFF). In this chapter we discuss the reconstruction problems with traditional shape from focus and give a short motivation to our work. We describe particular achievements and give an outline of the following chapters.

One of the most popular method of 3D reconstruction in optical microscopy is the SFF algorithm. Here the specimen is moved with respect to the imaging system and a stack of images with different levels of object focus is acquired. Because the exact movement is known in every step, the depth is estimated by identification of the focus maximum of every object point through the image stack. The focus is measured through detection of high frequency intensity variations in the image domain.

1.1 Problem Statement

In practice, industrial specimen produce a wide range of brightness variations in their reflection. A typical microscopic camera is limited to a resolution of only 8 – 12bit of intensity levels at each pixel. As a result, this can easily lead to under exposed or over saturated parts in the images, where the image information is lost. Therefore, the low

dynamic range of conventional image detectors limits the accuracy and robustness of the SFF algorithm, as illustrated in Figure 1.1(a,c). One possible approach to overcome these limitations is to increase the dynamic range of the sensor through sequential exposure change [14] or through different hardware [43], commonly addressed as high dynamic range (HDR) imaging. But these acquisition systems lead either to an essential increase of the total measurement time or to a major decrease of the camera resolution.

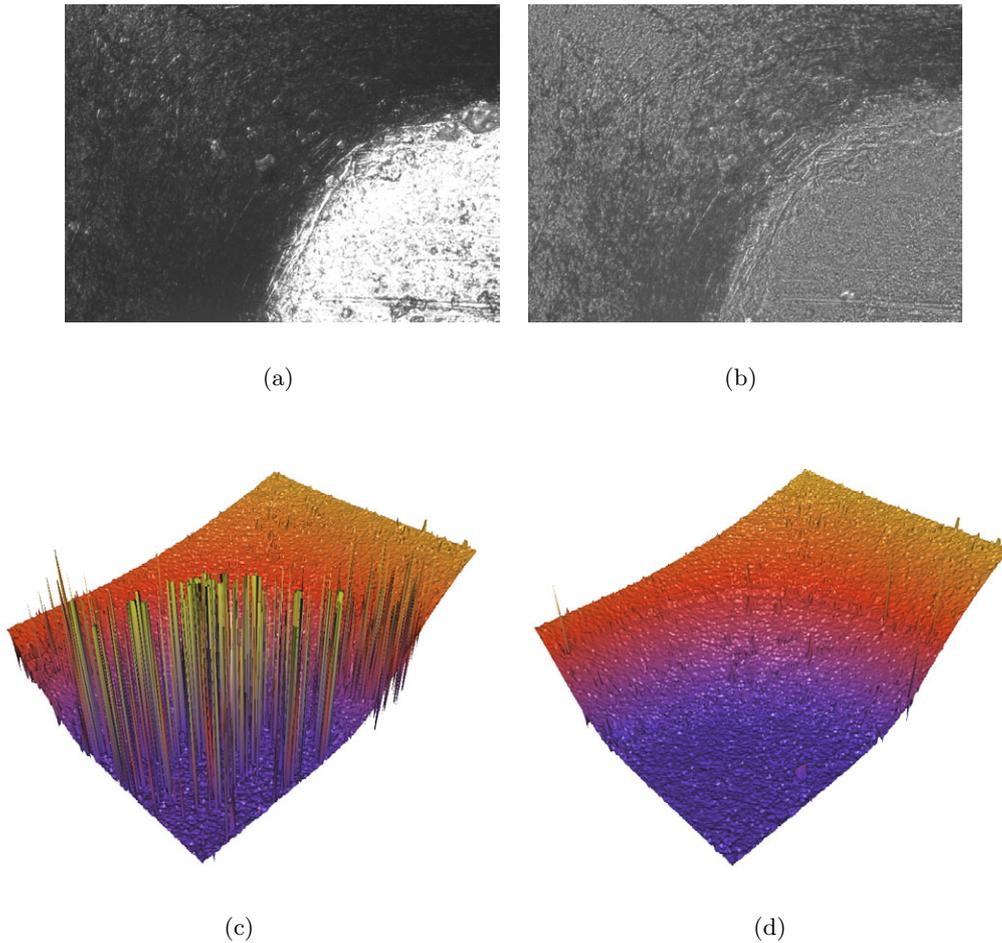


Figure 1.1: SFF reconstruction. All-in-focus texture and reconstruction result for standard SFF (a,c) and SFF with adapted illumination (b,d).

Another limitation of the traditional SFF algorithm is the reliance on enough detectable surface texture on the object. The focus operators in the SFF system measure the focus through high frequency intensity variations. During defocusing these variations get blurred. Therefore, this algorithm is only effective if the focused system has enough

high frequency content to detect a difference between in and out of focus. However, there exist many surfaces that are smooth even at the microscopic level and consequently are lacking in detectable texture. This leads to an inaccurate reconstruction and a dense depth map can not be calculated, as illustrated in Figure 1.2(a,c). Examples of such surfaces are lenses, silicon wafers or all transparent objects. In [48] it was shown, that texture can be forced on low-textured objects with a static patterned filter mask which was placed directly after the light source.

One main disadvantage in SFF depth estimation is its low speed, but due to the small working range and the small depth of field (DOF) in optical microscopy it is one of the most common reconstruction method in optical microscopy. In traditional SFF the depth map is calculated through a focus measurement for every pixel in every acquired image, resulting in a stack of focus measure images. The focus maximum is searched for every focus measurement through every stack image. This huge amount of computational cost leads to a very slow surface reconstruction compared to other methods like shape from stereo (SFS), where the surface is reconstructed by point correspondences in only two images, acquired simultaneously. In [1] and [3] the SFF depth estimation improved by a coarse focus maximum search, e.g. only for every tenth image, to get initial boundaries for calculating the fine focus maximum, but this leads only to marginal savings.

1.2 Motivation

In this work we present different approaches to overcome the problems mentioned in the last section with a microscopic projector-camera system.

With the projector as the light source the illumination intensity can be adapted locally. Therewith, the scene reflectance is compensated to be fully in the representable range of the camera sensor, which shows the same properties with respect to SFF as an image with an extended dynamic range, while the measurement speed and the camera resolution will not decrease. In Figure 1.1 a sample reconstruction of an industrial specimen with homogeneous illumination and with adaptive illumination is illustrated.

To create additional texture this projector-camera system is also used to project additional texture on an arbitrary surface to enhance the natural texture. This projected texture can be used for SFF reconstruction because the projector and the camera have the same focus plane. In our work we calibrate a projector pattern, which creates measurable texture on low textured regions. With this additional texture the sensitivity of the focus measure operator to depth variations and the robustness of the measurement results are

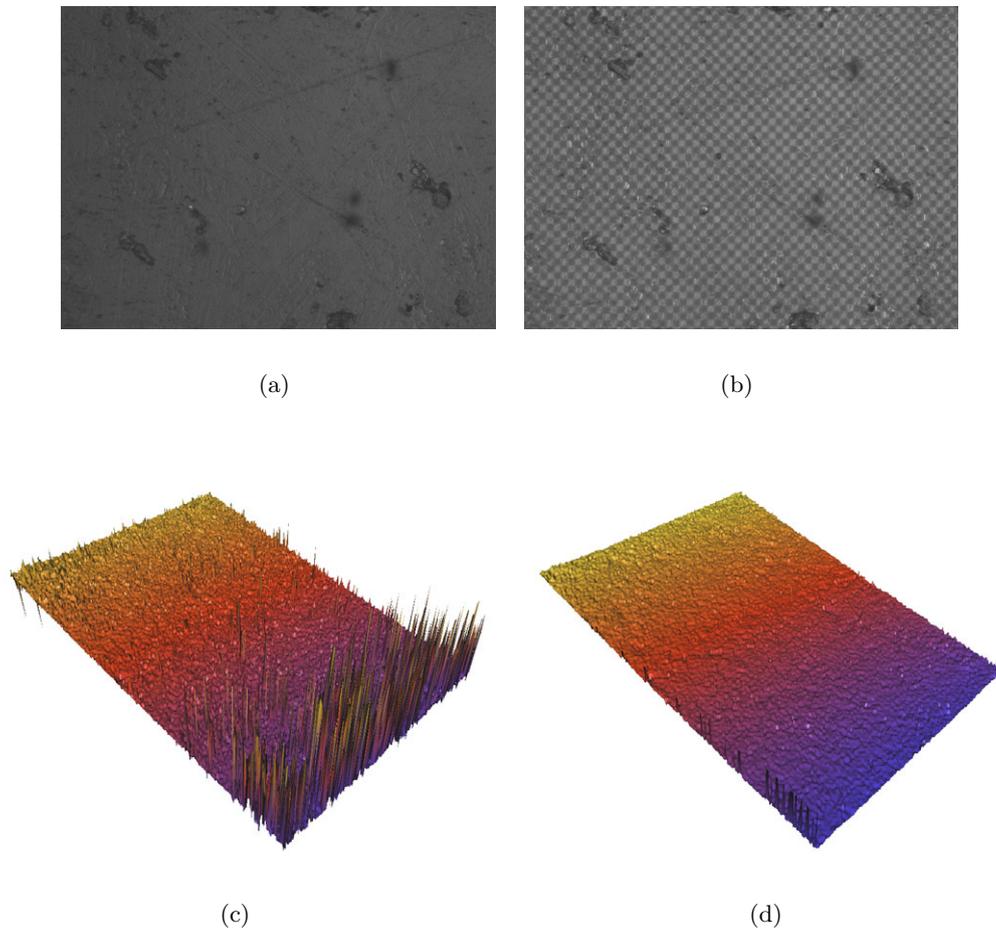


Figure 1.2: SFF reconstruction of a low-textured surface. All-in-focus texture and reconstruction result for standard SFF (a,c) and SFF with projected texture (b,d).

increased. Additionally the intensity can be adjusted according to the natural texture to achieve optimal reconstruction results. In this work we show that this method increases the reconstruction accuracy on low-textured specimens and even specimens with no measurable natural texture can be reconstructed. An example of the performance of texture projection used to enhance the natural object texture is illustrated in Figure 1.2.

Another real advantage of the projector-camera system is to combine the method to compensate the object reflectance and the method to create additional texture to become robust to high differences in the object reflectance and to the absence of texture information. In this work we present a method for the combination of a calibrated texture pattern and the adaption of illumination to reduce the scene dynamic range. Therewith,

the SFF system becomes generally applicable on objects without any previous knowledge of reflectance and texture. On specimen with both low-textured regions and regions with changing reflectance, we show that this method is more robust to wrong focus measurements than the compensation algorithm or the texture projection alone.

As mentioned above, for the 3D reconstruction in industrial quality inspection the measurement time of the reconstruction algorithm plays a major role. Therefore, we invented a new focus measure to speed up the focus maximum search through the image stack. This method is based on locally comparing the image intensities to a calculated all-in-focus image. This all-in-focus image is estimated by a convolution of the integrated image stack with the integral point spread function (IPSF). The IPSF is the blurring function of one surface point in an integral focus image and is pre-calibrated. By a patch-wise normalized cross correlation (NCC) of this texture image with the images from the image stack, the new focus measure is calculated. Because the NCC delivers normalized correlation values between -1 and 1 the focus maximum search can be stopped, if a local maximum above a defined threshold is found. Although the convolution is noise sensitive and the reconstructions may get less accurate, this method delivers a significant speed up of the SFF algorithm.

1.3 Overview

In this work we use an optical microscopic projector-camera system for 3D reconstruction with the SFF algorithm. Therefore we give a general overview of common hardware setups and algorithms for microscopic reconstruction and an introduction of projector-camera systems in relation to this work in Chapter 2. The theoretical background of the 3D reconstruction with SFF is explained in Chapter 3. In this chapter the various focus measurement techniques as well as the different methods for depth interpolation are presented. Additionally, the HDR imaging algorithm and a method for a radiometric compensation are explained. In Chapter 4 the projector-camera system used in this work is presented, where the optical path of the projector and the camera is delineated. Furthermore, both the geometric calibration that maps the projector-camera pixel correspondences and the radiometric calibration that maps the projector-camera intensity correspondences, are presented. In Chapter 5 the methodology of this work is described in detail. This includes the adaptive illumination, the texture projection and a new SFF measurement - the integral shape from focus (ISFF) algorithm. These methods are evaluated concerning accuracy, speed and robustness in Chapter 6. This chapter delivers experiments that shall prove

the applicability of our system for industrial quality measurements. A detailed discussion and an outlook to further work is given in Chapter 7.

Chapter 2

Related Work

Contents

2.1	Microscopic Systems for 3D Reconstruction	8
2.2	Reconstruction based on Small Depth of Field	9
2.3	Projector-Camera Systems	11

The purpose of this project is to reconstruct 3D objects in microscopic dimension for industrial quality inspection. The main target is to enhance an existing reconstruction method to increase speed and accuracy. The hardware setup to achieve this enhancements is a coaxial projector-camera system in optical microscopy, where the projector is used as a locally adaptive illumination source. Due to the small DOF of microscopic lenses the reconstruction method used in this project is the SFF algorithm, whereby a resolution beyond one micrometer is achieved. To give a general view of the possibilities in the field of industrial microscopic reconstruction and projector-camera systems, an overview of related projects in a microscopic operation range is given in this chapter. In Section 2.1 the most relevant hardware solutions and methods for microscopic reconstruction are presented. Section 2.2 outlines a concrete insight in algorithms for object reconstruction for acquisition setups with small DOF, especially concerning light modification and filtering to increase the dynamic range and/or accuracy. The projector-camera system in this work is used to realise local varying illumination intensities, either for compensating the natural object reflection or for increasing the object texture. Therefore an overview of the existing fields of application for projector-camera systems is given in Section 2.3.

2.1 Microscopic Systems for 3D Reconstruction

In this section related methods and microscopic setups for 3D reconstruction in small working ranges, from millimeter down to nanometer range, are presented.

One main type is optical sectioning in confocal microscopes, as presented in Jordan et al. [28]. In confocal microscopy the depth is measured through the intensity of a reflected point light source. Due to blurring, the light energy is distributed over a local circular patch. Therefore, the reflected intensity at the acquisition pinhole is directly related to the degree of focus. Although the measurement for one point is very accurate, the main disadvantage is the restriction to the measurement of only one point at the same time. An overview of applications of confocal microscopy is given in Dixon et al. [15].

Confocal microscopy can also be modelled to reconstruct whole surfaces by measuring point lights at different positions on the object in parallel, often realized with a rotating pinhole mask. Through transverse or axial scanning the confocal measurements can be fused to one dense depth-map, as presented in [28, 64].

Another microscope type for possible reconstructions with a higher magnification (nanometer range) is the scanning electron microscope (SEM) or the atomic force microscope (AFM). Due to the single projection of one electron ray at each scanning step the projection is comparable to confocal microscopy with the difference that it has a very large DOF. This provides the opportunity to use shape from stereo SFS [32] or multi-view [50] methods for 3D reconstruction. A microscopic multi-view setup is vulnerable to occlusions, but the results can be improved through additional shape from shading (SFSh) as presented in [5].

Finally also the most common type, the optical microscope, can be used for surface reconstruction. The traditional optical microscope consists of a single image sensor, e.g. digital sensor or human eye for acquisition and an illumination source, which is usually coaxial. There exist lots of methods for the reconstruction of objects in optical microscopy. Neil et al. [47] presented an optical sectioning approach by projecting a structured light grid on the object. Greenberg and Boyde [17] presented a method for stereo imaging. Here, the object is captured by a single sensor but with different illumination directions. Due to the reflection differences, the acquired images can be used for stereo reconstruction. Real stereo microscopy with two different objectives is limited by the small DOF of optical microscopes. Possible solutions of stereo light microscopy are presented by Kim et al. [29] and more precisely calibrated by Danuser [12].

of defocused object points on the sensor plane are presented in [25, 34].

The SFF method reconstructs the depth through finding a focus maximum, while SFD estimates the depth with the varying image blur. Pentland [51] first used the degree of image blur to calculate a focal gradient when changing the distance from an object to a camera system with fixed intrinsic parameters. From these focal gradients the depth is roughly derived. Grossmann [19] first discovered the possibilities of depth estimation in a scene using measurements of the degree of blur. Based on this method Darrel and Wohn [13] presented the first depth from focus model when using images of one scene with different focus levels and calculated the depth using Laplacian and Gaussian Pyramids. Subbarao [61] acquired images with a different degree of focus with a constant distance from the object to the camera lens, but with changing intrinsic camera parameters, such as changing the focal length of the lens, changing the distance from sensor to lens or changing the lens aperture. Nayar and Nakagawa [44, 45] and Nayar [41] presented a method for SFF reconstruction, where a sequence of images from an object is acquired with different distances between the object and the camera. The focus was calculated with a gradient maximization operator for each pixel in the image stack. Therewith, the depth was determined by finding the focus maximum of each pixel through the image stack. To get a refined focus maximum, the curve progression was interpolated by a Gaussian distribution.

More recently, a possible method for a robust estimation of the image surface out of the focus measures from [45] is the interpolation of a focused image surface (FIS) into the focus image stack. Subbarao et al. [63] first introduced a method, where a piecewise planar surface is fitted into the focus image stack. Yun et al. [72] enhanced this method by replacing the piecewise planar surface with a curved window interpolation using Lagrange polynomials. A different method for FIS optimization into the focus image stack was presented by Asif et al. [3] by approximating the shape of the FIS using a feedforward neural network. The weights of this network are learned such that the focus measure of the surface through the focus stack is maximized. In order to avoid the computation time for training the weights in a neural network, Ahmad et al. [1] presented a method to solve the optimization model (searching optimal focus measure) by a dynamic programming (DP) approach.

Further work is presented for optimization of focus measure response. Noguchi and Nayar [48] presented a method where additional texture is created on low textured surfaces by placing a patterned filter in front of the coaxial light source of the acquisition system

to increase the robustness and sensitivity of the focus measure operator.

Another method to increase the robustness against object reflection artefacts was presented by Nayar and Mitsunaga [43]. By placing an optical mask directly in front of the image sensor array of the camera, a broader range of brightness variations can be captured. The optical mask is a transparent pattern with spatially varying transmittance. This gives adjacent sensor pixels different sensitivity to the scene illumination. Through a high dynamic range (HDR) imaging algorithm this different exposure levels are calculated to one image which measures a wider range of scene radiances, whereas the resolution of the resulting image decreases with the number of different exposure levels.

2.3 Projector-Camera Systems

Our hardware setup consists of a camera for acquisition and a projector used as a locally adaptive light source. In this section an overview of the existing fields of application for projector-camera systems in relation to our project is given. Usually these systems have the purpose of changing the projector image to compensate geometry and reflectance characteristics of the projection screen.

This compensation is used in the context of object reconstruction through structured light by Gu et al. [21] and Koninckx et al. [30, 31] to compensate reflections and shadows of the structured light pattern on the object, which leads to a reconstruction robust against object reflection and ambient light. Another operation area for projector-camera systems is the projection of multimedia content on arbitrary surfaces in [7, 20, 68] and [16, 18, 42], where the geometry and the spatially varying reflectance due to object properties was compensated. Compensation techniques that are content dependent and optimized for human perception have been described in [2, 66].

Compensation can also be applied when using multiple projectors used as a single high-resolution monitor to compensate the overlap of projector images, as described in [27, 35, 52].

Another approach of performing a radiometric compensation is to directly connect the projector input image with the camera output image via acquiring the full light transport between the projector and the camera. This method computes the influence of one projector pixel to the whole camera image. Seitz et al. [56] first explored the impact of inter-reflections in photographs of real scenes with unknown and arbitrary geometry and illumination for scenes with arbitrary spatially varying bidirectional reflectance distribution functions (BRDFs). Sen et al. [57] presented a method to calculate the full

light transport used for dual photography to relight and synthesize a scene using a light transport matrix. This method was adapted by Wetzstein et al. [67–69] where this light transport was used for radiometric compensation of screen imperfections.

Chapter 3

Theoretical Background

Contents

3.1	Shape From Focus	13
3.2	High Dynamic Range Photography	24
3.3	Radiometric Compensation	30
3.4	Summary	34

In this chapter an overview of the algorithmic and theoretical background for our work is given. The main reconstruction method is the shape from focus (SFF) algorithm, described in Section 3.1 for 3D reconstruction. To avoid illumination artefacts with a static illumination source a method called high dynamic range imaging is presented in Section 3.2. Reflection artefacts are compensated by adding a projector to the acquisition system used as light source, presented in Section 3.3.

3.1 Shape From Focus

Our method for 3D reconstruction is based on a method called SFF, according to Nayar [41, 44] and Nayar and Nakagawa [45]. On the supposition that in practice all objects have a rough surface at some level of detail, high frequency texture indicates focused image parts. On the other hand the defocus of an imaging system plays the role of a low-pass filter. Since optical microscopy has a very small depth of field (DOF), which causes high focus variances at slight depth differences, SFF is the most popular method for reconstruction.

The method requires a stack of images acquired at different distances from an object

As illustrated in Figure 3.1, the defocused patch of our sample point increases by the distance between the sensor and the image plane. Assume that the distance between sensor and image plane is d . The radius of the patch r is calculated as

$$r = \frac{dR}{i}, \quad (3.2)$$

where R is the radius of the lens. The uniform distribution of the light energy over the circular patch with radius r referred to as the PSF, is ideally modelled with the pillbox function

$$h(x, y) = \begin{cases} \frac{1}{\pi r^2} & \text{if } x^2 + y^2 \leq r^2 \\ 0 & \text{otherwise.} \end{cases} \quad (3.3)$$

But in the presence of optical aberrations the PSF deviates from the pillbox function and is often approximated by a bivariate Gaussian model depending on i and o as described in Pentland [51] and Nayar [44]. An example of the bivariate Gaussian distribution model is illustrated in Figure 3.2.

The blurred image I_d is calculated by convolution of the focused image I_f with a Gaussian blurring function h by

$$I_d(x, y) = h(x, y) * I_f(x, y), \quad \text{where } h(x, y) = \frac{1}{2\pi\sigma_h^2} e^{-\frac{x^2+y^2}{2\sigma_h^2}} \quad (3.4)$$

In the frequency domain the defocusing process is described as

$$I_D(u, v) = H(u, v)I_F(u, v), \quad \text{where } H(u, v) = e^{-\frac{u^2+v^2}{2}\sigma_h^2} \quad (3.5)$$

I_F , I_D and H are the Fourier transforms of I_f , I_d and h .

In the frequency domain it can be seen that in $H(u, v)$ the attenuation of the high frequencies is dependent on the spread parameter σ_h of the Gaussian function. This parameter increases with the distance of the sensor to the focused image plane d . When considering that the distance between the lens and the sensor is constant in an optical image acquisition system, one way to defocus an image is to move the object with respect to the object plane. Li and Agathoklis [34] presented a detailed measurement and analysis of the PSF and the optical transfer function (OTF) in a microscope.

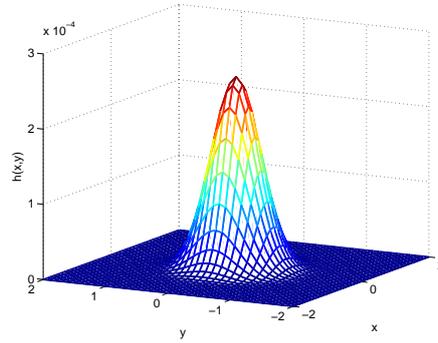


Figure 3.2: Bivariate Gaussian distribution model $h(x, y)$ for approximating the PSF.

3.1.2 Shape From Focus

In the system illustrated in Figure 3.3 the unknown object has to be placed in front of the sensor plane, to reconstruct depth with SFF. A translational stage moves the object, starting from the reference plane in Δd steps along the optical axis. As mentioned in Section 3.1.1, the optical system defines one plane where each object point is in focus. The position d of the translational stage and the position d_f of the focused plane are known in every step. To illustrate the reconstruction in detail we take a closer look on a single object point S . Considering d is increasing and the translational stage is moving towards the focused plane, the degree of focus of the object point S on the sensor will also increase, until it is directly "in focus". An additional movement of the stage towards the

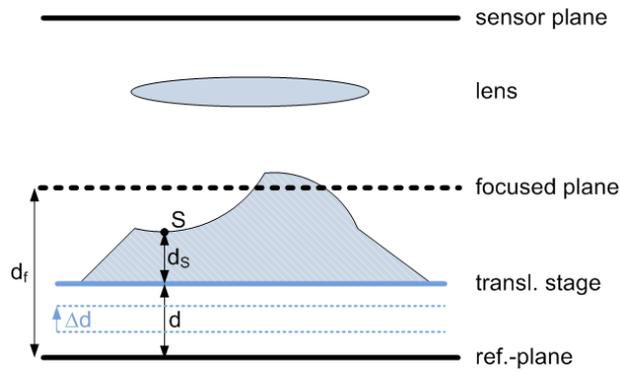


Figure 3.3: System setup for shape from focus acquisition, according to [45].

sensor will again decrease the degree of focus. As mentioned above, the stage translation d and the distance between the reference and the focused plane d_f is known. We can now easily calculate the height of our sample point S in our object with $d_s = d_f - d$.

The depth of every point on the object is calculated independently, resulting in a dense surface reconstruction.

3.1.3 Focus Measures

Usually focus measure algorithms are aimed on finding high frequency parts in an image. The algorithms for detecting these high frequency parts analyse the intensity variations in sub-windows and calculate the focus measure for the center of the sub-window. In [23] and [37] a good overview of the different methods is given.

Laplacian Focus Measure

The Laplacian operator, as presented in [53], calculates the second derivative, which gives the gradient magnitude and highlights therefore the intensity variations:

$$\text{FM}_L(x, y) = \frac{\partial^2 I(x, y)}{\partial x^2} + \frac{\partial^2 I(x, y)}{\partial y^2}. \quad (3.6)$$

The focus measure for the point x, y is then calculated by the sum of the second derivatives of the image grey-values $I(x, y)$ in x and y direction. A discrete approximation of the second derivative is given by convolving the image with the 3×3 Laplace Operator:

$$\mathbf{H} = \begin{pmatrix} 0.5 & 0 & 0.5 \\ 0 & -2 & 0 \\ 0.5 & 0 & 0.5 \end{pmatrix}. \quad (3.7)$$

When using the simple Laplace operator as focus measure operator in context of textured images, the second derivatives in x and y directions can have opposite signs. In this case they would cancel each other and the focus measure yields no response. To overcome this problem a slightly modified Laplace measure, called Modified Laplacian, was introduced.

Modified Laplacian

To avoid the cancellation of Laplace components when $I_{xx} = -I_{yy}$ we the second derivatives of (3.6) are squared:

$$\text{FM}_{ML}(x, y) = \left(\frac{\partial^2 I(x, y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 I(x, y)}{\partial y^2} \right)^2. \quad (3.8)$$

For the discrete model a separate convolution of the Modified Laplacian operator with the image in x and y direction is required. The Modified Laplacian is computed as:

$$\text{FM}_{ML} = (2I(x, y) - I(x - \Delta x, y) - I(x + \Delta x, y))^2 \quad (3.9)$$

$$+ (2I(x, y) - I(x, y - \Delta y) - I(x, y + \Delta y))^2 \quad , \quad (3.10)$$

where Δx and Δy are the pixel step size in x and y direction. The discrete operator is a 3×1 respectively 1×3 filter mask for the two dimensions:

$$\mathbf{h}_x = \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} \quad \mathbf{h}_y = \begin{pmatrix} -1 & 2 & -1 \end{pmatrix} \quad (3.11)$$

Sum of the Modified Laplacian

For rich textured images with high variability at each pixel, a focus measure can be calculated with the Modified Laplacian for a single pixel. Problems occur when the operator has to deal with weakly textured images. Nayar [44] first introduced a focus measure as the sum of modified Laplacian values in a local window to improve the robustness:

$$\text{FM}_{SML}(x_0, y_0) = \sum_{p(x,y) \in U(x_0, y_0)} \left(\frac{\partial^2 I(x, y)}{\partial x^2} \right)^2 + \left(\frac{\partial^2 I(x, y)}{\partial y^2} \right)^2 \quad (3.12)$$

Variable $p(x, y)$ declares a pixel in the local neighborhood U around the center coordinates x_0, y_0 . Typically a small window size of 3×3 or 5×5 is used.

Tenengrad Focus Measure

Another popular method for calculating the degree of focus is the Tenengrad Focus measure, first introduced by Tenenbaum [65]. It first calculates the approximated gradients in x and y direction by applying Sobel operators (see (3.13)) in the horizontal (\mathbf{S}_x) and vertical (\mathbf{S}_y) direction separately.

$$\mathbf{S}_x = \begin{pmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{pmatrix} \quad \mathbf{S}_y = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} \quad (3.13)$$

To avoid the cancellation of the x and y components the sum of squared responses after applying the Sobel masks is built, similar to the calculation of the FM_{ML} , as declared in [70]. With the additional requirement for robustness at weak-textured images the summation in a local window around the center-pixel (x_0, y_0) is built:

$$FM_T(x_0, y_0) = \sum_{p(x,y) \in U(x_0, y_0)} G_x(x, y)^2 + G_y(x, y)^2, \quad (3.14)$$

whereas G_x and G_y are the convolutions of the input image $I(x, y)$ with the Sobel masks in x and y direction.

Variance Focus Measure

In the case of a focused image region the variance of gray-values is higher than in case of unfocused regions. Therefore the degree of focus can be measured out of the variance in a local image region. The resulting focus criterion function is given by:

$$FM_{Var}(x_0, y_0) = \frac{1}{N-1} \sum_{p(x,y) \in U(x_0, y_0)} (I(x, y) - \mu_{U(x_0, y_0)})^2, \quad (3.15)$$

where $I(x, y)$ are the gray values in the image region $U(x_0, y_0)$ around the pixel (x_0, y_0) and $\mu_{U(x_0, y_0)}$ denotes the mean of the grey values in U .

Mean Focus Measure

Helmlí and Scherer [23] first presented a method for calculating the focus measure from the ratio of the mean grey value to the center grey value. The principle behind this method is similar to the variance focus measure: If an image is getting sharper the variance of the grey values increases. A ratio of 1 implies a constant gray-level or absence of texture. The ratio is higher in the case of high variations.

$$FM_{mean}(x_0, y_0) = \sum_{p(x,y) \in U(x_0, y_0)} \begin{cases} \frac{\mu_{U(x_0, y_0)}}{I(x, y)} & \mu_{U(x_0, y_0)} > I(x, y) \\ \frac{I(x, y)}{\mu_{U(x_0, y_0)}} & \text{else.} \end{cases} \quad (3.16)$$

To be more robust against weak-textured images, the focus measure is summed in a local window.

Focus Measure based on Chebyshev Moments

A relatively new method for measuring image focus is computing the focus based on discrete orthogonal Chebyshev moments. The low and high spatial frequencies of an image are represented as the low- and high-order Chebyshev moments. Consequently, the focus measure itself is calculated out of the ratio of the norm of the high-order moments to the low-order moments, presented by Yap and Raveendran [71]. The Chebyshev moment of order $(m + n)$ for an image $I(x, y)$ with size $M \times N$ is therefore defined as:

$$T_{m,n} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} \tilde{t}_m(x; M) \tilde{t}_n(y; N) I(x, y) \quad (3.17)$$

where $\tilde{t}_m(x; M)$ and $\tilde{t}_n(y; N)$ are the normalized Chebyshev polynomials defined by

$$t_m(x; M) = \sum_{k=0}^{M-1} a_{k,m} x^k, \quad (3.18)$$

$$\tilde{t}_m(x; M) = \frac{t_m(x; M)}{\sqrt{\rho(m; M)}} \quad \text{and} \quad \tilde{t}_n(y; N) = \frac{t_n(y; N)}{\sqrt{\rho(n; N)}} \quad (3.19)$$

Term $\rho(m; M)$ is the normalisation for

$$\rho(m, M) = (2n)! \binom{N+n}{2n+1}. \quad (3.20)$$

To speed up the computation time, the three-term recurrence relation can be used to calculate $t_n(x, N)$ with

$$t_n(x; N) = [(2n-1)(2x-N+1)t_{n-1}(x; N) - (n-1) \quad (3.21)$$

$$\times (N^2 - (n-1)^2)t_{n-2}(x; N)]/n \quad (3.22)$$

where $n = 2, 3, \dots, N-1$ and $t_0(x; N) = 1, t_1(x; N) = 1 - N + 2x$.

3.1.4 Depth Estimation - Interpolation Methods

As mentioned in Section 3.1.2, a stack of images of the object with different distances d to the optical system is acquired, whereas distance between two consecutive images Δd is known. To get a focus vector $\mathbf{F}(d)$, the focus values $f(x, y, d)$ for every image pixel

(x, y) over the whole acquired image stack ($\forall d \in D$ where D is the number of images) is calculated. This vector can be calculated for every pixel separately, therefore we will focus our attention on a single image point. The discrete number of frames and the discrete stage displacement between two images results in a loss of depth accuracy.

In the following section we will describe the main techniques to get a maximum approximation of the focus vector $\mathbf{F}(d)$.

Maximum Method

A simple approach to find a coarse maximum of the focus vector is to take the discrete depth index d where the focus vector $F(d)$ has its maximum, as illustrated in Figure 3.4. The maximum is therefore calculated as

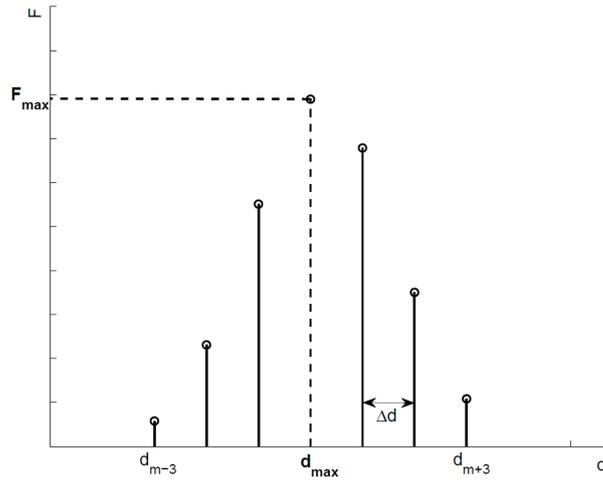


Figure 3.4: Maximum Method. Finding the depth of the maximum focus through the index of the discrete maximum value in the focus vector.

$$F_{max} = \max[F(d)], \text{ where } d = 1, \dots, D \quad (3.23)$$

$$d_{max} = d : F(d) = F_{max}. \quad (3.24)$$

Linefit Method

An depth interpolation method presented by Malik and Choi [36] is the line fitting algorithm. This method first identifies the discrete focus maximum F_{max} . With this maximum it separates the focus vector into two datasets $DS_1(t)$ and $DS_2(t)$, which contain j elements

before and after d_{max} including F_{max} , calculated as:

$$DS_1(t) = F(d_{max} - t), \text{ where } t = 0, 1, \dots, j \mid j < d_0 \quad (3.25)$$

$$DS_2(t) = F(d_{max} + t), \text{ where } t = 0, 1, \dots, j \mid j < d_{end} \quad (3.26)$$

After generating the two datasets a least squares linear regression of DS_1 and DS_2 is applied to approximate two lines. The approximated new focused depth $d_{linemax}$ is calculated as

$$d_1 = |d_i - d_c|, \quad d_2 = |d_i - d_{max}| \quad (3.27)$$

$$d_{sum} = d_1 + d_2 \quad (3.28)$$

$$d_{linemax} = w_1 d_c + w_2 d_{max}, \quad \text{where } w_1 = \frac{d_2}{d_{sum}}, \quad w_2 = \frac{d_1}{d_{sum}}. \quad (3.29)$$

The variable d_i is the intersection distance of the two lines and d_c is the closest point in the dataset DS_1 and DS_2 to d_i . With w_1 and w_2 the sum of the points is weighted depending on their distance to the intersection point. In Figure 3.5 the maximum interpolation

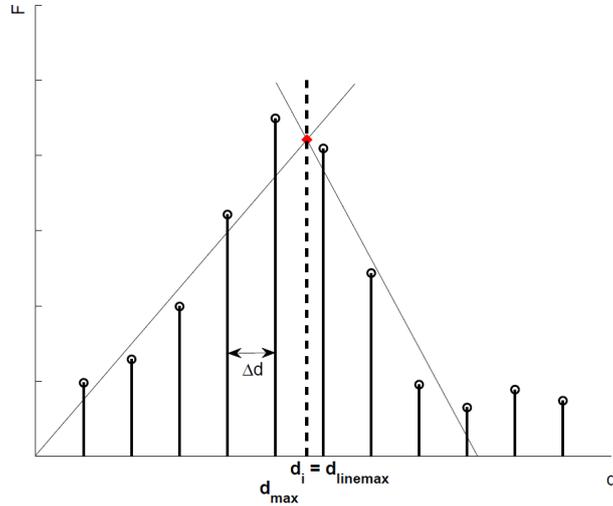


Figure 3.5: Linefit method. The depth is calculated image focus based on the linear regression model.

through the focus vector calculated with the linefit method is illustrated.

Gaussian Interpolation

In experimental verifications by Subbarao [62] and Pentland [51] it was shown that the peak of $F(d)$ can be approximated by a Gaussian curve, with mean value d_{max} and standard deviation σ_F . To find an approximate focus maximum a Gaussian interpolation on the focus vector is applied. To save computation time only three focus measures are used to interpolate the Gaussian, namely F_{m-1} , F_m and F_{m+1} , that lie on the largest mode of $F(d)$, such that $F_m > F_{m-1}$ and $F_m > F_{m+1}$. The focus measure vector can be

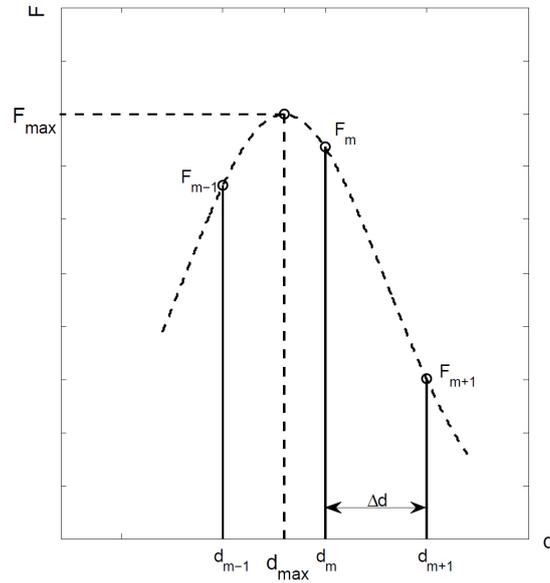


Figure 3.6: Approximation of the focus maximum by interpolating focus points with a Gaussian curve, according to [45]

expressed using the Gaussian model with:

$$F(d) = F_{max} \exp \left\{ -\frac{1}{2} \left(\frac{d - d_{max}}{\sigma_F} \right)^2 \right\} \quad (3.30)$$

where d_{max} is the mean value and σ_F the standard deviation of the Gaussian distribution. To get d_{max} (3.30) is rewritten as:

$$\ln(F(d)) = \ln(F_{max}) - \frac{1}{2} \left(\frac{d - d_{max}}{\sigma_F} \right)^2 \quad (3.31)$$

Thus, after substituting the function with the three sample points, three equations can be obtained. This equation system is solved for d_{max} and σ_F to:

$$d_{max} = \frac{(\ln(F_m) - \ln(F_{m+1}))(d_m^2 - d_{m-1}^2) - (\ln(F_m) - \ln(F_{m-1}))(d_m^2 - d_{m+1}^2)}{2\Delta d((\ln(F_m) - \ln(F_{m-1})) + (\ln(F_m) - \ln(F_{m+1})))} \quad (3.32)$$

$$\sigma_F = \frac{(d_m^2 - d_{m-1}^2) + (d_m^2 - d_{m+1}^2)}{2((\ln(F_m) - \ln(F_{m-1})) + (\ln(F_m) - \ln(F_{m+1})))} \quad (3.33)$$

With d_{max} and σ_F found in (3.32) and (3.33), F_{max} is calculated by substituting one of the sample points F_m, d_m in Equation (3.30) as:

$$F_{max} = \frac{F_m}{\exp\left\{-\frac{1}{2}\left(\frac{d_m - d_{max}}{\sigma_F}\right)^2\right\}} \quad (3.34)$$

With the found result we can characterize the image content at point (x, y) . If the peak of the Gaussian model is high and the standard deviation is small it can be implied that there is high texture content in the vicinity of the point (x, y) and vice versa. Therefore one can also divide the image content into regions of different texture content.

3.2 High Dynamic Range Photography

Image processing often suffers from the problem of a limited dynamic range. When acquiring an image, a range of irradiance values that are of interest and a suitable exposure time have to be chosen to measure these irradiance values correctly with the image sensor. Scenes with illumination artefacts due to different reflections or different luminance within a scene have extreme differences in irradiance values. These differences are impossible to be captured without under- or overexposed parts in the image, as shown in Figure 3.7. A straight forward idea to solve this problem would be to capture more than one image of a scene with different exposure times. But taking images with different exposure times leads to the problem that the brightness values are rarely true measurements of the relative irradiance in the scene. It has to be assumed that the mapping between irradiance and pixel intensities is non-linear due to several non-linear mappings in the photographic process in conventional cameras.

A method for recovering HDR radiance maps from photographs was first presented in [38] and [14]. There, image data is composed from one camera with different exposure



Figure 3.7: Scene with illumination artefacts. Due to illumination artefacts the scene produces a wide range of brightness variations. It is not possible for the image sensor to capture this high dynamic range.

times or image data is merged from images obtained by different cameras. This method provides a convenient and robust way of determining the overall response curve of an imaging process, allowing also images from different processes to be used consistently as radiance maps.

Nayar and Mitsunaga [43] first presented a method for acquiring high dynamic range images by placing an optical mask directly in front of the image detector array of a camera. The mask is divided into patterns with different spatially varying pixel transmittance. This gives adjacent pixels different sensitivity to the scene illumination. Hence, one high resolution image can be divided into images with a smaller resolution and different exposures, from which the HDR radiance map can be calculated.

In the following sections the algorithm to recover the camera response function and a method of reconstructing the high dynamic range irradiance image from multiple images are presented.

3.2.1 Recovering the Camera Response Function

In this section a method for calibrating the characteristic curve of a camera is presented. The characteristic curve maps incoming radiance values to pixel intensities, and hereby models the image acquisition process. To understand the non-linearities of this function the image acquisition process has to be considered first (see Fig. 3.8). The exposure X is the product of the irradiance E at the film and the exposure time Δt . In this context the non-linearities in the acquisition process are represented in a characteristic curve which is only determined by the exposure and therefore by $X = E\Delta t$. This assumption only

breaks down for extrema in the exposure time (very large or very low Δt).

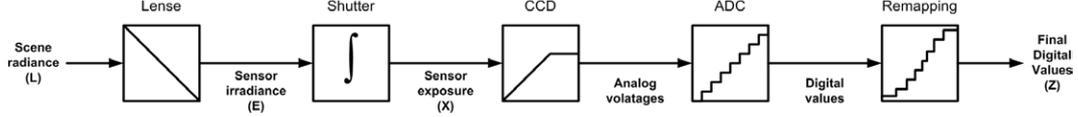


Figure 3.8: Process of image acquisition of digital cameras. Non-linearities can occur during exposure, scanning, digitization and remapping [14].

After the sensor output is digitized, we get the final digital values Z , which are calculated by a discrete non-linear function f of the original exposure X at one pixel. Knowing that f is monotonically increasing, we can compute X at each pixel with $X = f^{-1}(Z)$. The irradiance can be recovered as $E = X/\Delta t$. It can be assumed that the function f is a composition of the characteristic curve of the image sensor and the non-linearities introduced by further processing steps.

To recover f , we have to take a number of digital sample photographs captured with a static camera with different known exposure times. We assume that the irradiance values at each pixel are constant through this process. The reciprocity equation is written as:

$$Z_{i,j} = f(E_i \Delta t_j), \quad (3.35)$$

where i is the pixel index of image j . As mentioned above, it is possible to invert this equation to

$$f^{-1}(Z_{i,j}) = E_i \Delta t_j. \quad (3.36)$$

For linearisation we define $g = \ln f^{-1}$ and get

$$g(Z_{i,j}) = \ln f^{-1}(Z_{i,j}) = \ln E_i + \ln \Delta t_j. \quad (3.37)$$

The pixel intensities $Z_{i,j}$ and the exposure times Δt_j are known. The unknowns of (3.37) are therefore the irradiance values E_i and the inverse response function g , which has to be smooth and monotonic. Letting N be the number of pixels and P the number of images, we now have a set of $N \times P$ equations. To recover the function g and the irradiances E_i the set of equations has to be rewritten to a least-squared error solution to:

$$\operatorname{argmin}_g \left\{ \sum_{i=1}^N \sum_{j=1}^P [g(Z_{i,j}) - \ln E_i - \ln \Delta t_j]^2 + \sigma \sum_{z=Z_{min}+1}^{Z_{max}-1} g''(z)^2 \right\}. \quad (3.38)$$

Z_{min} and Z_{max} are the minimum and maximum pixel intensity values (e.g. for 8bit $0 - 255$). The first term of (3.38) includes a set of equations from (3.37) in a least squares sense, whereas the second term ensures that the function g is smooth, through penalizing the sum of squared second derivative of g . The variable λ is a scalar weighting term for the smoothness relative to the data fitting term. It should be chosen dependent on the amount of noise expected in the $Z_{i,j}$ measures. Because of the discrete setting the second derivative is approximated by

$$g''(z) = g(z - 1) - 2g(z) + g(z + 1). \quad (3.39)$$

The over determined equation system 3.38 is solved using the singular value decomposition (SVD). An additional offset factor α is introduced to set the unit exposure to the pixel intensity in the middle of the range between Z_{min} and Z_{max} , mathematically expressed as:

$$g(Z_{mid}) = 0, \quad \text{where } Z_{mid} = \frac{1}{2}(Z_{max} - Z_{min}). \quad (3.40)$$

Therefore, the logarithmic irradiance $\ln E_i$ and the function g are replaced by $\ln E_i + \alpha$ and $g + \alpha$. Further on, since $g(z)$ typically has a steep slope near Z_{min} and Z_{max} , it should be expected that $g(z)$ is less smooth and fits the data more poorly near the minimum and maximum values. A weighting function w has to be introduced, which weights pixel intensities near the middle intensity Z_{mid} the most and decreases towards the extremes:

$$w(z) = \begin{cases} z - Z_{min} & \text{for } z \leq \frac{1}{2}(Z_{min} + Z_{max}) \\ Z_{max} - z & \text{for } z \geq \frac{1}{2}(Z_{min} + Z_{max}). \end{cases} \quad (3.41)$$

Plugging in these weights in (3.38), the quadratic least squares solution becomes:

$$\operatorname{argmin}_g \left\{ \sum_{i=1}^N \sum_{j=1}^P \{w(Z_{i,j}) [g(Z_{i,j}) - \ln E_i - \ln \Delta t_j]\}^2 + \sigma \sum_{z=Z_{min}+1}^{Z_{max}-1} [w(Z_{i,j})g''(z)]^2 \right\} \quad (3.42)$$

Debevec et al. [14] evaluated that a set of $P = 11$ photographs with a choice of $N = 50$ pixel is more than adequate for calibrating g over a pixel range of 255 ($Z_{max} - Z_{min}$). In Figure 3.9 twelve images of a sample scene with different exposure times are illustrated. The pixel locations for reconstruction should be chosen to be evenly distributed between Z_{min} and Z_{max} . In Figure 3.10 the response function recovered from the sample scene is illustrated.



Figure 3.9: HDR Grayscale photographs. Twelve images of an sample scene with different exposure times Δt are acquired for calibrating the response function g .

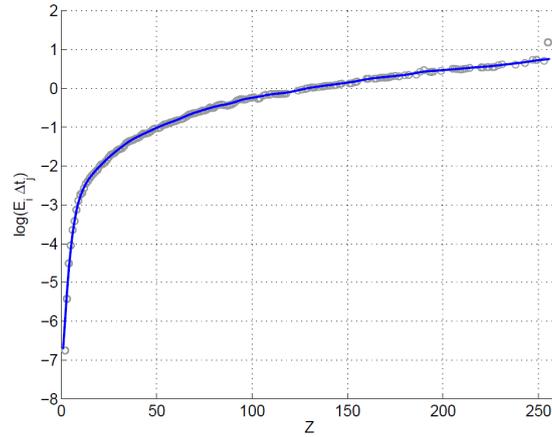


Figure 3.10: Example response function. The response function g is reconstructed from $P = 12$ photographs and $N = 50$ pixel from each photograph. The \circ symbol marks the pixel samples of the input images.

3.2.2 HDR irradiance Reconstruction

With the camera response function g , the high dynamic range radiance map is calculated easily through (3.37), assuming the exposure times Δt_j are known:

$$\ln E_i = g(Z_{i,j}) - \ln \Delta t_j. \quad (3.43)$$

To reduce noise in the recovered irradiance values, all the available exposures should be used. The weighting function from (3.41) is reused, to give higher weight to exposures in which the pixel intensity is closer to the middle of the response function. The HDR irradiance image is reconstructed through

$$\ln E_i = \frac{\sum_{j=1}^P w(Z_{i,j}) (g(Z_{i,j}) - \ln \Delta t_j)}{\sum_{j=1}^P w(Z_{i,j})}. \quad (3.44)$$

To fully recover the radiance map, the number of images should be chosen, so that every part of the scene is imaged correctly in at least one photograph. The HDR reconstruction



Figure 3.11: Image reconstruction with HDR imaging. A HDR image reconstructed using $N = 5$ images.

result with $N = 5$ differently exposed images is shown in Figure 3.11.

3.2.3 Disadvantages of HDR Imaging

Although an image including a high dynamic range can be generated with HDR imaging it also has its problems. One main problem - especially in amateur photography - is that HDR imaging requires a static camera and a static scene during acquisition. Another problem relevant to industrial inspection, where speed is important, is the number of photographs. Depending on the dynamic range in the scene, two or more images need to be acquired, and the quality increases with the number of images, so the time for acquisition and calculation of HDR images also rises.

3.3 Radiometric Compensation

Acquiring images of an object with a constant illumination source often leads to the problem of high variations of the reflectance on the object surface. This reflectance depends on the surface texture, the material properties and the angle of the surface to the optical axis. It can lead to over and under saturated image-parts. The concept of radiometric compensation (RC) addresses this problem. With the assistance of a projector, the RC tries to compensate spatially varying reflectance properties to get a desired camera image. Furthermore, it enables the projector to display images onto an arbitrary surface, while the influence of the projection surface itself is compensated.

Before calculating a compensation pattern for the projector, a radiometric model has to be built to be able to determine the radiometric chain from projector values to image intensities, as described in Section 3.3.1. Further, the geometric mapping between camera and projector pixels has to be known. In Section 3.3.2 a method for calculating the radiometric model is described, whereas the simple radiometric compensation through known geometric correspondences is described in Section 3.3.2. Another approach is to calculate the radiometric compensation through inverse light transport. Here, a model of the influence of projector pixel to camera pixel is calibrated directly, as described in Section 3.3.4.

3.3.1 The Projector-Camera Radiometric Chain

As elaborated in Section 3.2, the mapping between irradiance and pixel intensities in a camera is typically non-linear. Similar, the radiant flux generated by the projector is also non-linear with respect to the applied pixel intensity. According to Koninckx et al. [30, 31] these non-linearities tend to boost the perceptual quality, but concurrently complicate the calculation of correct projector irradiance from given pixel intensities, as compared to a linear model.

There, the projector-camera radiometric chain for a single projector pixel was developed (see Fig. 3.12(a)). The same model is used for all other points equally. The response curves for camera and projector map an irradiance value E to a discrete pixel intensity Z (8bit: $f : \mathbb{R}^+ \rightarrow [0, 255]$). The transition between the projector pixel intensity to the projected irradiance value is calculated through:

$$E_p = f_p^{-1}(Z_p). \quad (3.45)$$

Note that in this case, the non-linear mapping between the input and output of the projector includes multiple non-linearities in the projection process and also depends on the wavelength λ . The projector screen irradiance is then reflected by an object with the reflectance properties R . The reflected irradiance in the viewing direction of the camera can be written as

$$E_c = R(E_p). \quad (3.46)$$

The reflected irradiance E_c is measured by the camera for an exposure time Δt . The discrete camera pixel intensities are calculated with

$$Z_c = f_c(E_c \Delta t), \quad (3.47)$$

where f_c is the non-linear mapping between camera irradiance E_c and the acquired pixel intensities Z_c . With this correspondences, we have built a relation between the input

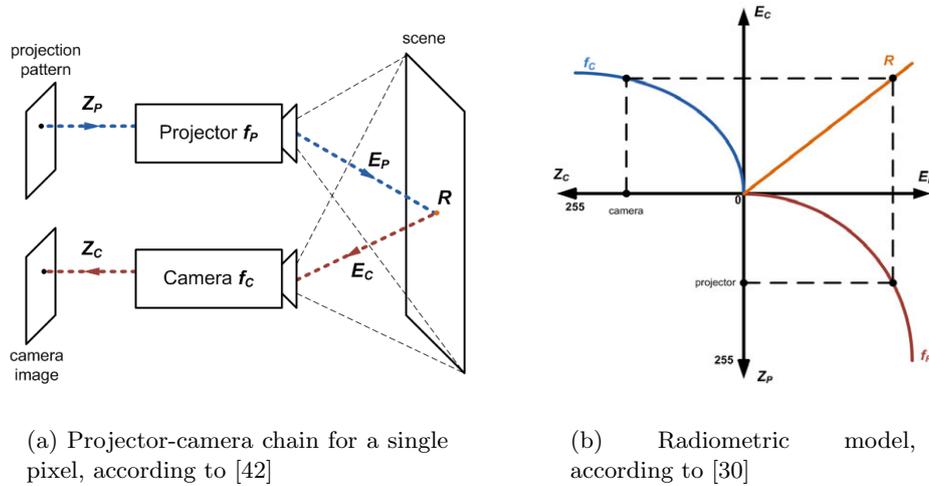


Figure 3.12: Projector-camera radiometry model. The projection value is transformed by f_p^{-1} , reflected by the scene and acquired by the camera by transformation with f_c .

pixel-values of the projector to the output pixel-values of the acquired camera image (see Fig. 3.12(b)).

3.3.2 Calibration of the Projector Response Function

The determination of the camera response function f_c is done by the algorithm presented in Section 3.2, respectively in [14]. To calculate the projector response function f_p we go the inverse way from the camera image Z_c to the projector irradiance E_p :

$$E_p = R^{-1} \left(\frac{f_c^{-1}(Z_c)}{\Delta t} \right). \quad (3.48)$$

The reflectance model R is eliminated by projecting a known uniform pattern onto a planar diffuse surface. The projector response f_p is estimated by mapping projected uniform patterns with different intensity levels with their corresponding projector irradiances calculated with (3.48). The exposure time Δt has to be adjusted properly to avoid over or under exposed camera images.

3.3.3 Compensation Algorithm through Geometric Correspondence

In this section the radiometric compensation through the camera and projector response functions is developed for a direct geometric projector-camera pixel correspondence. It is assumed that each projector pixel only influences one corresponding camera pixel. The complete mapping from camera to projector intensities is now represented using a single non-linear response function including the camera and the projector response function:

$$Z_c = h(R, Z_p) = f_c (R (f_p^{-1}(Z_p)) \Delta t) \quad (3.49)$$

Projector pixel intensities for a desired camera image brightness can be theoretically determined through h with an unknown reflectance R which should be compensated.

When projecting a novel display pattern $Z_{p,n}$ the camera acquires an uncompensated image $Z_{c,n}$. The compensation requires that the measured image exactly equals the original projector image. The calibrated inverse response h^{-1} is used to compute the compensation image $\tilde{Z}_{p,n}$. This image is projected to get the final compensated output image $\tilde{Z}_{c,n}$.

3.3.4 Compensation through Light Transport Matrix

In this section a method of a radiometric compensation for a projector-camera system, which has no direct mapping between projector and camera pixels is presented. One method to calculate the correspondence between the two coordinate frames is through a geometric calibration of the pixel correspondences as presented in Nayar et al. [42, 46]. In

this method the pixel correspondence is modelled through a transformation of second-order polynomials between the two coordinate frames. The final mappings between projector and camera coordinate frame are stored as look-up tables, where each point in one domain is used as an index to obtain the corresponding point in the other. A problem with this direct mapping is that global illumination effects such as reflections, scattering, refractions, dispersion or diffraction are ignored. One pixel only affects its corresponding pixel in the other system.

Another approach of performing a radiometric compensation with an unknown projector-camera pixel correspondence is to directly connect the projector input image with the camera output image via acquiring the full light transport between the projector and the camera. This method computes the influence of each projector pixel to the whole camera image a light transport matrix (LTM), as used in [67–69].

The camera image \mathbf{Z}_c has a resolution of $m \times n$ and the projector pattern \mathbf{Z}_p a resolution of $p \times q$. Therefore, the LTM \mathbf{T} has a size of $mn \times pq$. To calculate the matrix, Sen et al. [57] presented a method where a set of illumination patterns is projected onto the scene and captured with the camera. The projector patterns are designed to refine the matrix hierarchically. With the calculated light transport matrix \mathbf{T} the camera image can then be calculated with

$$\tilde{\mathbf{z}}_c = \mathbf{T}\tilde{\mathbf{z}}_p + \mathbf{e}, \quad (3.50)$$

where $\tilde{\mathbf{z}}_c$ is the vectorized camera image, $\tilde{\mathbf{z}}_p$ is the vectorized projector image and \mathbf{e} is a vectorized camera image at the projector's black level. The radiometric compensation can than easily be calculated by solving 4.11 for $\tilde{\mathbf{Z}}_p$ where $\tilde{\mathbf{Z}}_c$ is the desired camera image by

$$\tilde{\mathbf{z}}_p = \mathbf{T}^{-1}(\tilde{\mathbf{z}}_c - \mathbf{e}). \quad (3.51)$$

The biggest problem when calculating the RC through the LTM is its enormous size ($mn \times pq$). \mathbf{T} has a high level of sparsity but this does not infer that its inverse \mathbf{T}^{-1} has also a high level of sparsity. When calculating the inverse matrix the memory of normal personal computer comes to its limitations (with full-hd projector, a 2megapixel (MP) camera and *float* notation $\sim 14\text{GB}$). Therefore Wetzstein et al. [68, 69] clustered the equation system into independent subsystems that are solved individually. In this algorithm another main problem is that the LTM has to be acquired for every scene separately, because the exact light transport varies with the scene.

3.4 Summary

In this chapter the theoretic and algorithmic background for our work was given. First, the SFF reconstruction method with the different focus measure operators was explained in detail. Because the SFF algorithm acquires images at defined depth steps, the resolution of the reconstruction result is limited. To achieve a resolution beyond one depth step various interpolation methods were presented.

Second, two methods to avoid illumination artefacts when capturing scenes with a higher dynamic range were presented. The first one is high dynamic range photography, where the static scene is captured with different exposure times. These images are fused to get one image where the whole dynamic range is illustrated. The second method uses an additional projector to compensate the illumination artefacts of a scene. Through a correspondence between the projected pattern and the acquired camera image the scene illumination can be adjusted locally to compensate high differences in the object reflectance.

Chapter 4

Microscopic Projector-Camera System

Contents

4.1	Hardware System	35
4.2	Optical Path	39
4.3	Calibration	41
4.4	Summary	53

In this chapter an overview of the projector-camera system for the acquisition of shape from focus images is given. Section 4.1 gives an introduction in the hardware specification of the acquisition system. The optical path through the projector camera system is described in Section 4.2. Calibration algorithms for radiometry and geometry are presented in Section 4.3.

4.1 Hardware System

A traditional SFF system uses a constant light source to illuminate the object. Since the SFF algorithm calculates focus information out of high frequencies in texture, this global illumination leads to a loss of information when it comes to great differences in the object reflectance within an image. This occurs, when the object reflectance reaches the maximum or the minimum of the displayable range of the sensor. Another problem occurs if the object surface itself is lacking in detectable texture (see Fig. 4.1). To overcome these problems we use a projector as an adaptive light source instead of constant illumination.

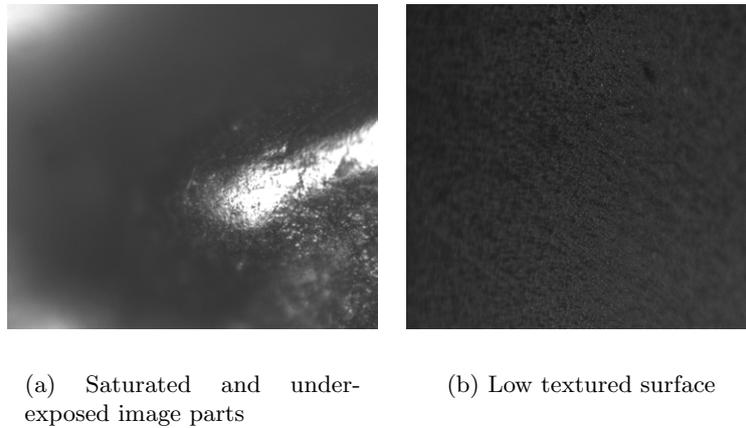


Figure 4.1: Lack of focus information due to saturated or under-exposed image parts (a) and low-textured surfaces (b).

On the one hand, in the case of under-exposed or saturated parts on the object, the projector can adapt the light intensity locally. On the other hand, the projector can be used to project a strong texture on the focused object surface to create a texture on homogeneous objects.

In the hardware setup, the camera and the projector are placed coaxial, as illustrated in Fig. 4.2. The reason for the coaxial illumination is that the focus plane of the camera

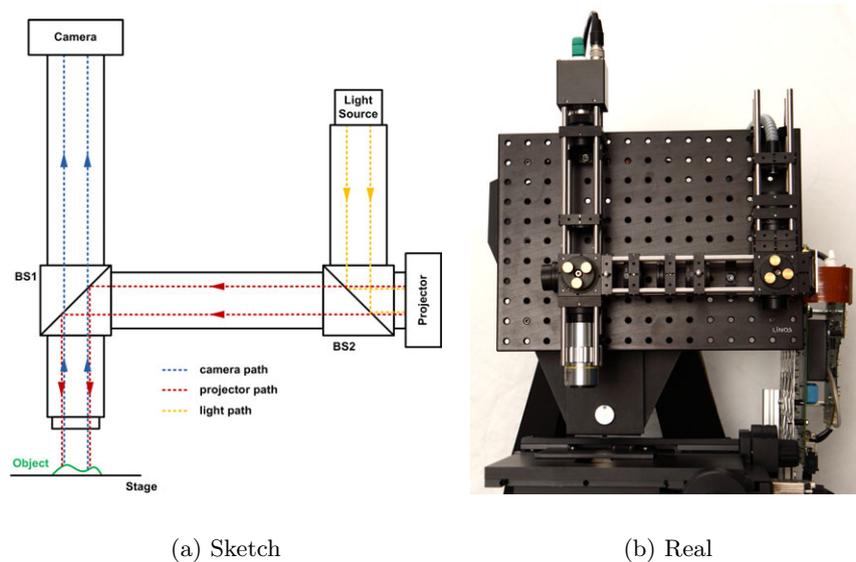


Figure 4.2: Coaxial camera projector system. The light rays from the projector to the object and from the object to the camera are parallel.

has to be exactly the same plane where the illumination pattern of the projector is in focus. To achieve a coaxial setup we use a beam-splitter (BS1 in Fig. 4.2) to redirect the light rays from the projector to the camera axis. A beam-splitter is an optical device that splits the incoming beam of light in two light rays with the half of the incoming intensity (see Fig. 4.3).

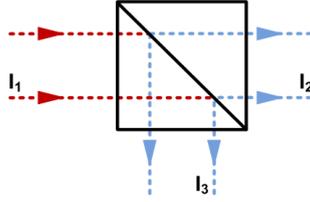


Figure 4.3: Schematic representation of a beam-splitter. I_1 incoming light 100%, I_2 - 50% transmitted light and I_3 - 50% redirected light.

The technology of our projector is a spatial light modulator (SLM) based on a liquid crystal on silicon (LCoS) display. The LCoS is a reflective micro-display which consists of a silicon backplane, where the liquid crystals are directly mounted. This whole plane is coated with a highly reflective surface, as illustrated in Figure 4.4. With this technology

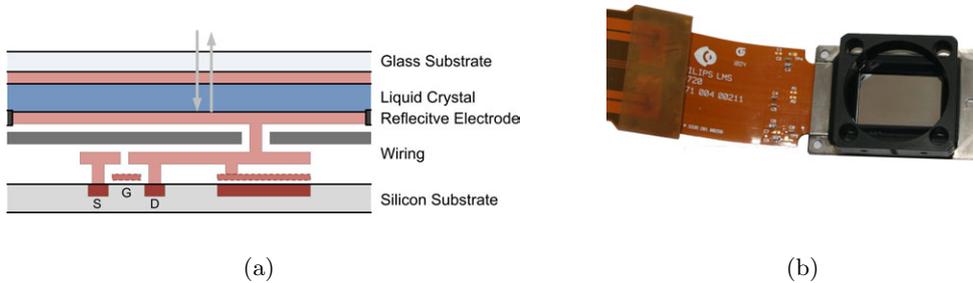


Figure 4.4: LCoS Display. A field effect transistor (FET) controls the reflection, whereas the liquid crystal acts as a SLM for illumination passing through. In (a) a schematic illustration of one LCoS pixel and in (b) the LCoS display used in this project is shown.

the phase of the reflected light can be modulated spatially [4, 24]. The LCoS display technology has some key advantages for microscopic solutions compared to other micro projector technologies. One main advantage compared to the liquid crystal display (LCD) technology is its high fill factor ($\sim 92\%$) with nearly no loss of brightness. Another advantage compared to direct light processing (DLP) projectors is the continuous light projection. DLP projectors generate pixel values by a sequence of tilting micro mirrors to

the light source and away from it.

The SLM system consists of a light source, a twisted nematic LCoS display and a polarizer P before and an analyzer A after the display, as illustrated in Figure 4.5. A light

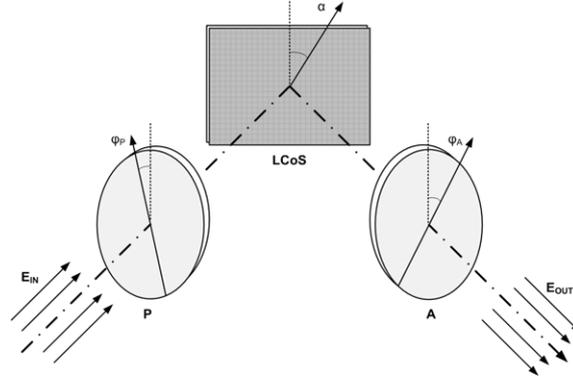


Figure 4.5: Spatial light modulation with a LCoS display. Light from a light source E_{in} is filtered to a single polarization by P , reflected with a shift of the light phase according to the pixel values. The analyzer A converts the light intensity according to this shift.

beam sourcing from a halogen cold light source is first polarized by P . This polarizer converts the unpolarized light beam into a beam with a single linear polarization angle φ_P .

This polarized light is then reflected by the LCoS display. It changes the polarisation of the incoming light by an angle α , proportional to the 8bit grey value of the projected pixel. The angle is therefore between 0 and α_{max} at grey values between 0 and 255. The analyzer A is another polarization filter, which filters the light beam corresponding to the difference between its angle φ_A and the polarization angle of the incoming light. The intensity of the transmitted light is proportional to the difference between the polarization of the light beam and the angle of A . This intensity can be calculated with the Malus' law by

$$E = E_{in} \cos^2(\theta), \quad (4.1)$$

where E_{in} is the incoming intensity and θ the angle between the incoming polarisation $\varphi_P + \alpha$ and the angle φ_A of the analyzer. Therefore the theoretical optimal angle between polarizer and analyzer is α_{max} , as presented in [10]. To calibrate the polarizer and the analyzer, uniform patterns with maximum 255 and minimum values 0 are projected. The angles of P and A are adjusted iteratively to get the radiance maximum on the object

at the projector's maximum intensity and the radiance minimum when projecting the minimum intensity.

4.2 Optical Path

The optical path and its lens system are designed to project a focused illumination pattern on the object which is captured on a focused camera image. The optical system can be roughly divided into two main parts, the camera path and the projector path. The projector path defines the optical system from the light source to the specimen whereas the camera path defines the optical system from the specimen to the camera. The focus plane of the projector and the camera have to be coplanar. A model of the optical system is shown in Figure 4.6.

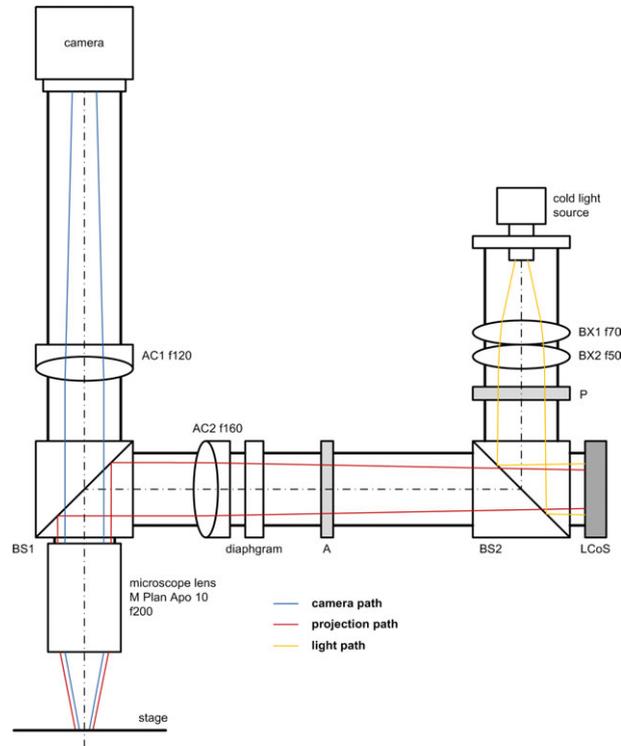


Figure 4.6: Optical path and lens system. The camera sensor is focused by AC1 and the microscope lens on the object. The light beam of the cold light source is focused through BX1 and BX2, reflected by the LCoS and focused by AC2 and the microscope lens. The focus of both camera and projector path is coplanar.

The projector path starts with the light rays of the cold light source which are far away from parallel. Therefore they are focused by two biconvex lenses BX1 and BX2 so

that the light rays falling on the LCoS display are approximately parallel. The polarizer P and the analyzer A are part of the SLM system as described in Section 4.1 and have no influence on the light ray focus. Afterwards, from the projector display, the light rays pass through an iris diaphragm. With this diaphragm the intensity of the projected light can be dimmed. After this regularization the light beams are refracted by the positive achromatic lens AC2 to become parallel. At normal biconvex lenses the refraction of light beams differ with their distance to the optical axis and their wavelengths (colors) (see Figure 4.7). An achromatic lens compensates this effect. This lens doublet consists

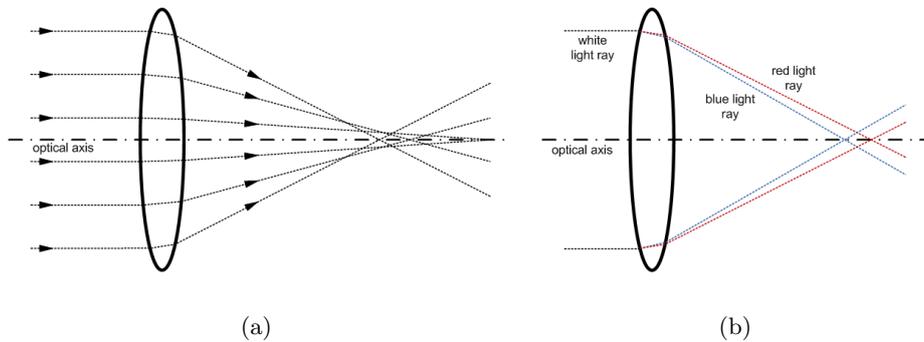


Figure 4.7: Simple biconvex lens with aberration, according to [59]. (a) Spherical aberration: rays closer to the optical axis have their focus farther from the lens. (b) Chromatic aberration: short wavelengths are more strongly refracted, the focus is closer to the lens.

of a positive lens with an additional negative element in order to balance out both the spherical and the chromatic aberration. The parallel light rays are reflected by the beam splitter BS1 and focused on the specimen by an infinity corrected objective (*Mitutoyo M Plan Apo 10*). The parallelism of the light rays before the objective is necessary to avoid artefacts as a result of converging light rays passing through the beam splitter. The *M Plan Apo 10* infinity corrected lens is designed to focus on objects with an increased distance from objective nose-piece to the specimen than normal microscopic objectives and performs an additional spherical and chromatic compensation. The working distance is 33.5mm at a magnification of $10\times$. This is necessary to have enough free space to move an object in front of the objective.

The light rays reflected by the specimen are refracted by the same infinity corrected objective in the inverse way. Therefore, the light rays are again parallel afterwards. These parallel light rays are transmitted by the beam splitter BS1 and refracted by the achromatic lens BX1 which focuses the light rays directly on the image sensor of the camera.

The parts of the object that are on the focus plane are sharp in the image. And, because the camera and the projector system have the same focal length, the projected pattern is also in focus on the parts of the object that are on the focused plane.

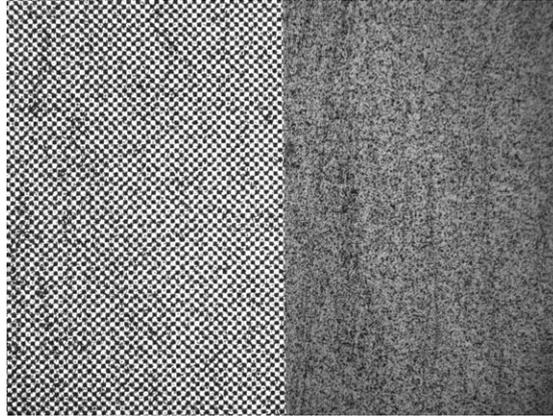


Figure 4.8: Camera image of a textured plane, perpendicular to the optical axis, which is in focus. On the left side a pattern is projected sharp on the surface, while the left side is homogeneous illuminated.

In Figure 4.2 a camera image of a focused textured plane, perpendicular to the optical axis, is illustrated. The right side of the surface is homogeneous illuminated while on the left side a checker pattern is projected. Because both sides are focused in the image one can infer that the focal planes of camera and projector are coplanar.

4.3 Calibration

This section explains the calibration of the projector-camera system. In order to capture the object by a camera and use the projector as light source, the influences of the projector on the camera images has to be known. The system has to be calibrated in two ways: First, the radiometric influence of focused projector intensities on focused camera intensities have to be mapped (see Section 4.3.1). This includes all the radiometric non-linearities during the projection and the acquisition process. Second, a geometric mapping between the camera and the projector coordinate frame in the focused case has to be defined, to get correspondences between projector and camera pixels (see Section 4.3.2). Additionally, the calibration of a scatter matrix is explained in Section 4.3.4. This matrix maps both, the radiometric and geometric correspondence between projector and camera system.

4.3.1 Radiometric Calibration

This section gives the theoretical overview of the radiometric calibration of a camera and a projector. Hence, we use a method similar to Debevec et al. [14] for camera response calibration and Koninckx et al. [30] and Nayar et al. [42] for projector response calibration.

Camera Response Function

The radiometric calibration of the camera results in a correspondence between the incoming light irradiance \mathbf{E}_c and the pixel values at the camera image \mathbf{Z}_c . As presented in Section 3.2, the camera is handled as a black box where the exposure time Δt and the discrete pixel values of the output images are known and the incoming irradiance values \mathbf{E}_c are unknown. The definition of the non-linear mapping of the camera response function (*crf*) is given by

$$\mathbf{Z}_c = f_c(\mathbf{E}_c \Delta t), \quad \text{where } Z_c : \mathbb{R}^+ \rightarrow [0, 255] \quad (4.2)$$

$$crf(\mathbf{Z}_c) = \ln(f_c^{-1}(\mathbf{Z}_c)) = \ln(\mathbf{E}_c) + \ln(\Delta t). \quad (4.3)$$

To calculate the camera response function *crf* we define a quadratic optimization problem from (4.3) to

$$\underset{crf}{\operatorname{argmin}} \left\{ \sum_{i=1}^N \sum_{j=1}^P \{w(Z_{i,j}) [crf(Z_{i,j}) - \ln E_i - \ln \Delta t_j]\}^2 + \sigma \sum_{z=Z_{min}+1}^{Z_{max}-1} [w(Z_{i,j}) crf''(z)]^2 \right\} \quad (4.4)$$

where i is the pixel- and j the image-index as specified in Section 3.2. The first term includes an equation system of N points in P images in a least squares sense. The second term in (4.4) ensures the smoothness of *crf*.

To get enough sample pixels at different exposure times, we acquire an image series of a static scene with constant illumination. This is necessary to calculate the mapping of the direct pixel differences in a scene where only the exposure times vary. The scene is an uniform plane normal to the optical axis. To model the curve over the whole working range, we captured the calibration images with an exposure time from $500\mu s$ to $100000\mu s$ in $500\mu s$ steps. The optimization result of the *crf* is illustrated in Figure 4.9. With this camera response function the discrete 8bit pixel brightness values with an assigned exposure time can be easily converted into radiance values with

$$\mathbf{E}_c = \frac{\exp(crf(\mathbf{Z}_c))}{\Delta t}. \quad (4.5)$$

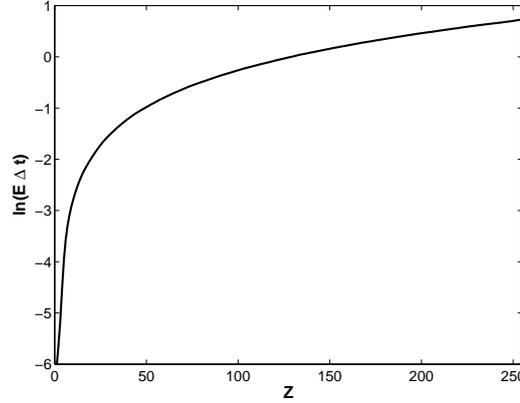


Figure 4.9: Camera Response Function. Mapping between pixel values of camera images and the camera exposure $X = \ln(E\Delta t)$.

Projector Response Function

The projector response function (*prf*) is a mapping which includes all non-linear effects during the projection process. This is necessary to calculate the correct projector irradiances from given input pixel values. This function incorporates all non-linearities which occur during the projection process. It gives a direct relation from the discrete pixel intensity to the irradiance on the object.

$$\mathbf{Z}_p = prf(\mathbf{E}_c), \quad \text{for } Z_p(x, y) : \mathbb{R}^+ \rightarrow [0, 255] \quad (4.6)$$

The calibration is similar to the determination of the *crf*. The projector can be interpreted as an inverse camera. Hence, the camera only detects the grey values of the irradiance intensities \mathbf{E}_c reflected by the object. We can calculate these intensities via the previous defined *crf*. The *prf* is estimated with a mapping of the camera irradiance values \mathbf{E}_c with the pixel values of the projected pattern \mathbf{Z}_p and the reflectance map of the surface \mathbf{R} . The projector radiance values are calculated with

$$\mathbf{E}_p = \mathbf{R}^{-1}(\mathbf{E}_c). \quad (4.7)$$

Normally, the reflection of objects has unknown local differences due to different texture, color or geometry. In our projector irradiance calculation, the linear reflection model \mathbf{R} is basically eliminated by projecting uniform patterns onto a planar, uniform and diffuse surface, as figured out in [31].

To calibrate this mapping, uniform patterns with different intensity levels are projected

onto the surface and acquired by the camera. The intensity levels i of the projector pattern $\mathbf{Z}_{p,i}$ range from 0 to 255. To avoid under-exposed or saturated images the exposure time has to be set according to this intensity. It ranges from Δt_{min} (exposure time for minimal intensity) to Δt_{max} (exposure time for maximal intensity) in logarithmic steps. This can be calculated with

$$a = \ln(\Delta t_{max}) \quad b = \ln(\Delta t_{min})$$

$$\Delta t_i = \exp\left(a + \frac{b-a}{255}i\right), \quad \text{where } i = 0, 1 \dots 255. \quad (4.8)$$

The camera radiance image $\mathbf{E}_{c,i}$ at the projected intensity i can be calculated according to (4.6). The value in the projector response function corresponding to the image intensity i is calculated as the mean of the radiance image normalized by the radiance image at the maximum intensity 255 with

$$prf(i) = \mu\left(\frac{\mathbf{E}_{c,i}}{\mathbf{E}_{c,255}}\right). \quad (4.9)$$

The calculated projector response function is illustrated in Figure 4.10.

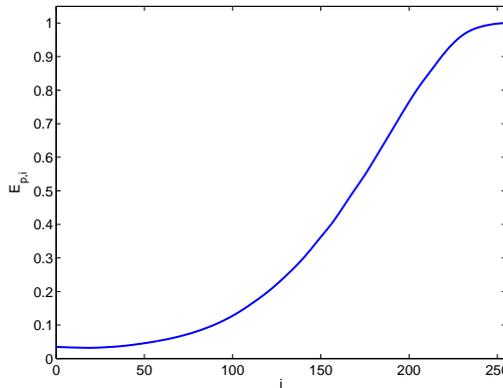


Figure 4.10: Projector Response Function. Mapping between pixel intensity levels i of projector pattern $\mathbf{Z}_{p,i}$ and the projected radiance $\mathbf{E}_{p,i}$

With the *crf* and the *prf* we have a radiometric connection from projector pixel intensities to their radiance values and from camera pixels to camera radiance values. This gives us the opportunity to derive the radiance map from an arbitrary object and to adapt the projector intensity to achieve a desired camera intensity.

4.3.2 Geometric Calibration

In this section the calibration of the geometric mapping between camera and projector coordinate systems is presented. This mapping is necessary to correctly perform a radiometric compensation between projector and camera. In contrast to Brown et al. [8] who calculated a geometric correspondence through a homography matrix with a homography estimation presented in [22], we store a fixed geometric pixel mapping in look-up tables (LUT). This direct method gives the opportunity of a very accurate mapping including all possible small variations of non-linearities in the optical system (e.g. lens distortion). Hence, the point correspondences between the projector and the camera coordinate system are stored in both ways. Each point coordinate in one domain is used as an index to obtain the corresponding point coordinate in the other.

To record the coordinates of the direct pixel mapping, we capture a camera image for each activated projector pixel separately. To get the point representation of the projector pixels in the camera image as sharp and accurate as possible, a mirror plane normal to the optical axis is chosen as the reflecting object. A sample projector-camera pixel correspondence is shown in Figure 4.11.

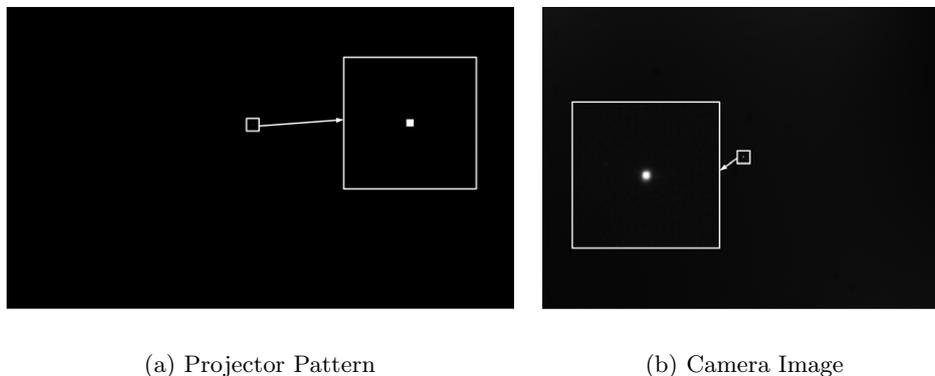


Figure 4.11: Geometric projector-camera pixel correspondence. One sample projector point is activated in (a). The corresponding camera image is shown in (b).

Because one projector point influences more than one camera point and is not ideally rectangular in the camera image, both the direction of the axis and the pixel scatter have to be incorporated in our mapping. The approximate position of the camera image in the projector pattern on the object is shown in Figure 4.12. Here we can see that the x-axis of the projector roughly corresponds to the inverse y-axis of the camera and the y-axis of the

projector roughly corresponds to the inverse x-axis of the camera. These differences occur due to the mechanical setup of the system. Also one reflected projector pixel is shown

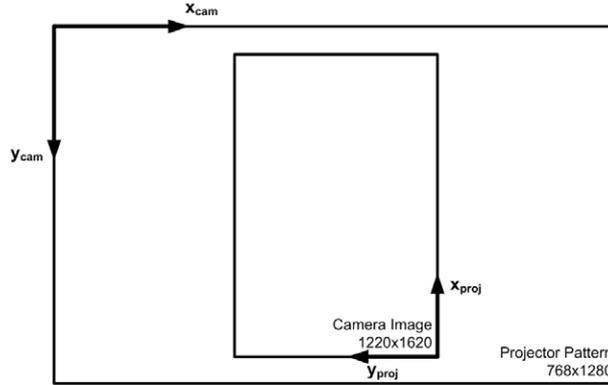


Figure 4.12: Geometric position of the camera image in the projector pattern.

on more than one camera pixel because of the different scale, where 1×1 pixel on the projector are $\sim 3.4 \times 3.4$ pixel on the camera image. This leads to the problem of finding the exact center of each projector pixel in the camera image. We extract a sub-image of the activated camera pixels (see Fig. 4.13). The coarse center of this patch in the camera

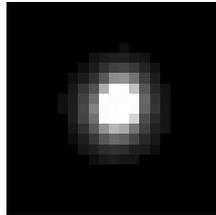


Figure 4.13: Patch of one projector pixel in the camera image

image is found by calculating the center of mass of a thresholded binary camera image. The width and height of the patch is also determined through a binary image with a lower threshold, to make sure that all influenced camera pixels are included. Therewith, the sub-pixel center is calculated through a weighted arithmetic mean of the pixel values in both x and y direction calculated with

$$\bar{x} = \frac{\sum_{i \in I(x,y)} \hat{x}_i I_i}{\sum_{i \in I(x,y)} I(i)} \quad \text{and} \quad \bar{y} = \frac{\sum_{i \in I(x,y)} \hat{y}_i I_i}{\sum_{i \in I(x,y)} I(i)}, \quad (4.10)$$

where I_i is the grey value at index i . The variables \hat{x} and \hat{y} are weights ascending in x respectively y direction. The resulting variables \bar{x} and \bar{y} are the new centroid coordinates

of a projector pixel in the camera image and are stored in the projector look-up tables (one for x and one for y direction). On the other side, the coordinates of the activated projector pixel in the projector pattern are stored in the camera look-up tables.

This procedure has to be done for every projector pixel that is shown in the camera image. To speed up this process, we project a grid where multiple projector pixels are activated at the same time. In this grid we have to pay attention on the distance between the activated pixels so that the pixels into the camera image would not influence each other. A sample projection grid is shown in Figure 4.14(a) with the corresponding camera image shown in Figure 4.14(b). With these geometric correspondences we can directly

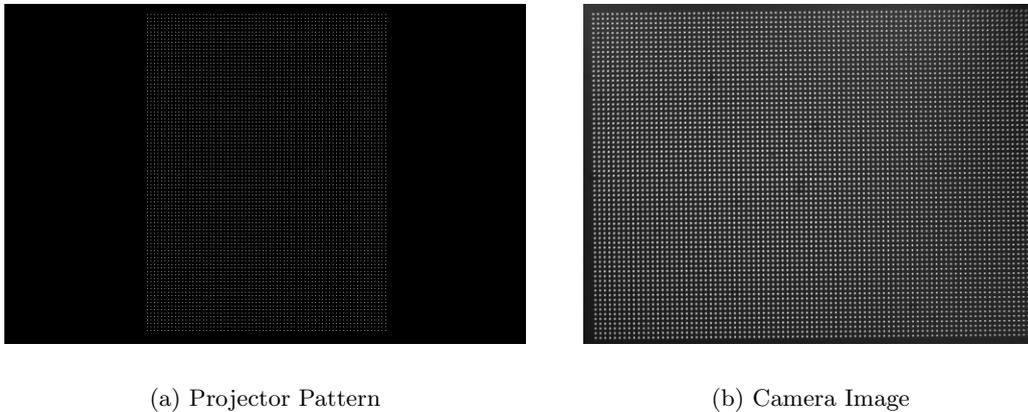


Figure 4.14: Projection of a grid for correspondence calculation. The calibration grid pattern (a) is projected on a mirror plane, normal to the optical axis. The distance between consecutive projector points is selected to avoid interactive influence in the camera image (b).

map each projector pixel in the camera coordinate system.

The results were compared to an ideal rectangular grid and to a grid calculated through a homography. The evaluation results in a slight skew of 0.1 degree of the projector grid to the camera grid. Furthermore, the comparison of the LUT to coordinates calculated through a homography results in an absolute mean deviation of 0.266 pixel with a standard deviation of 0.084 pixel. The error of the look-up table (LUT) to the calculated homography is illustrated in Figure 4.15. The pixel error is very low for the center pixel and increases towards the edges. With a restriction to map only a central region of corresponding pixels, the homography is also a possible model. The increasing distance error towards the edges is a consequence of the radial lens-distortion of the optical path in the microscope. This barrel distortion can also be modelled in a radial distortion model, in

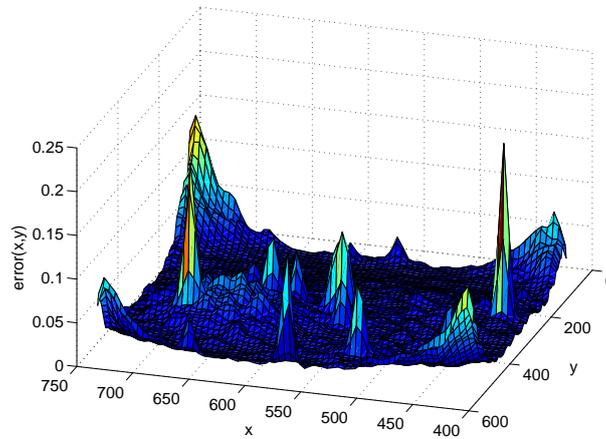


Figure 4.15: Error of Homography vs. LUT. Pixel error of coordinate correspondences calculated directly through LUT and through homography.

addition to the homography.

Although the mapping through a homography and a radial distortion is faster, we choose the calibration of a direct projector-camera mapping through a LUT, because we have no time restriction in the calibration process and this calibration is robust against all sorts of changes in the optical path. Furthermore, the calibrated mapping is more accurate over the whole acquisition area than a calculated distortion model.

4.3.3 Calibration of the Point Spread Function

In this section the response of the optical system to a point light source, namely the PSF, is determined. This response function is important in the characterization of the acquisition system. The theoretical background of the PSF is described in Section 3.1.1.

In our projector-camera system the determination of the real PSF is important to calculate the distribution of light energy of one projector pixel on the specimen. In literature the ideal PSF is described as a pillbox function, according to [40, 41]. Where the spatial size of the PSF depends on the radius of an ideal blur circle according to the optical system. The PSF of a focused point is an ideal point response. But due to optical aberrations and a finite size of the point light source the ideal PSF can not be realized. In SFF algorithms this pillbox function is approximated as a bivariate Gaussian distribution function, as presented in [40, 41, 45]. In Figure 4.16 a cross-section through theoretical PSF functions is illustrated. Its Fourier transform, the OTF, describes the PSF in the frequency domain. The analytical form of the OTF was further developed and calculated in

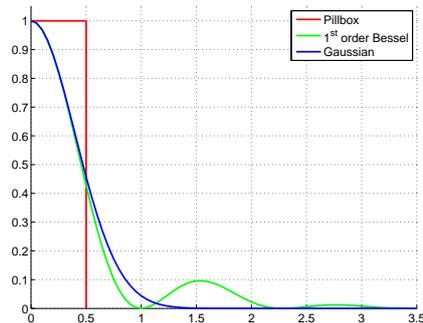


Figure 4.16: Cross-section through PSF approximations. The ideal PSF is represented as a pillbox function. Analytical calculations for the OTF with a circular aperture result in a first order Bessel function. For SFF algorithms this function was approximated with a Gaussian distribution.

[25] and [60] for an optical system with a circular aperture as a first order Bessel function.

In more complex optical systems the PSF and the OTF can not be calculated so easily, because necessary information such as the exact focal length, the exact aperture and the exact optical aberrations are unknown. In our hardware system another problem for the exact calculation of the PSF arises. The smallest possible projection spot in our system is one projector pixel. This illumination source is bigger than one camera pixel so that it could not be considered as a point light source per definition (according to [34]). Furthermore, this projector pixel is rectangular which also influences the appearance of the real PSF in the optical system. Hence, instead of calculating the PSF through an approximation the PSF is calibrated through the acquisition of one activated projector pixel through the blurring process. The acquired scene is a planar surface normal to the optical axis. To compensate the object texture the mean of the circular patch at different surface position is taken.

In Figure 4.17 and Figure 4.18(a) the acquired PSF at different depth steps is shown. In the PSF calibration it can be seen that for slight defocus the PSF can be approximated with a Gaussian distribution function, but at higher blur the PSF converges to the Fourier series approximation of a pillbox function. But in the frequency domain the OTF can be coarsely approximated with a first order Bessel function, as illustrated in Figure 4.18(b).

4.3.4 Construction of the Scatter Matrix

As presented in Section 3.3.4 and in Wetzstein et al. [69], both the radiometric and geometric mapping can be represented through a full light transport matrix. A modelled

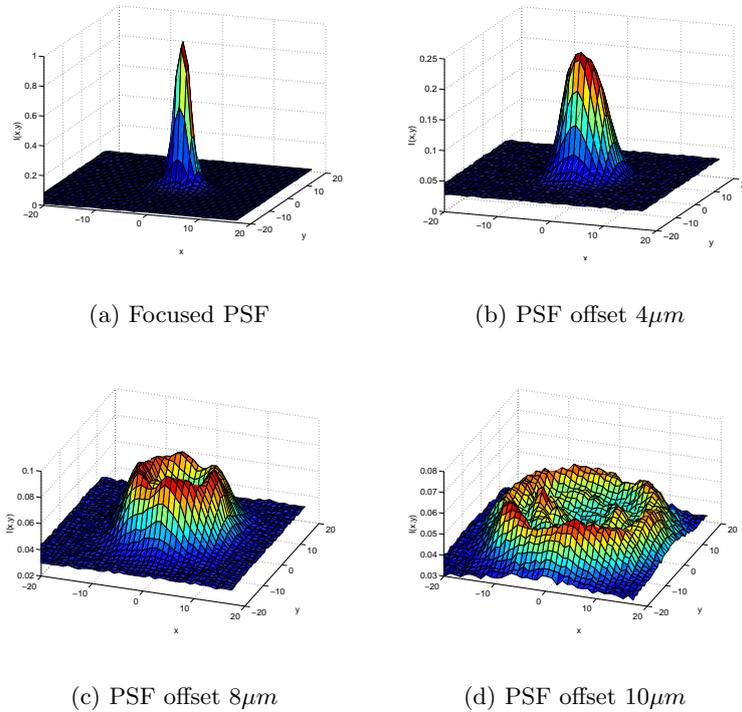


Figure 4.17: Acquired PSFs at different depth steps. The focused and slightly defocused PSFs can be roughly approximated with a bivariate Gaussian distribution (a,b) but at higher defocus the PSF converges to a Fourier series of the pillbox function (c,d).

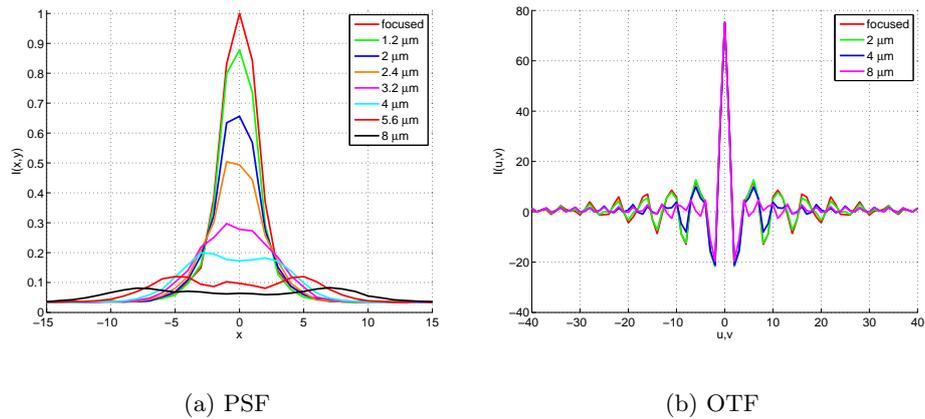


Figure 4.18: Cross-section through calibrated PSF and OTF at different blur.

camera image can be calculated through a simple multiplication of the LTM with the projected pattern with

$$\tilde{\mathbf{z}}_{\mathbf{c}} = T\tilde{\mathbf{z}}_{\mathbf{p}} + \mathbf{e}, \quad (4.11)$$

where $\tilde{\mathbf{z}}_{\mathbf{c}}$ is the vectorized camera image, $\tilde{\mathbf{z}}_{\mathbf{p}}$ is the vectorized projector image and \mathbf{e} is a camera image at the projector's black level. Through this matrix the full light transport between projector and camera can be calculated at once. The matrix also includes any kind of reflections on the object and during the projection process. A projector pattern for a desired camera image can be calculated with the inverse of this matrix. In Wetzstein et al. [69] this leads to a huge matrix with very low sparsity with an immense computation time of the matrix.

The scatter matrix, proposed in this work, is very similar to the full light transport matrix with the difference that we assume that each projector pixel has only an influence on a local patch around its exact position in the camera image. This is necessary because a single projector pixel, which is ideally rectangular, always affects several camera pixels and has no sharp border due to defocus and scattering, as described in Section 4.3.3. Neighbouring projector pixels influence each other in the camera image as illustrated in Figure 4.19. During the radiometric compensation, these influences have to be taken into account. The scatter matrix has the same size as the LTM but here only the point scatter for each projector pixel is saved. With this restriction the matrix has a high sparsity and can be calculated very fast compared to the method presented in [67].

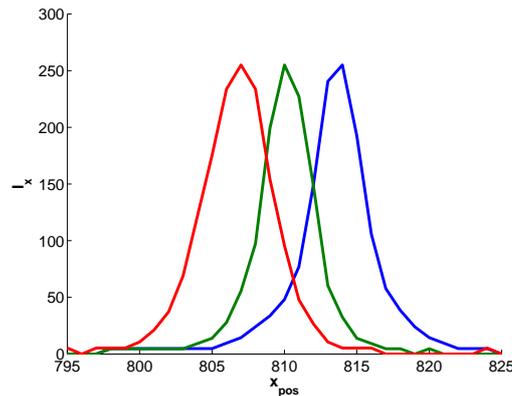


Figure 4.19: Profile of the camera image values of three neighbouring projector pixels

The computation of the scatter matrix is very similar to the geometric calibration and is calculated in three steps:

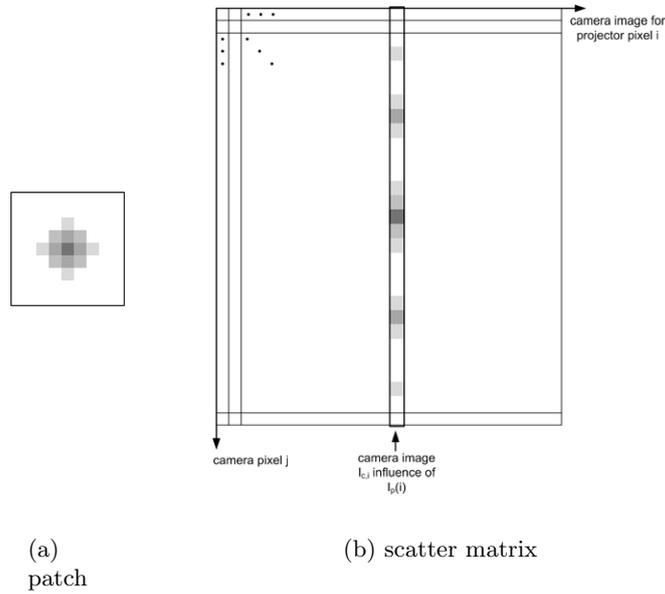


Figure 4.20: The scatter patch in x and y direction is put into the scatter matrix as vectorized camera image.

1. Each projector pixel is projected and captured individually. The specimen is a diffuse plane which lies normal to the optical axis and is in focus. In order to eliminate the object texture each projector pixel position is captured several times while moving the object and the mean image is taken. Again we can project a grid where multiple unaffected projector pixels are activated at the same time to speed up the computation.
2. The patches have to be cut out of the acquired image. Therefore we can use the spatial look up tables calculated during the geometric calibration to set the centroids of the patches. To separate the patch from the projector black level, the background has to be subtracted first. The remaining non-zero values of the patch are converted into radiance values with the camera response function (see (4.5)). A model of a scatter patch is illustrated in Figure 4.20(a).
3. This scatter patch \mathbf{P}_i , which represents each activated projector pixel at index i in the camera image $\mathbf{I}_{c,i}$, has to be put into the scatter matrix \mathbf{A} . The scatter matrix is constructed like the LTM. The ordinate represents the index of each activated projector pixel whereas the corresponding vectorized camera image is inserted along the abscissa, as shown in Fig. 4.20(b). The coordinates of the patch have to be

converted back into the camera image system to place it in position.

The scatter matrix \mathbf{A} has a high level of sparsity, since a single projector pixel only influences a small region of the observed scene. In our implementation the computation time of the scatter matrix \mathbf{A} is approximately one hour, whereas the density of the matrix is 0.00065%. After the construction of the scatter matrix, the camera radiance image $\hat{\mathbf{E}}_c$ can be modelled out of any radiance corrected projector pattern \mathbf{E}_p with

$$\hat{\mathbf{e}}_c = \mathbf{A}\mathbf{e}_p, \quad (4.12)$$

where $\hat{\mathbf{e}}_c$ and \mathbf{e}_p are the vectorized irradiance images form $\hat{\mathbf{E}}_c$ and \mathbf{E}_p .

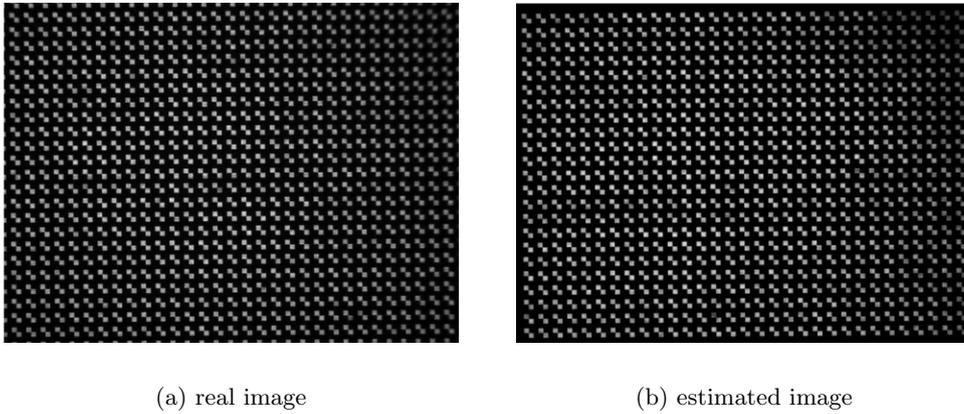


Figure 4.21: Comparison of the camera image of a projected sample pattern. (a) illustrates the real camera image and (b) the estimated camera image by multiplication of the scatter matrix with the projected pattern.

To illustrate the function of our scatter matrix, we project a sample code pattern on an uniform plane normal to the optical axis and compare it with the calculated camera image. In Figure 4.21 we can see that the real camera image versus the estimated camera image calculated with the scatter matrix. The calculated absolute average error is roughly 6% of the maximum image intensity.

4.4 Summary

In this chapter an overview of the projector-camera system to acquire a SFF image stack was presented. A coaxial LCoS display to project a pattern on the object was used. The optical path was designed to achieve a coplanar focus of the camera and the projector.

Therewith, the projected pattern is in focus on the parts of the object that are on the camera focus plane.

To use the projector as an illumination source the camera and the projector were calibrated radiometrical and geometrical. At the radiometric calibration the direct relationship between discrete projector pixel values and discrete projector pixel intensities was established. With this calibration a projector intensity can be calculated to achieve a desired camera intensity. At the geometric calibration a mapping between the camera and the projector coordinate system was established to get a geometric mapping between camera and projector pixels. These calibrations can be used to create a scatter matrix, which enables an estimation of the camera image through a multiplication with the projector pattern.

Chapter 5

Dynamic Illumination for Robust Microscopic 3D Metrology

Contents

5.1 Adaptive Illumination	56
5.2 High Frequency Texture Projection	67
5.3 Combination of Adaptive Illumination with High Frequency Pattern	75
5.4 Integral Shape From Focus	79
5.5 Summary	86

In this chapter our methods for increasing the accuracy and the robustness of shape from focus (SFF) reconstruction and a new way of focus measurement are presented. The instrument we use to achieve the accuracy improvements is a projector used as local adaptive light source. The projector focus plane is exactly coplanar with the focus plane of the camera. Therewith the projector patterns are sharp on the surface which is focused in the camera image.

To enhance the reconstruction results from traditional SFF we have to address the following problems. One major problem, especially in industrial inspection, is the information loss due to great differences between dark and bright pixels in the object reflection, that exceed the resolution of the camera sensor. This leads to under-exposed or saturated image regions. The focus is not measurable in such image regions. To overcome this problem we introduced a method where the illumination is adapted locally to equalize the reflection intensity over the whole object region which is captured by the camera. The ra-

radiometric differences are compensated to achieve a largely uniform camera image without under- or over-exposed parts. This method is presented in Section 5.1.

Another problem is the failure of a dense focus calculation of low textured objects. Many microscopic objects especially at the inspection of very small electronic components such as semiconductors or micro processors have a very smooth surface. Due to the loss of detectable texture it is also very hard to calculate an accurate reconstruction. To conquer this problem we use a method, first introduced by Noguchi and Nayar [48], where a high frequency pattern is projected on the object. Because the focus plane of camera and projector are equal, this pattern is imaged sharp on the focused image region and is used as object texture, which is detected by the focus measure operator. The methodology of this texture enhancement is presented in Section 5.2.

By merging these two methods both the dynamic range of the reflectance is reduced and the focus is calculated more accurately especially in low textured regions, as presented in Section 5.3.

Thirdly, we present a novel focus measure operator based on a normalized cross correlation (NCC) of the acquired image stack with a precalculated focused texture image of the scene. The all-in-focus texture image is calculated through an integral image of the image stack deconvolved with a previously calibrated integral point spread function. The result of the NCC is a correlation map, where the focus is expressed through the similarity of an image part with the same part of the focused image. The correlation map is further used to calculate a depth map like in traditional SFF, as presented in Section 5.4.

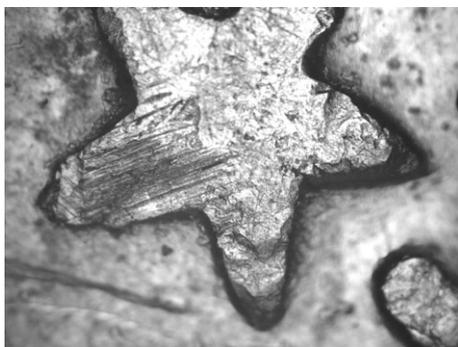
5.1 Adaptive Illumination

As mentioned above, one problem in traditional SFF methods is that high differences in the irradiance values which exceed the radiometric resolution of a camera sensor can not be displayed in one image. This overrun of the dynamic range of the sensor leads to under-exposed or saturated image regions. To overcome this problem, a method for the illumination adaption through a projector is presented in this section.

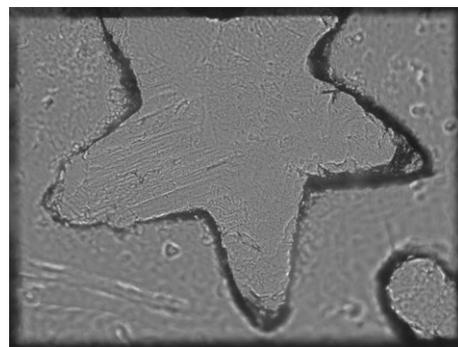
A previous method for increasing the dynamic range of sensors is high dynamic range (HDR) imaging (according to Debevec and Malik [14]). This method leads to more robust texture models for SFF reconstruction. But the main disadvantage especially for industrial inspection is its high acquisition time. To obtain one radiance image, several camera acquisitions with different exposure times are necessary to get an acceptable result. This increases the measurement time also by a minimum of two. Furthermore the potential

acquisition of extremely over-exposed images may cause blooming effects which degrades the quality of HDR images.

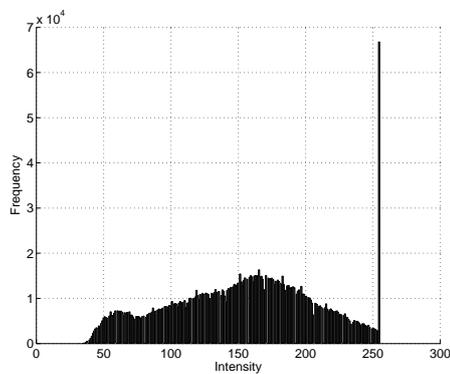
Because the projector is able to change the illumination locally, the adaptive illumination is used to compensate the reflection properties, whereas the measurement time does not increase. This leads to an equalisation of the reflected radiance which indirectly reduces the dynamic range of the reflectance for the camera sensor. Consequently this algorithm converts regions on the object which are outside the acquirable dynamic range back into regions with measurable focus by adapting its illumination.



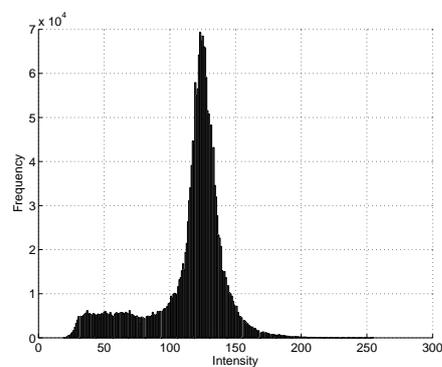
(a) Normal illumination



(b) Adapted illumination



(c)



(d)

Figure 5.1: Radiometric Compensation. A projector used as adaptive light source is used for adaptive illumination to compensate the reflection differences. Under-exposed or saturated regions in the image were compensated and are again in a measurable range for the camera sensor. The scene is a star of a two Euro coin, approximately $0.83 \times 1.12\text{mm}$.

There are many reasons why the dynamic range of the camera exceeds the representable

range:

- Different materials: When high reflective and diffuse parts of an object are constantly illuminated, the high reflective parts can get either over-saturated or the diffuse parts under-exposed in a camera image (According to [11]).
- Different surface texture: When the texture of the surface changes also the degree of reflection will change. This leads to the same problem as different materials. Dark regions are less reflective than bright ones.
- Material roughness changes: If the roughness within an object changes also the degree of reflection changes. A reflection model to calculate and represent this was presented by Oren and Nayar [49].
- Different angles of the object surface to the optical path: The reflected radiance which reaches the camera sensor differs with the angle between the optical axis and the surface normal. These differences in reflection can be calculated by the Lambertian cosine law.

With the knowledge of the occurrences of the illumination artefacts we built the theoretical image formation process. This process models the pipeline from projector pattern to camera image in Section 5.1.1. With the knowledge of the physical image formation, we calculate the radiometric model to adapt the projector pattern in Section 5.1.2. To reduce the computation time of the compensation we present a method to predict a possible initial projector pattern in each step, as described in Section 5.1.3. This will lead to a reduction to one camera iteration per step of the stage, which is equal to traditional SFF. Figure 5.1 shows an example of light adaption for radiometric compensation the object reflection to a desired value.

5.1.1 Image Formation Process

In this section the image formation process within the SFF image acquisition is described. This is necessary to understand the adaption process of the projector pattern. Figure 5.2 models this image formation from the input projector pattern to the discrete camera image.

The discrete projector pattern \mathbf{Z}_p is mapped to an irradiance pattern \mathbf{E}_p by the projector response function prf , as determined in Section 4.3.1, which models the non-linearities

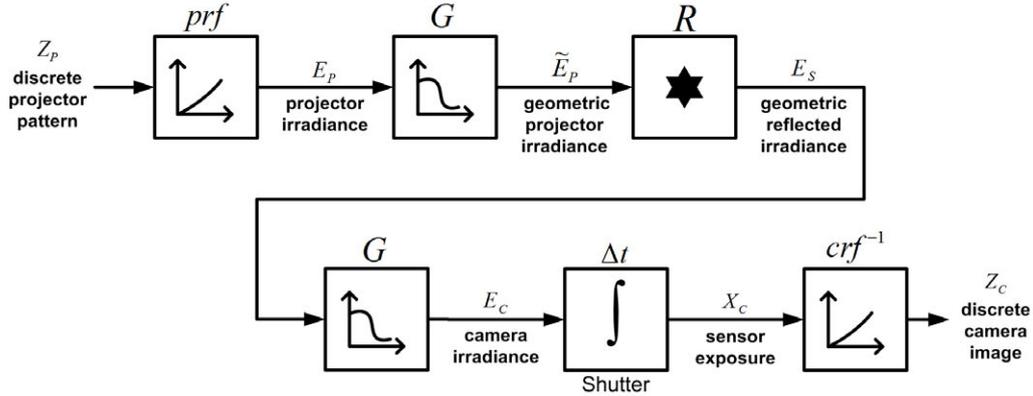


Figure 5.2: Image formation pipeline: The discrete projector image \mathbf{Z}_p is projected via irradiance pattern \mathbf{E}_p . The projector irradiance undergoes a geometry dependent defocusing on the object to $\tilde{\mathbf{E}}_p$ and is reflected on the scene surface. The reflected irradiance \mathbf{E}_s undergoes the defocus again and is finally mapped to a discrete camera image \mathbf{Z}_c .

between the pattern gray value and the projected irradiances.

$$\mathbf{E}_p = \text{prf}(\mathbf{Z}_p). \quad (5.1)$$

Due to a non planar surface geometry not every pixel of the projected pattern is in focus. This defocus is expressed as a blurring with a Gaussian filter, whereas the amount of defocus is relative to the offset from the projector focus plane. Furthermore, the amount of defocus depends on the surface depth at each pixel. This surface depth \mathbf{G} is previously unknown and leads to the defocused pattern $\tilde{\mathbf{E}}_p$. The object reflection \mathbf{R} models the factor of the reflection of incoming light at each pixel. It can be assumed that it is linear to incoming irradiance on materials following the Lambertian or Phong reflectance model as declared in [30] and [42]. Consequently, the reflected irradiance \mathbf{E}_s is calculated with

$$\mathbf{E}_s = \mathbf{R}\tilde{\mathbf{E}}_p. \quad (5.2)$$

As the focus planes of projector and camera are coplanar, the reflected irradiance \mathbf{E}_s undergoes the same amount of defocus according to \mathbf{G} , which leads to the defocused camera irradiance \mathbf{E}_c . The camera acquires the image through the exposure \mathbf{X}_c which is defined as the integral of the incoming irradiance over an exposure time Δt . We assume

a constant irradiance over the exposure time to simplify the calculation to

$$\mathbf{X}_c = \int_0^{\Delta t} \mathbf{E}_c(t) dt \approx \mathbf{X}_c = \mathbf{E}_c \Delta t. \quad (5.3)$$

The non-linearities in the camera acquisition process are taken into account by the camera response function *crf*. Because the *crf* is monotonic increasing (inverse is well defined) the discrete camera image is calculated with

$$\mathbf{Z}_c = \text{crf}^{-1}(\ln(\mathbf{X}_c)), \quad (5.4)$$

according to the definition of the camera response function in Section 4.3.1.

Based on the physical model of the projection-acquisition process, a projector image is estimated in the following section that compensates the different reflection factors of the reflectance map \mathbf{R} in a closed loop system.

5.1.2 Compensation of the Reflection

From the image formation process we know the influence of the projector illumination on the camera image. By adjusting the projector illumination we are able to radiometrically compensate the unknown object reflectance \mathbf{R} . The compensation should lead to an object illumination that results in a predefined uniform reflectance which avoids exceeding the dynamic range of the camera. The compensating illumination is optimized to obtain a desired camera image $\hat{\mathbf{Z}}_c$. To eliminate the over- and under-exposed parts in the same degree, $\hat{\mathbf{Z}}_c$ is typically an uniform image of grey value 128. The input for the estimation algorithm are the projector pattern \mathbf{Z}_p and the camera image \mathbf{Z}_c resulting from the image formation as declared above. The camera response function (*crf*) and the projector response function (*prf*) are estimated in a previous calibration step (see Section 4.3.1) without the exact knowledge of the surface properties. Because of the unknown object geometry this optimization leads to an ill-posed problem. Therefore we have to estimate the reflectance map iteratively from an initially acquired image $\mathbf{Z}_{c,0}$ and the initial illumination pattern $\mathbf{Z}_{p,0}$. The initial pattern is also an uniform image with a grey value of 128. This initial value leaves enough range to correct the illumination in both positive and negative direction.

Before we start the compensation we make some simplification for the ill-posed problem:

- It is neither possible nor necessary to compensate the illumination of the entire image in the microscopic setup. Due to the co-planarity of the focus plane of camera and projector, only those parts of the image that are roughly in focus have to be compensated radiometrically.
- The illumination must not eliminate high frequency parts in the image which are necessary for focus calculation. Therefore the projector pattern has to be of considerably lower spatial resolution compared to the camera resolution.

An illustrative model of the iterative radiometric compensation is shown in Figure 5.3. Because the connection of camera and projector is the object, two crucial parts in our system are unknown, the scene geometry \mathbf{G} and its reflectance \mathbf{R} . For the adaptive illumination we have to compensate the scene reflectance \mathbf{R} out of projector pattern \mathbf{Z}_c and camera image \mathbf{Z}_p . For simplification we will consider a single pixel first. According to

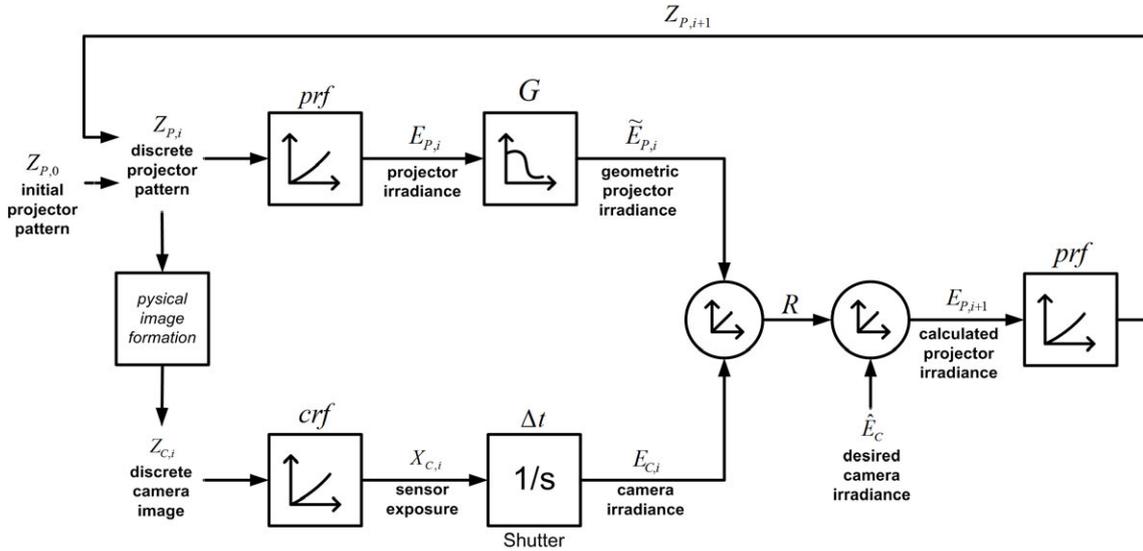


Figure 5.3: Iterative pattern adaption. Estimation of the reflectance R through calculation of the camera and the projector irradiance. Through calculated R and desired camera irradiance \hat{E}_c a compensating projector pattern $E_{p,i+1}$ is estimated. The iteration leads to a stepwise refinement of the pattern.

(5.2), the reflectance map is calculated with the camera irradiance $E_{c,i}$ and the projector

irradiance $E_{p,i}$ with

$$R_i = \frac{E_{c,i}}{E_{p,i}}. \quad (5.5)$$

This connection between the projector and the camera intensity is illustrated in Figure 5.4. With this reflectance estimation the compensated projector irradiance is estimated through

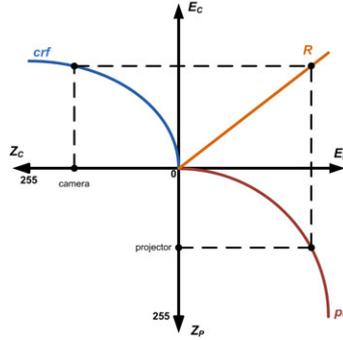


Figure 5.4: Projector-camera model. The projector intensity Z_p is transformed by prf and reflected by the scene reflectance R . The reflected irradiance is transformed to the pixel intensity via the crf .

$$E_{p,i+1} = \frac{\hat{E}_c}{R_i}, \quad (5.6)$$

where the irradiance value E is mapped from the according grey value Z through the response functions crf and prf . Because of possible non-linearities in the reflectance function, or bad initialisation of $I_{p,0}$, the compensated camera intensity $I_{c,1}$ may differ from the desired camera intensity \hat{I}_c . To get to an optimized solution the iteration process is repeated until the compensation is accurate enough or the reflectance correction reaches its optimum, whereas $\|E_c - \hat{E}_c\|$ or $\|R_{i-1} - R_i\|$ falls below a predefined threshold.

To transform this theory in the 2D case we have to consider the different resolution and mapping of camera and projector. As described in Section 4.3.2, a single projector pixel, which is ideally rectangular, directly affects several camera pixels. Furthermore, the affected area is not a sharp region due to slight defocus and scatter. We model this system response of a single projector pixel on the object surface through a point spread function:

$$E_s(x_c, y_c) = psf(x_c, y_c, x_p, y_p, E_p(x_p, y_p)). \quad (5.7)$$

This function models the response to one projector pixel (x_p, y_p) with the irradiance E_p

in the camera pixel (x_c, y_c) . We determined the scatter matrix \mathbf{A} where the point spread of all projector pixels in the camera image is mapped, according to Section 4.3.4. This matrix is calculated in a previous calibration step. Therewith the influence of all projector pixels on the surface of the specimen is calculated through the linear relationship

$$\mathbf{e}_s = \mathbf{A}\mathbf{e}_p, \quad (5.8)$$

where \mathbf{e}_p is the vectorized projector irradiance image

$$\mathbf{e}_p = [E_p(0,0), \dots, E_p(x_{p,m}, y_{p,n})]_{mn \times 1}^T \quad (5.9)$$

and \mathbf{e}_s the vectorized scene irradiances in camera space

$$\mathbf{e}_s = [E_s(0,0) \dots, E_s(x_{c,u}, y_{c,v})]_{uv \times 1}^T. \quad (5.10)$$

Hence, the size of the projector pattern is $m \times n$ and the size of the camera image is $u \times v$.

The inverse problem of mapping scene irradiance to the projector irradiance is more complex, since the inverse of the sparse scatter matrix \mathbf{A} is required for this calculation. Since \mathbf{A} is not invertible the compensation projector pattern is calculated with the pseudo inverse by

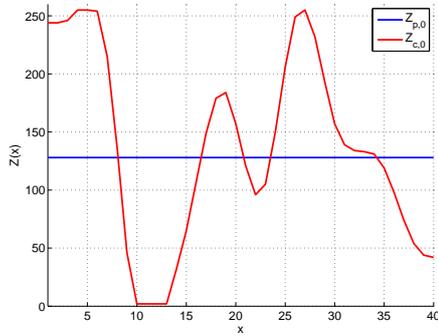
$$\mathbf{e}_p = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{e}_s. \quad (5.11)$$

Due to the huge size of the matrix \mathbf{A} ($mn \times uv$) this inversion can lead to memory problems. To overcome this problem the inverse calculation can be theoretically solved with a linear least-squares solution, as presented in [31]. Where the calculation of $\mathbf{A}^T \mathbf{A}$ and $\mathbf{A}^T \mathbf{e}_s$ is made without any problems. An approximation of the inverse matrix is calculated iteratively through a conjugate gradient method, as presented in [58].

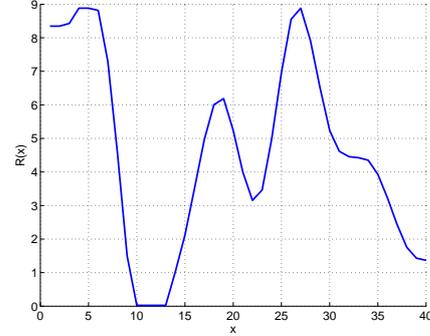
The calculation of the optimized projector pattern is only correct for radiometric compensation of perfectly focused objects. In practice, this assumption is not correct for non-planar surface geometry. In defocused surface regions the illumination pattern is also defocused. One projector pixel illuminates a bigger patch of the surface depending on the scene depth, which leads to inaccurate compensation.

For extremely defocused image regions, a bad illumination is of lesser concern, but on slightly defocused surface regions high-frequency artefacts could be introduced through compensation, which influence the focus measurement in SFF calculation. To avoid these

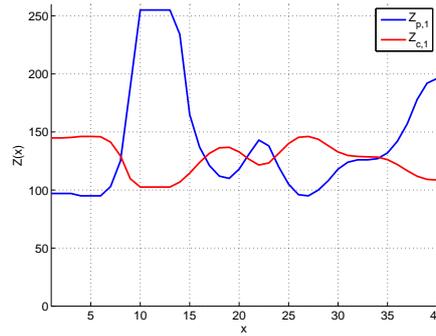
high frequency artefacts, the projector irradiance pattern is low-pass filtered before estimating the defocused reflectance maps.



(a) Initial projector pattern and camera image



(b) Calculated reflectance map



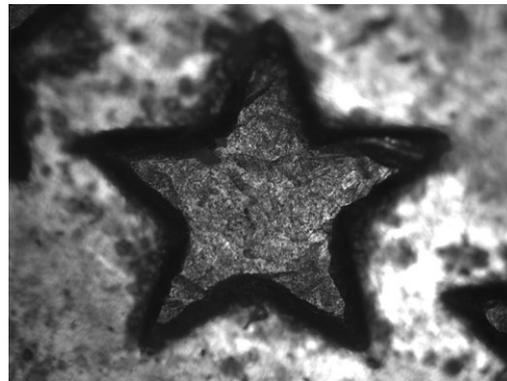
(c) Resulting projector pattern and camera image

Figure 5.5: One iteration of the radiometric adaption process. The reflectance map R is calculated out of initial projector pattern $Z_{p,0}$ and initial camera image $Z_{c,0}$. The new projector pattern $Z_{p,1}$ is estimated through the desired camera image (uniform 128 grey value) and the calculated radiance map R , resulting a new camera image $Z_{c,1}$.

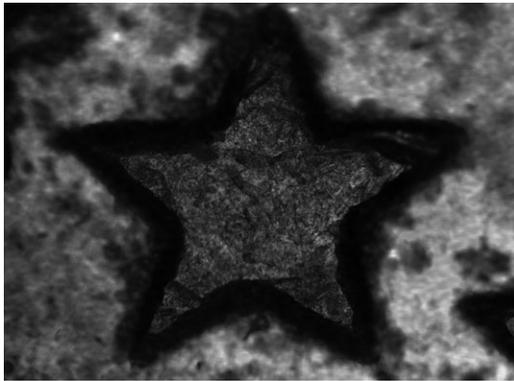
In Figure 5.5 an example adaption process through one iteration is illustrated on a one dimensional example. The reflectance map is estimated with the initial projector pattern and the resulting camera image. With this reflectance map and the desired camera image a projector pattern is calculated for radiometric compensation, resulting a compensated camera image. The radiometric compensation for the two-dimensional case is illustrated in Figure 5.6 in a real-world example.



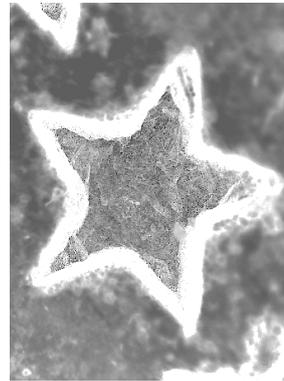
(a) Initial projector illumination



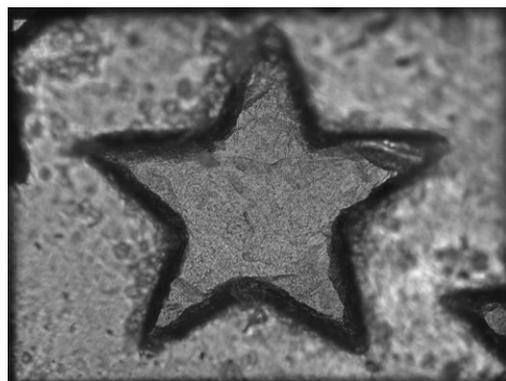
(b) Initial camera image



(c) Calculated reflectance map



(d) New projector illumination



(e) Improved camera image

Figure 5.6: Radiometric compensation in one iteration. Calculate reflectance map R out of initial projector irradiance $E_{p,i}$ and initial camera irradiance $E_{c,i}$. The new projector pattern $E_{p,i+1}$ is calculated through the desired camera image and the reflectance map R .

5.1.3 Predicted Adaptive Illumination

The SFF algorithm takes the focus measures of at least three consecutive images out of the stack to interpolate the sub-image focus maximum. To ensure that this calculation leads to an optimal solution, every object point has to be slightly in focus on at least three images. The step-size of the translational stage has to be small enough to ensure this constraint. From this constraint we can infer, that direct neighbours in the image stack just slightly differ from each other. Therefore, also the adapted illumination of the actual step just slightly differs from the illumination pattern of the previous step.

In the iterative adaption at least two iterations are necessary if the initial pattern is uniform. Consequently the acquisition of three frames is necessary in every step and the measurement time will also increase by a factor of three, which is prohibitive for most practical applications. This would not be a real improvement compared to high dynamic range imaging.

But we can take advantage of the similarities of the sequenced frames, because the scene appearance only changes a little bit between neighbouring images. First we take the illumination pattern from a previous step as initial illumination in the optimization of the actual step. This typically leads to an optimized solution in one iteration. Furthermore, we track the compensation mask through the image stack in a prediction-correction approach. There, the step prediction is computed through

$$\mathbf{E}_{p,i+1} = \mathbf{E}_{p,i} + \lambda(\mathbf{E}_{p,i} - \mathbf{E}_{p,i-1}), \quad (5.12)$$

where the variable λ is a scaling factor for the prediction. Therewith the illumination pattern for the next step is predicted with the information of the previous and the last step. The correction is calculated through the iterative compensation in one iteration. With this prediction-correction approach only one camera acquisition is necessary in each step. Additionally two initialisation frames are needed, which are negligible on a typical shape from focus acquisition where about sixty frames are needed.

Figure 5.7 shows an example reconstruction with standard shape from focus with homogeneous illumination compared to SFF with adapted illumination through the image stack.

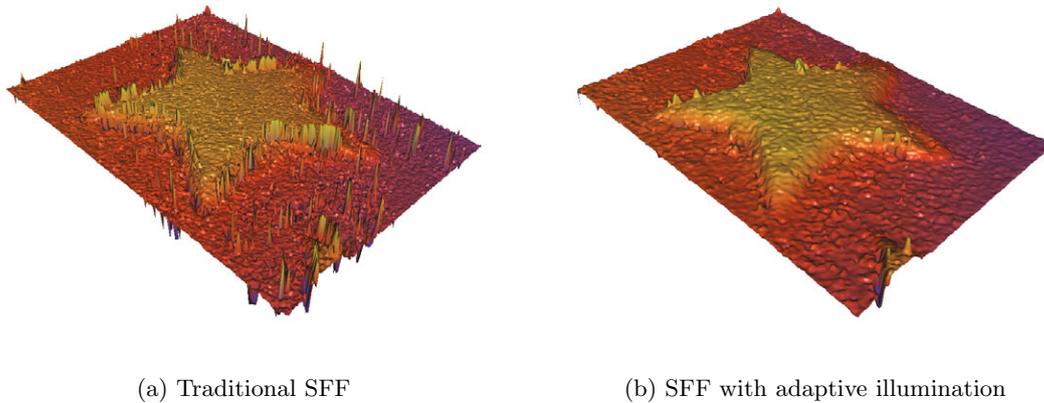


Figure 5.7: Depth reconstruction of the presented sample in Figure 5.6. (a) SFF depth with homogeneous illumination, (b) the same object reconstructed with adapted illumination.

5.2 High Frequency Texture Projection

An essential problem of traditional shape from focus is the requirement of enough measurable texture all over the acquired surface. On many objects, the surface is very smooth or shaded which leads to a lack of detectable texture. Examples of such surfaces are solder joints or silicon wafers, which are especially of practical importance.

This leads to the idea of projecting a strong texture on the object to create texture through projection. Because the focus of both camera and projector are coplanar in our hardware setup the projector pattern is sharp on the regions of the object, that are in focus on the camera image. The high frequencies of the pattern are detectable when the specimen is in focus and get equally defocused with the specimen. In Figure 5.8 a sample texture pattern is projected on a smooth surface without own detectable surface texture. Therewith the degree of focus is again measurable.

Noguchi and Nayar [48] presented a similar approach for the reconstruction of shape from focus using active illumination. There, a static patterned filter mask was placed directly after the light source. The problem of this method is the loss of the remaining object texture, because the filter mask only sets surface regions to fully illuminated or absolutely not illuminated.

The main challenge in texture enhancement is to find an optimal pattern, which creates perfect measurable texture on untextured regions. Ideal is a maximisation of the sensitivity of the focus measure to depth variations, maximizing the robustness and reducing the size

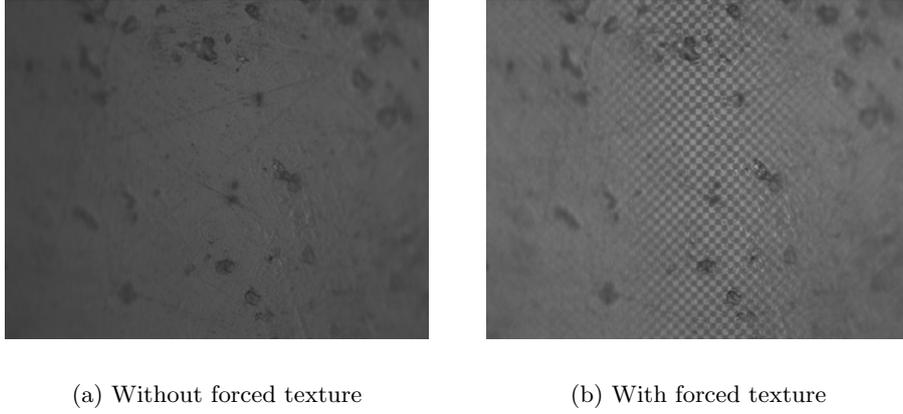


Figure 5.8: Sample texture enforcement on a glass surface. Image (a) has no detectable texture information, where in (b) the same scene was illuminated by a rough texture pattern. The focus and defocus is again detectable.

of the focus operator to achieve a maximum spatial resolution. The estimation process of finding this pattern is presented in Section 5.2.1. The application of the texture pattern to support the natural object texture to increase the spatial resolution or the robustness is presented in Section 5.2.2.

5.2.1 Calibration of the Optimal Pattern

In this section the calibration of the best pattern for the reconstruction of surfaces that lack in detectable texture is presented. We assume that the projected pattern is stronger than the natural object texture of the surface. Our evaluation emanates from traditional focus measure operators that detect high frequencies by gradient maximization through a filter kernel and sums up the detections over a local patch (e.g. sum of modified Laplacian, Tenengrad, ...). In [48] an optimal illumination filter pattern is found to be a checkerboard whose pitch is the same size as the distance between adjacent elements in the discrete Laplacian kernel. This results a checkerboard frequency t_x, t_y of 2 or 4 times the size of the Laplacian focus measure operator p_x, p_x in x and y direction:

$$t_x = \lambda p_x \quad (5.13)$$

$$t_y = \lambda p_y, \quad \text{where} \quad \lambda = \{2, 4\}. \quad (5.14)$$

Because in our hardware setup we have a larger projector-to-camera pixel ratio, we

follow a more intuitive way to get the optimal size and appearance of the projection pattern. In the following, we assume an absolutely smooth surface and therefore ignore the object texture. To get a high frequency pattern it will be only composed of pixel intensity extrema of 0 and 255.

With this assumption and according to [48] we want to optimize the pattern in three ways in order to get the best possible response from the focus measure operator:

- Maximization of sensitivity: Get high focus measure variation on small focus variations.
- Maximize robustness: The same degree of focus leads to the same focus measure response.
- Maximize spatial resolution: Reduce smoothing through minimizing the size of the focus measure operator.

Because we want the texture to be as dense as possible on the specimen we first have to take the scatter influences of neighbouring focused projector pixels into account. Projected texture can only be measured through illumination differences in neighbouring camera image pixels. The influence of one projector pixel in the camera image can be roughly modelled as a thin-plate spline. An influence radius σ_p is set, where the illumination

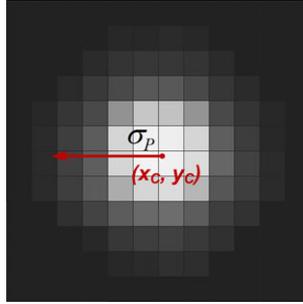


Figure 5.9: Influence radius σ_p of one projector pixel in the camera image with sub-pixel center of (x_p, y_p) in the camera image.

intensity has lost $\sim 95\%$ compared to the center value, as shown in Figure 5.9. It is depending on the projector-to-camera pixel ratio, which defines how many camera pixels are directly illuminated with one projector pixel, and the surface characteristics, which defines the scatter and the reflections around directly illuminated camera pixels. This influence deviation defines the minimal distance between two illumination peaks where a difference in the sharp texture can be detected, to $2\sigma_p$.

To get a dense depth reconstruction we also have to determine the maximum distance between two illumination peaks. This is necessary to recognize the texture on every pixel to get a more robust focus map over the whole camera image. Therefore we have to take a closer look on the focus measure operator. The focus measure operator in traditional SFF detects high frequencies through an approximation of the first or second derivative, usually in a 3×3 discrete sliding window through each image pixel. These focus measures are summed up in a local window $m_x \times m_y$ (see Section 3.1.3), whereas the spatial resolution decreases with the size of this local summation window. Figure 5.16 illustrates the influencing area of the focus measure operator for one image pixel. For a

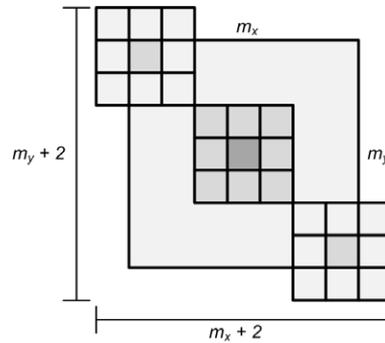


Figure 5.10: Focus measurement area of a focus measure operator. The traditional focus measure operator calculates the gradient in a local 3×3 patch which is summed up in a local $m_x \times m_y$ window. To get a dense focus measurement the distance between neighbouring texture elements has to be less than $m_{x,y} + 2$.

5×5 summation window the frequency variations are effectively measured in a 7×7 pixel window.

Subsequently the periods t_x and t_y of the illumination pattern in x and y direction are defined as

$$2\sigma_p < \frac{1}{2}t_x = \frac{1}{2}t_y < m_{x,y} + 2. \quad (5.15)$$

The periods t_x and t_y are calculated in the camera coordinate system. Thus, the resolution of the projector and the projector-to-camera pixel resolution difference have also to be incorporated. We assume that camera and projector pixels are quadratic ($t_x = t_y$) and we choose a quadratic focus measurement window ($m_x = m_y$).

When we take a closer look at the most common focus measures we see that the gradient is calculated in both principal directions. To create an illumination pattern

with the highest spatial resolution in a given periodicity while considering the scatter of projected pixel we evaluated a periodic checker board as optimal pattern. The illumination period of this pattern is $t_x/2$ and $t_y/2$ according to (5.15).

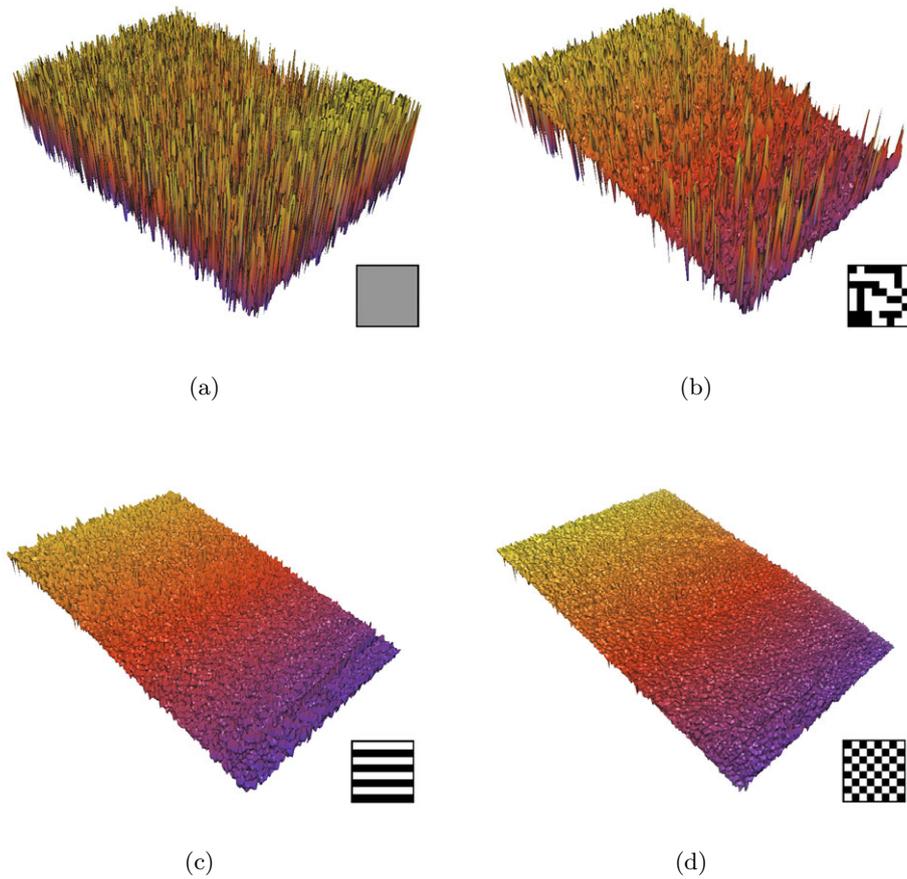


Figure 5.11: Reconstruction results of a planar surface, illuminated with different patterns. A textureless plane is illuminated with different projected texture pattern. Standard homogeneous illumination (a), random pattern (b), stripe pattern (c) and checkerboard pattern (d).

In Figure 5.11 a transparent plane is reconstructed with different illumination patterns. The surface of this plane is texture-less, which produces many outliers in traditional SFF reconstruction. The checkerboard pattern illumination produces more accurate shape information that is superior to that produced by a random pattern or a stripe pattern.

The different reconstruction results with different pattern frequencies at a fixed focus measure operator size are illustrated in Figure 5.12.

In a given system with a projector-to-camera pixel ratio of ~ 3.4 the influence diameter

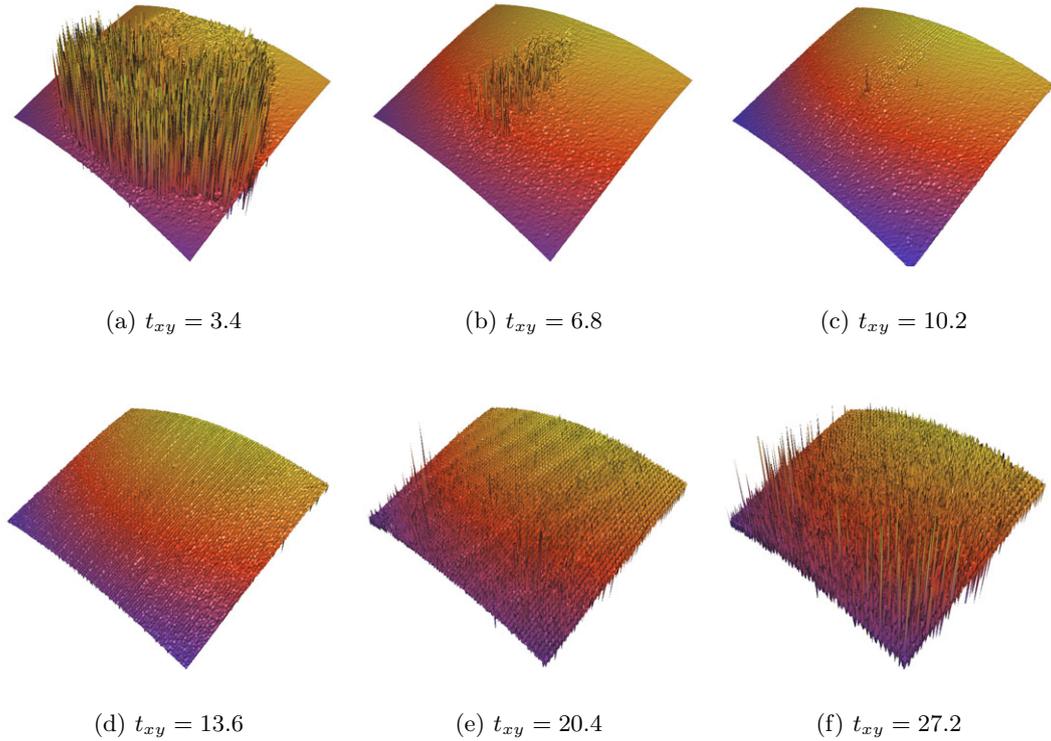


Figure 5.12: Reconstruction of a glass lens with illumination patterns of different frequencies. One projector pixel directly influences $\sim 3.4 \times 3.4$ camera pixels, $2\sigma_p \approx 8.4$ at a Tenengrad window size of 9×9 .

$2\sigma_p$ for a focused projector pixel is ~ 8.4 . The possible pattern frequencies in the camera image are also determined to be a multiple of projector-to-camera pixel ratio to $t_{x,y} = \{3.4, 6.8, 10.2, 13.6, \dots, n3.4\}$. Therefore, we chose the lowest possible focus measure size of $m_{x,y} = 9$ to fulfil the constraint form (5.15). If the illumination period of the pattern is beneath σ_p the illuminated texture can not be measured correctly and if it is above $m_{x,y} + 2$ the distance between two pattern peaks is too big to reconstruct a smooth surface. In Figure 5.12 the reconstruction with different illumination pattern frequencies is illustrated.

5.2.2 Application of a HF Pattern for Illumination

In this section the application of the optimal illumination pattern to create additional texture on an object is presented. On the one side, this texture enables a reconstruction of low textured objects and on the other side, the spatial resolution and the robustness

can be increased in textured objects. The intensity variations of the optimal illumination pattern, as calibrated in Section 5.2.1, have to be adapted according to the natural texture. The illumination pattern should create additional texture, whereas the natural texture, if existing, should remain detectable.

The checkerboard pattern is expressed as a two-dimensional grid. The intensities of the illuminating checkerboard are calculated with

$$Z_{HF}(x, y, \gamma) = \begin{cases} 255 \gamma, & \text{if } \text{CB}(x, y, t_x, t_y) = 1 \\ 255 (1 - \gamma), & \text{if } \text{CB}(x, y, t_x, t_y) = 0, \end{cases} \quad (5.16)$$

where $\text{CB}(x, y, t_x, t_y)$ is the binary checkerboard pattern with frequency t_x, t_y . The intensity term $\gamma: \mathbb{R}^+ \rightarrow [0, 0.5]$ weighs the projected texture pattern. This value is chosen according to the natural object texture. If the surface has no detectable texture γ is 0 and if it has enough natural texture γ is 0.5. In Figure 5.13 the focus measurement of

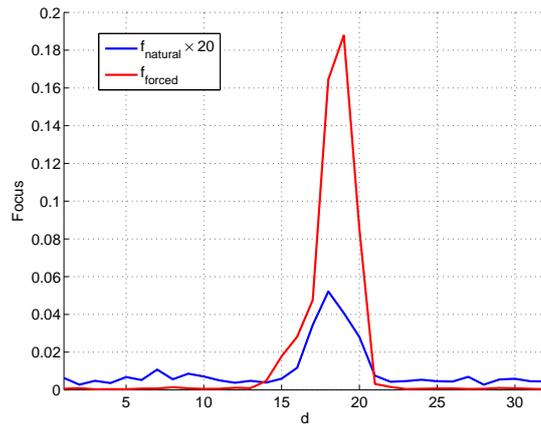
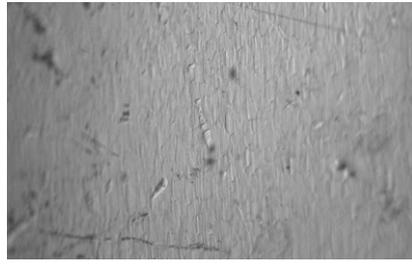


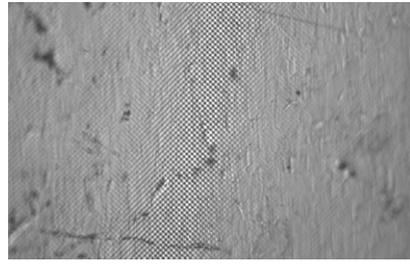
Figure 5.13: Focus measure improvement with HF pattern illumination. Focus measure of one point through the SFF image stack, with natural texture only, and with a projected texture pattern. The dataset of the focus measurement was multiplied by 20 to increase the visibility in the plot.

one pixel with and without projected texture through the image stack is illustrated. The focus measurement with projected texture produces a higher maximum when the pixel is in focus and is lower when it is defocused, which maximizes the sensitivity and the robustness of the focus measure operator.

In Figure 5.14 a reconstruction of a sample plane is shown. The surface texture is supported by pattern illumination. Therewith both the robustness and the sensitivity



(a) Stack image with standard illumination



(b) Stack image with HF pattern illumination

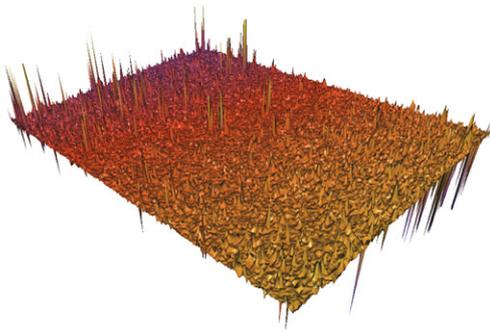
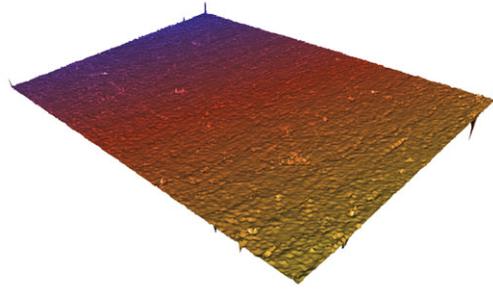
(c) Low natural texture $m_{x,y} = 5$ (d) Supporting texture $m_{x,y} = 3, t_{xy} = 6.8$

Figure 5.14: Support of the natural texture by a frequency pattern to increase accuracy and spatial resolution. The intensity term γ is 0.2.

are increased, while minimizing the size of the focus measure operator. The minimum and maximum irradiance values of the projected pattern are adapted to be not stronger than the natural texture, in this case to intensity term γ is 0.2. The additional texture increases the accuracy, whereas the size of the focus measure operator can be reduced to the minimum of 3×3 .

5.3 Combination of Adaptive Illumination with High Frequency Pattern

In this section a combination of the radiometric compensation from Section 5.1 and the enhancement of natural texture with texture pattern projection from Section 5.2 is presented. As the radiometric compensation does not handle lack in measurable texture and the projected texture pattern can easily lead to under-exposed or over-saturated image parts due to high dynamic range we combined these two methods to overcome both problems. In Figure 5.15 the advantages of this combination is illustrated. On surfaces with

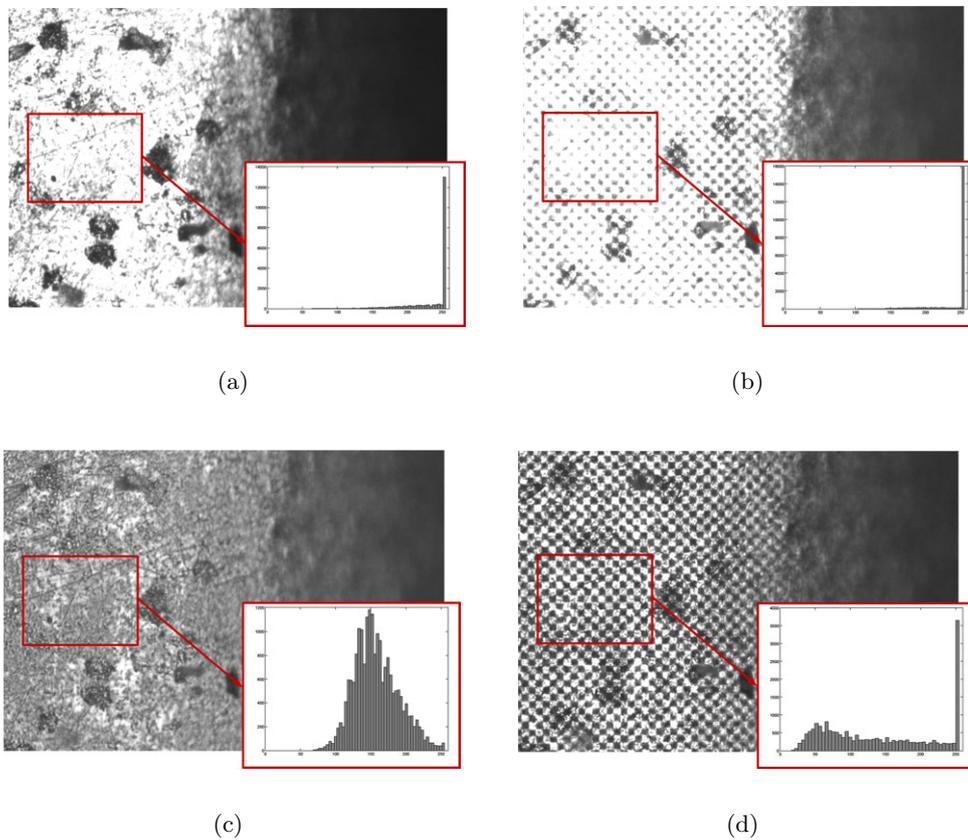


Figure 5.15: Combination of intensity adaption and HF pattern projection. Texture pattern projection alone leads to over exposed parts (b), whereas intensity compensation does not create additional texture (c). The combination creates additional texture and compensates the object reflectance, which leads to an increasing image contrast (d).

high reflection differences and low texture, standard texture pattern projection leads to extremely over exposed parts, whereas standard compensation does not create additional

texture. A combination leads to additional texture while the natural surface reflectance is compensated, which increases the contrast of the image.

In the following two possible combinations of the adaption and the high frequency (HF) pattern are presented. In Section 5.3.1 an approach is presented, where the illumination pattern is created by a slightly modified HF pattern added to the compensation pattern. Another approach is to modify the adaption process itself, and optimize for a desired HF pattern instead of a desired constant illumination, as presented in Section 5.3.2.

5.3.1 Adaption with Added HF pattern

In this section a combination of standard adaption for radiometric compensation and a modified HF pattern for texture enhancement through addition is presented. The com-

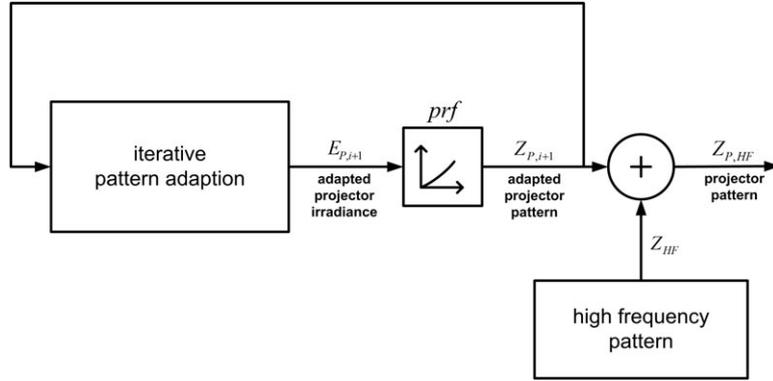


Figure 5.16: Iterative pattern adaption $Z_{p,i}$ combined with a high frequency pattern Z_{HF} to get the final projector pattern $Z_{p,HF}$.

combination of the iterative pattern adaption and the high frequency pattern is shown in Figure 5.3.1. To merge the results of both techniques a high frequency pattern $Z_{HF,mod}$ which has to be slightly modified from the pattern calculated above $Z_{HF,norm}$ is added to the calculated iterative adapted pattern, where

$$Z_{HF,mod}(\psi) = \begin{cases} 255\psi, & \text{if } Z_{HF} = 255, \\ 255(-\psi), & \text{if } Z_{HF} = 0. \end{cases} \quad (5.17)$$

The influence variable $\psi: \mathbb{R}^+ \rightarrow [0, 1]$ controls the weighting between the high frequency pattern and the adapted pattern. The resulting pattern is calculated with

$$Z_{p,HF} = Z_{p,i+1} + Z_{HF,mod}(\psi), \quad \text{where } Z_{p,HF}: \mathbb{R}^+ \rightarrow [0, 255]. \quad (5.18)$$

In the experimental evaluations we determined the weighting variable ψ to be between 0.25 and 0.3. The exact value of ψ is set so that the percentage of under- or over-exposed image regions is beneath one percent over the whole surface. Generally, the higher the difference in the surface reflectance, the lower is ψ and vice versa. The problems of this method are that the HF pattern and the adapted pattern cancel each other out in some degree. On the one hand, the intensity of the HF pattern is reduced through ψ . On the other hand, the adapted pattern loses the amount of adaptivity through the HF pattern. For example if a region of the surface has to be fully illuminated to compensate the reflectance an additional texture pattern results in a worse compensation.

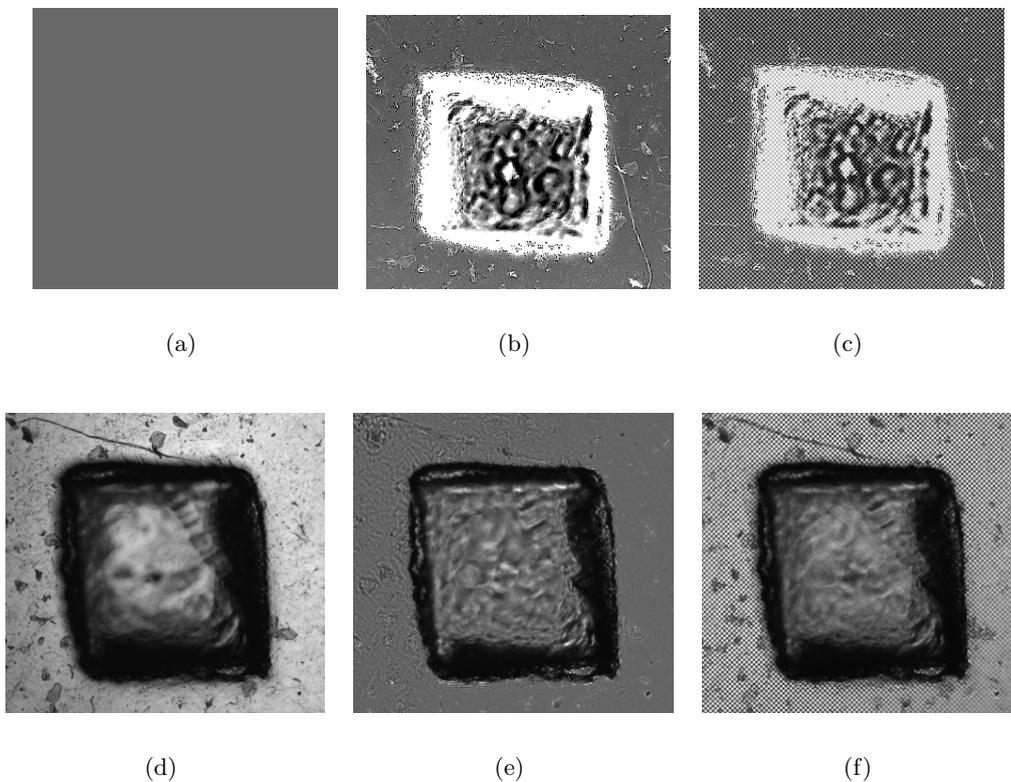


Figure 5.17: Example of a specimen with different illumination. In (a-c) the projection pattern and in (d-f) the camera images of a specimen with different illumination are shown. (a,d) homogeneous illumination, (b,e) adapted illumination and (c,f) combination of adapted and high frequency illumination with $\psi = 0.2$.

In Figure 5.17 the combination of the projector pattern and the corresponding camera images are shown. In Figure 5.18 the reconstruction results of the surface which is illuminated with an addition of the adapted pattern and the HF pattern. It can be seen, that

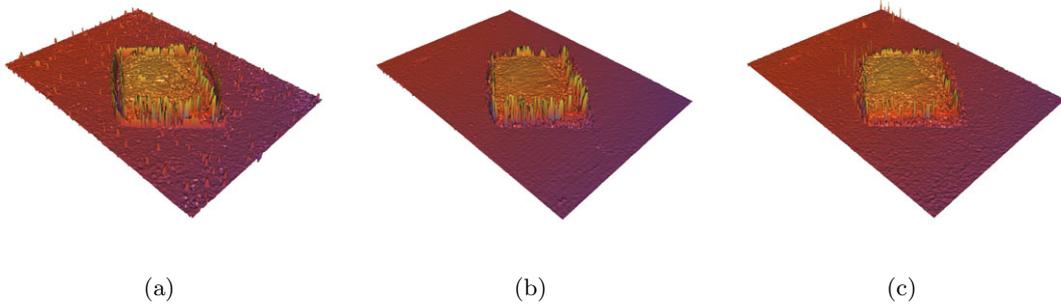


Figure 5.18: SFF reconstruction of a surface with homogenous illumination (a), with HF texture projection (b) and with an adaption with an added HF pattern (c). The low textured plane and the under- or over-exposed parts of the surface are reconstructed more accurately.

the HF part of the illumination pattern creates texture on the low-textured plane parts, which can be measured more accurate (fewer outliers than in standard SFF). Through the compensation the under- and over-exposed parts are better illuminated which also leads to a more accurate focus measurement in these regions.

5.3.2 Adaption of a HF pattern

In this section a method where the high frequency pattern for texture enhancement is used as the desired camera irradiance \hat{E}_c in the adaption process is presented (see Fig. 5.3). There, the HF pattern does not directly influence the adaption pattern but indirectly through the optimization process. In Figure 5.19 a sample iterative adaption process is shown. Because each projector pixel influences more than one camera pixel the frequency of the desired HF camera image has to be set according to the considerations made in Section 5.2.1. Therefore it has to be set to a value higher than the constant influence radius $2\sigma_p$ and lower than boundary from the focus operator $2 + m_{x,y}$. For instance at an influence radius of $\sigma_p = 4.2$ and a focus operator with the summation in a 9×9 window the frequency has to be set to be $8.4 < 1/2t_{x,y} < 11$. The low and the high value of the desired HF camera image have to be set to avoid over- or under-exposed image parts due to inaccurate compensation. To ensure this we set to set the low and high intensity of the checkerboard pattern to 25% and 75% of the maximum intensity (8bit gray-scale $\rightarrow \hat{Z}_{c,min} = 60$ and $\hat{Z}_{c,max} = 190$).

In Figure 5.20 the difference between homogeneous illuminated, HF pattern projection and the combination of reflectance compensation and HF pattern projection is illustrated.

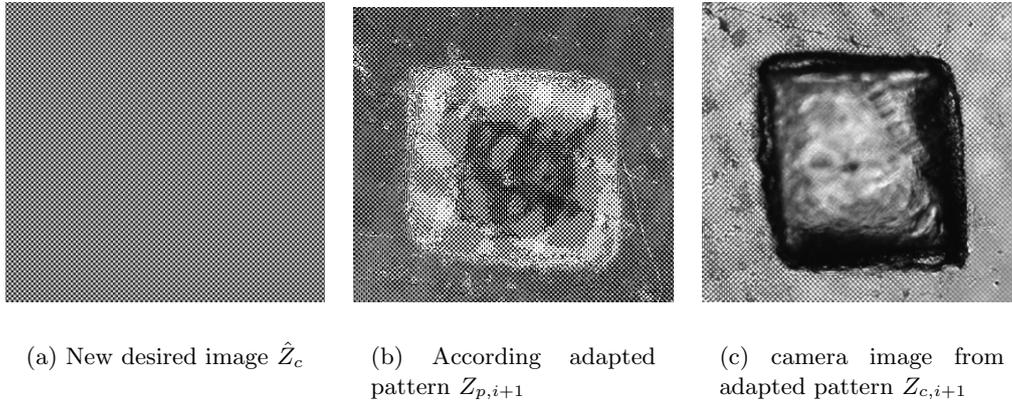


Figure 5.19: Example of the adapted illumination with a HF pattern as desired camera image. The projector pattern (b) is calculated through the iterative pattern adaption process with the HF pattern (a) as desired image, which results in a camera image (c) with adapted illumination.

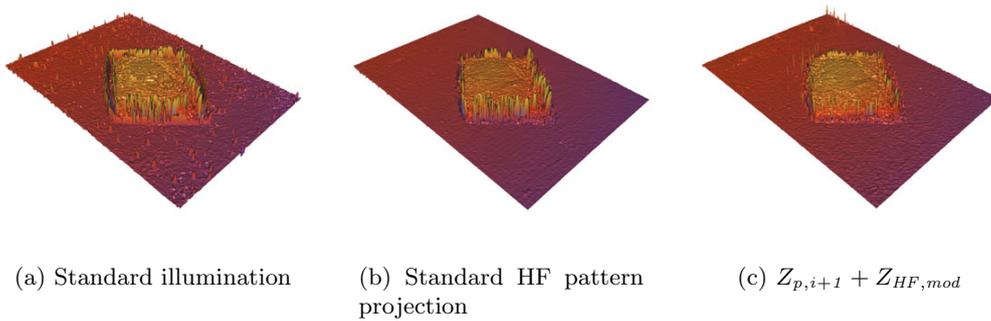


Figure 5.20: SFF reconstruction of a surface with homogeneous illumination (a), with HF pattern projection (b) and with an adapted HF pattern (c). Both the low textured and the under- or over-exposed parts of the surface are reconstructed with fewer outliers.

With a forced texture illumination a dense focus measurement in low-textured surface regions is enabled, whereas the amount of measurable texture in the over- and under-exposed image regions is increased, resulting in a more accurate and more robust depth reconstruction with less outliers than standard illuminated SFF.

5.4 Integral Shape From Focus

In this section a novel focus measure method for calculating a shape from focus reconstruction with the assistance of a focused texture image is presented. The basic principle

behind this approach is to compare the gray values of each image from the image stack with a known all-in-focus texture image. Through this comparison we get a similarity value for each pixel in the image stack to the focused image through correlation. This similarity value is comparable with the focus measure operator in traditional SFF. After the calculation of all similarity values through the stack the maximum value can be interpolated through an interpolation method, as presented in Section 3.1.4.

The advantage of this novel method is the characteristic of the NCC. Because the correlation value is normalized between -1 and 1 the focus calculation through the focus stack can be stopped at a local maximum above a predefined threshold. If the focused surface points are evenly distributed in the image stack, this increases the reconstruction speed by an average factor of two compared to traditional depth reconstruction.

The most challenging problem of this approach is to calculate a focused texture image out of the image stack without any knowledge of depth. Nagahara et al. [40] presented a method for extending the range of scene depths that appear focused in an image, known as the depth of field (DOF). In this method the position of the image detector to the lens varies during the integration time of a single photograph. This photograph is called the integral image. In this image, each scene point is captured under a continuous range of focus settings, including perfect focus. Applying a deconvolution to this integral image with a calibrated integral point spread function (IPSF) results an image where the entire scene appears focused.

We use this method to calculate an all-in-focus image out of an image stack. The IPSF is calibrated in Section 5.4.1. The application of the IPSF to calculate a focused texture image and the reconstruction of the surface using a correlation based focus measure is presented in see Section 5.4.2.

5.4.1 Calibration of the IPSF

The IPSF is a two-dimensional function which aggregates the point spread function (PSF) of one image point through the image stack. Thus, it models the defocusing of one focused surface point. Conversely, if an integral image is convolved with the IPSF, the all-in-focus image can be restored. We assume that the IPSF is constant in different image stacks and equal for every pixel in an image and is calculated in a previous calibration step.

To calibrate the IPSF we use a set of blurred images $I_{b,i}$ without any knowledge of the amount of defocus and a focused image I_t of a known a scene and want to know the blurring function h . If the calibration surface is a plane, normal to the optical axis, I_t

is the image of the stack, where the plane is perfectly in focus. Otherwise, the focused texture image I_t is calculated through traditional SFF. The blurred images from the image stack are summed up to an integral image I_{int} that incorporates different degrees of focus through the whole stack, including perfect focus (according to [40]). This is necessary to calculate a robust blurring function h . To model a simple estimation of the blurring function we first look at the simple filtering, which is calculated with

$$I_b(x, y) = I_t(x, y) * h(x, y) + n(x, y) \quad \text{spatial domain} \quad (5.19)$$

$$I_b^F(u, v) = I_t^F(u, v)H(u, v) + N(u, v) \quad \text{frequency domain,} \quad (5.20)$$

where $n(x, y)$ and $N(u, v)$ is the image noise and $h(x, y)$ and $H(u, v)$ are the degradation function of the spatial and frequency domain. I_b is the blurred image of the scene. Therewith, a simple approach for calculating the degradation function is

$$H(u, v) = \frac{I_b^F(u, v)}{I_t^F(u, v)} \quad (5.21)$$

The main problem of the filter estimation from (5.21) is the ignorance of noise, which is normally unknown. To overcome this problem we transform this linear filter estimation into an optimization problem. We want to optimize the IPSF in order to minimize the error between the known focused texture image I_t and the calculated focused image \hat{I}_t . The optimization problem can be expressed through a minimization of the quadratic error between the calculated focused image and the real focused image:

$$\operatorname{argmin}_{ipsf} \{ ||I_{int}(x, y) * ipsf(x, y) - I_t(x, y)||^2 \}. \quad (5.22)$$

This optimization problem is solved through non-linear least squares optimization which is presented in [39]. A more robust way of solving this problem is the total variation [54] where an additional smoothing term is introduced. A special algorithm for blind deconvolution based on total variation is presented in [9]. Figure 5.21 shows a cross-section of the optimized IPSF with different optimization methods, whereas the TV-L1 optimization approach delivers the best result. These optimized IPSFs are compared to the integration of the calibrated PSFs of the projector-camera system, as calibrated in Section 4.3.3. With this optimized IPSF, the focused texture image is calculated.

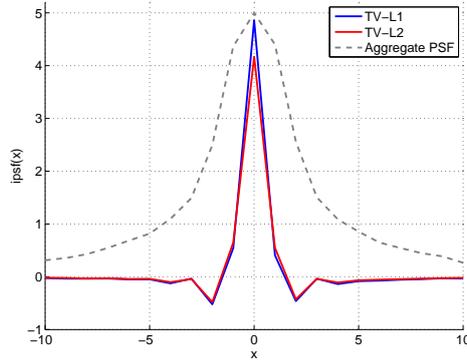


Figure 5.21: Cross-section of IPSF optimization results. The IPSF optimized with TV-L1 regularisation and with a TV-L2 approach compared to the integration of the calibrated PSFs.

5.4.2 Application of the IPSF for Correlation Based SFF Reconstruction

This section explains the calculation of a focused texture image out of an image stack without any information about depth and degree of defocus, which is used for a correlation based focus measure for SFF reconstruction.

Calculation of a Focused Texture Image

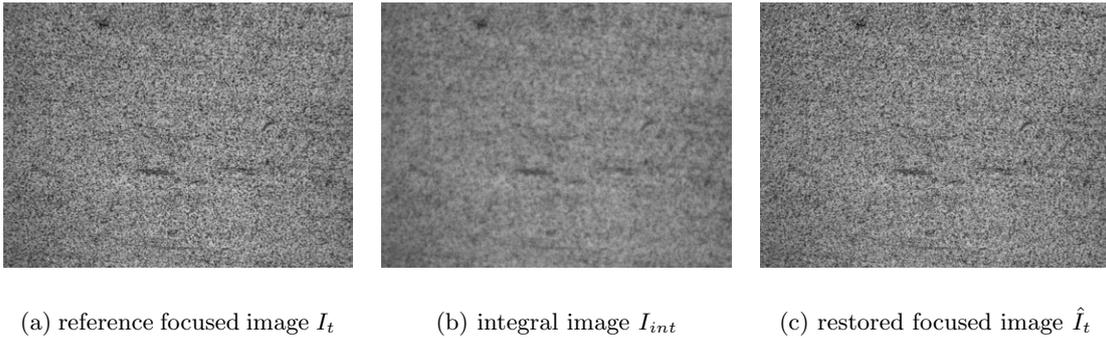


Figure 5.22: After optimization of the IPSF with the focused texture image I_t and the integral image I_{int} the focused image \hat{I}_t is calculated by convolution with the IPSF.

The IPSF is optimized with the reference focus image and the integral image. With the calculated IPSF the focused texture image \hat{I}_t is restored out of the integral image I_{int} ,

according to (5.22), through

$$\hat{I}_t(x, y) = I_{int}(x, y) * ipsf(x, y). \quad (5.23)$$

Because the IPSF is constant for different specimen, it is used to recover a sharp texture image out of every acquired image stack by convolution of the integral image from the stack with the degradation function. In Figure 5.22 the focused image is calculated from the calibration scene.

Focus Measurement through Cross Correlation

By correlation of the focused texture image with the grey value information of the acquired images a SFF focus measure is calculated. This is done by a patch-wise correlation of every image pixel through the image stack with the corresponding patch in the all-in-focus image. The correlation result is our new focus measurement. The correlation focus

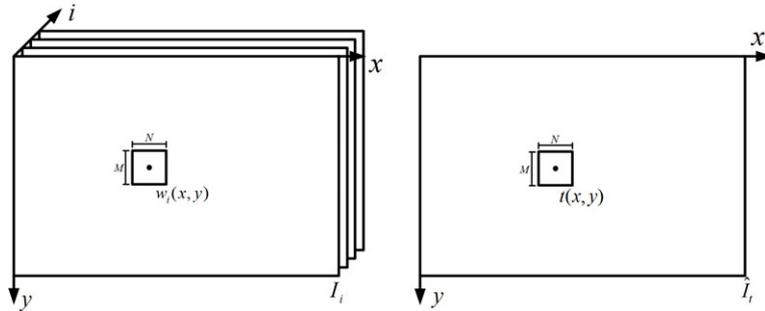


Figure 5.23: Corresponding correlation patches. The patch $w_i(x, y)$ from image i in the image stack around the pixel (x, y) is correlated with the corresponding patch $t(x, y)$ at the same position in the all-in-focus image.

measure algorithm calculates the cross correlation of a patch w_i of the size $M \times N$ around a stack image pixel $I_i(x, y)$ with the patch t of the calculated focused image \hat{I}_t with the same size and pixel position (see Fig. 5.23). This is done for every pixel and each image of the SFF image stack, resulting in a stack of correlation images C . The cross correlation is calculated with:

$$C_i(x, y) = \sum_m \sum_n w_{i,x,y}(m, n) t_{x,y}(m, n) \quad (5.24)$$

With the calculation of the traditional cross correlation several disadvantages occur, according to [33]:

- The simple cross correlation is not invariant to changes in amplitude. But, due to the integration of the images the focused image varies in image amplitude and mean compared to normal stack images.
- If the image energy $\sum w^2(x, y)$ varies between w and t the matching can fail. E.g. correlation between equal patches may be less than the correlation between a patch and a bright spot.

Instead we use the normalized cross correlation (NCC). The normalized correlation coefficients are calculated with

$$\gamma(x, y) = \frac{\sum_{m,n}[w_{x,y}(m, n) - \bar{w}_{x,y}] \sum_{m,n}[t_{x,y}(m, n) - \bar{t}_{x,y}]}{\left\{ \sum_{m,n}[w_{x,y}(m, n) - \bar{w}_{x,y}]^2 \sum_{m,n}[t_{x,y}(m, n) - \bar{t}_{x,y}]^2 \right\}^{1/2}}, \quad (5.25)$$

where $\bar{w}_{x,y}$ and $\bar{t}_{x,y}$ are the mean values of the patch around (x, y) . The calculated correlation coefficients $\gamma(x, y)$ are in the range $[-1, 1]$ and are thus normalized to changes in amplitude and energy of w and t . The maximum value occurs when normalized w and normalized t are identical, which in our case is, when the patch w is in focus.

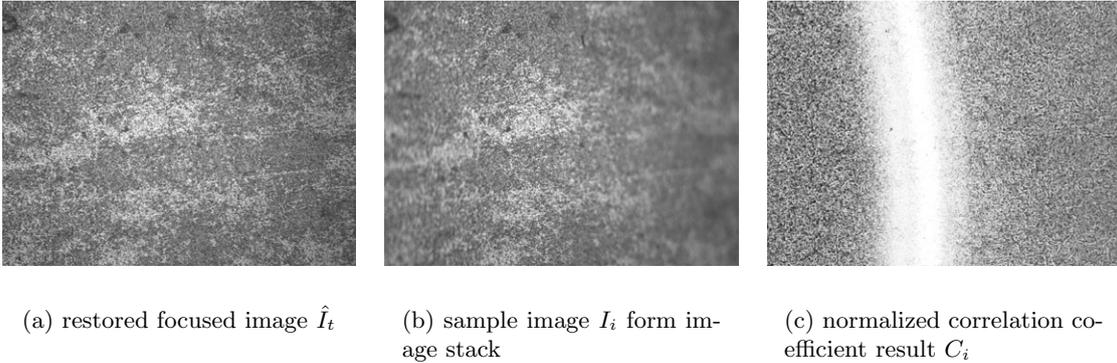


Figure 5.24: Calculating the normalized correlation coefficients of the calculated focused texture and a sample stack image. The resulting correlation image delivers a map of similarity values of each pixel region between the images.

In Figure 5.24 the normalized correlation coefficient result of a skew plane is illustrated. The sample image I_i from the image stack is correlated with the calculated focused image

\hat{I}_t . The values in the resulting correlation matrix C_i correspond to the focus. When calculating the normalized correlation coefficients for each stack image I_i we get a stack of correlation results. Like in traditional SFF, the depth is calculated with the maximum value of each pixel through this stack. To speed up the depth estimation of the image

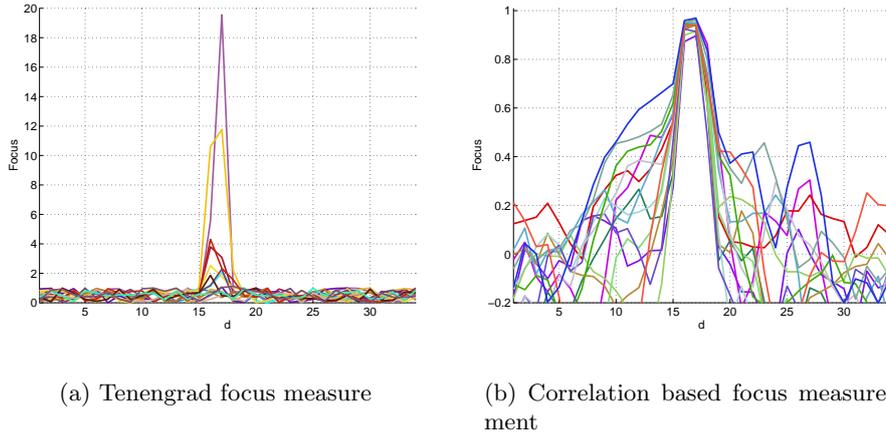


Figure 5.25: Focus measurement through the image stack. The focus measurement is calculated for a cross-section of an surface with the same depth through the image stack with the Tenengrad focus measure (a) and the NCC based focus measure (b).

which is in focus, the NCC focus measure has the advantage that it is normalized between -1 and 1 . Therefore a global threshold can be set to define the focus maximum. After a local maximum above the threshold is found, the depth calculation for this image point is finished. This additional constraint reduces the computation time in average by half. This method can not be used in traditional focus measures, because the focus values have no common dimension, as seen in Figure 5.25. To avoid holes in the depth map the global maximum is calculated if the maximum of correlation coefficients through the stack is beneath the threshold.

To increase the robustness to outliers, the correlation results are additionally summed up in a local window similar to other focus measure operations with

$$FM_{corr}(x_0, y_0) = \sum_{p(x,y) \in U(x_0, y_0)} (C(x, y) + 1), \quad (5.26)$$

where $p(x, y)$ a pixel in the summation patch around the center pixel (x_0, y_0) . The addition of one has to be made to transfer the correlation results into \mathbb{R}^+ . Otherwise positive and negative values within a patch would cancel each other out.

With this focus measure the depth map is calculated by interpolating the pixel vector of each image position through the stack, according to Section 3.1.4.

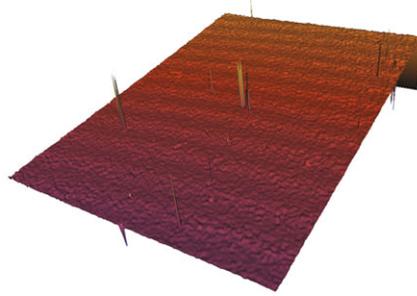


Figure 5.26: Reconstruction of a plane with correlation based SFF.

In Figure 5.26 a plane is reconstructed using correlation based SFF with a texture image calculated with a calibrated IPSF.

5.5 Summary

In this chapter novel methods to increase speed, accuracy and robustness of the SFF object reconstruction were presented. The SFF acquisition system was supported by a projector used as a locally adaptive light source.

Homogeneous illumination during the image acquisition often leads to over- or under-saturated regions in the image. In these regions an exact focus measurement is not possible. To avoid these illumination artefacts we used the projector to adapt the illumination intensity locally. The compensation algorithm equalizes the reflected irradiance over the whole acquired surface to achieve a largely uniform camera image without over- or under-exposed parts.

A second problem in traditional SFF reconstruction is the failure of a dense focus calculation of low textured objects. To overcome this problem we used the projector to create additional texture on the object by projecting a high frequency texture pattern. Through this additional texture the degree of focus becomes measurable again. Because these two methods did not address both problems at once we invented a combination of the illumination adaption and the texture projection, where both illumination artefacts due to differences in the object reflection and the absences of measurable texture are eliminated.

To increase the speed of the traditional SFF algorithm we introduced a novel depth focus measurement which is calculated through a patch-wise NCC of the images from the image stack with calculated all-in-focus texture image. This all-in-focus texture image is

calculated through the accumulated image stack, the integral image, convoluted with a calibrated deblurring function, the IPSF. Because the NCC result is normalized between -1 and 1 the focus calculation through the image stack can stop when a local maximum above a predefined threshold is found.

Chapter 6

Evaluation and Experiments

Contents

6.1 Adaptive Illumination	90
6.2 Projected Texture	100
6.3 Combination of HF Pattern with Adapted Illumination	106
6.4 Integral Shape From Focus	110
6.5 Summary	118

In this chapter the experiments and evaluations of our work are presented in detail. The accuracy and robustness of the presented algorithms are compared to traditional SFF reconstruction. All experiments are performed on a microscope prototype exhibiting a 2MP 8bit grey scale charge coupled device (CCD) camera and a 1280×720 pixel liquid crystal on silicon (LCoS) display as light source. To move the object through the focus plane we use a translational stage with an accuracy of $\pm 0.1\mu m$. The calibrated depth step is $5.059\mu m$ and the lateral resolution in the camera image is $0.740 \times 0.740\mu m$ per pixel. The SFF algorithm calculates the reconstruction with a Tenengrad focus measure operator with a window size of 7×7 . For the depth step interpolation a Gaussian model is used. No pre- or post-processing is performed, and parametrization is identical for all evaluations.

In the following sections, the methods presented in Chapter 5 are evaluated. In Section 6.1 the evaluation of the adaptive illumination is presented. The texture enhancement by projecting a HF texture pattern on the specimen is evaluated in Section 6.2. In Section 6.3 the experimental results for the combination of an adapted illumination with HF pattern projection is presented. The integral shape from focus (ISFF) method is evaluated

in Section 6.4.

6.1 Adaptive Illumination

In this section we want to evaluate the reconstruction results of the adaptive illumination, presented in Section 5.1. In this method, the reflectance of an arbitrary surface is compensated to get the same level of focus information over the whole acquired surface, because both saturated and under-exposed image regions do not deliver enough variations to calculate a dense focus measure.

The adapted illumination pattern is calculated and refined iteratively in every step. To increase the speed of this adaption process and to decrease the number of camera acquisitions per step, the compensation pattern is tracked through the image acquisition and an additional pattern prediction is calculated. This leads to a reduction to one camera acquisition and one adaption iteration per step. To show that the predictive adaption gives the same compensation results as the standard iterative adaption we evaluate the compensation differences in Section 6.1.2.

Through the evaluation of accuracy and robustness we show, that on industrial specimen with high dynamic range the adaptive illumination (*ADAPT*) delivers superior results compared to traditional SFF with homogeneous illumination (*STD*). Furthermore, we want compare *ADAPT* to comparable reconstruction methods like HDR imaging constructed from varying numbers of input images, where the dynamic range of the camera is extended. The reconstruction results for these evaluations are presented in Section 6.1.3.

In Section 6.1.1 the different specimen used for our evaluation are presented.

6.1.1 Reference Specimen

In this section the specimen used for accuracy evaluation are presented. These objects have a varying reflectance in order to analyse the performance of our work in various scenarios.

The main reference specimen are six regions on three cutter blades, namely *CUT1 - side*, *CUT1 - front*, *CUT2 - side*, *CUT2 - corner*, *CUT2 - edge* and *CUT3 - edge*, as illustrated in Figure 6.1. These cutter blades are examples of industrial fields of application for quality inspection. The angle of the surface normal to the optical axis and the texture on the cutter blades vary within the regions. For these specimen no external groundtruth is available. Because we want to evaluate the compensation of high dynamic ranges we use

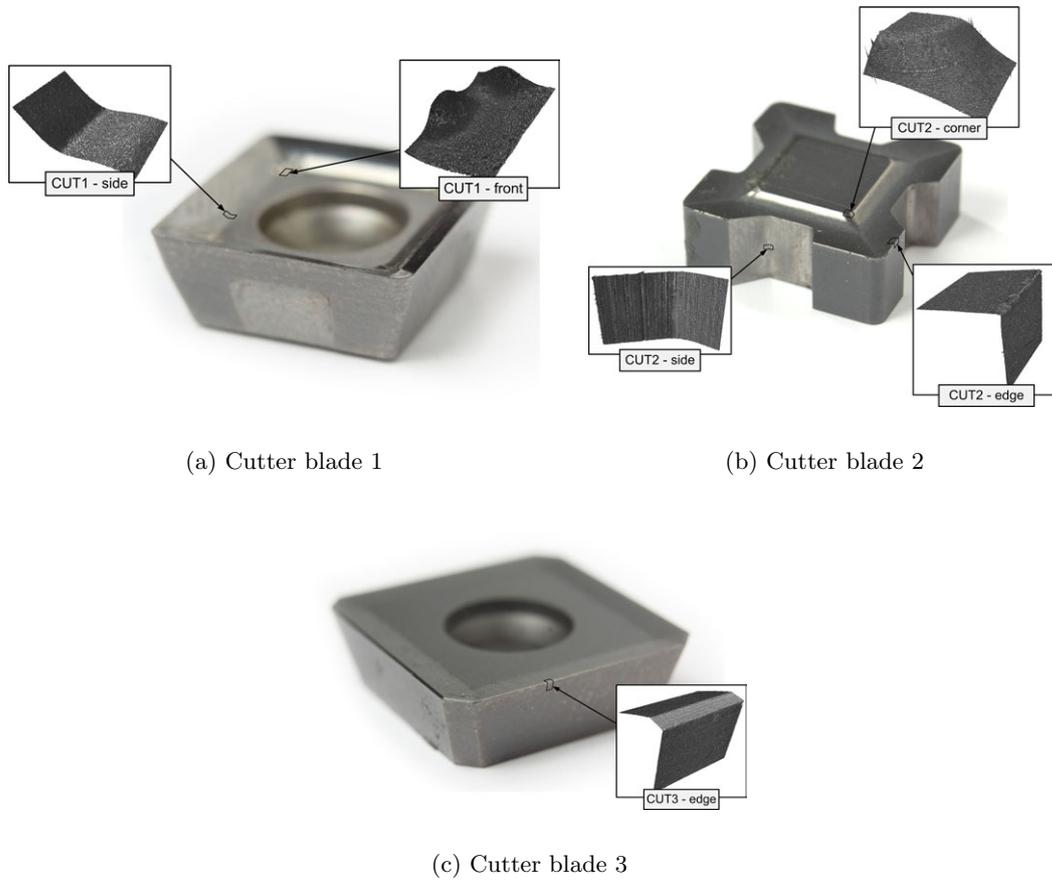


Figure 6.1: Reference cutter blade specimen. Three different cutter blades are chosen for quality inspection. The evaluated regions of the cutter blades are highlighted.

a filtered SFF reconstruction result of scans from HDR images with ten different exposure times (*HDR10*), because this reconstruction ensures that the whole dynamic range is mapped on the acquired images.

In addition to these industrial specimen, two defined specimen are used to calculate the geometric consistency. The specimen are two solid wolfram-carbide samples, resembling a *Wedge* and a *Campfer*. The geometric information is provided from an external measurement protocol, from a DIN EN ISO 5436-1 geometric inspection. With this certified geometric measurement we can exactly compare the geometric reconstruction to a known groundtruth profile. In Fig. 6.2 the profiles of the groundtruth samples are illustrated. These groundtruth profiles are compared against reference scans of reconstructed profiles. Because the geometric consistency of our reconstruction is independent from the illumina-

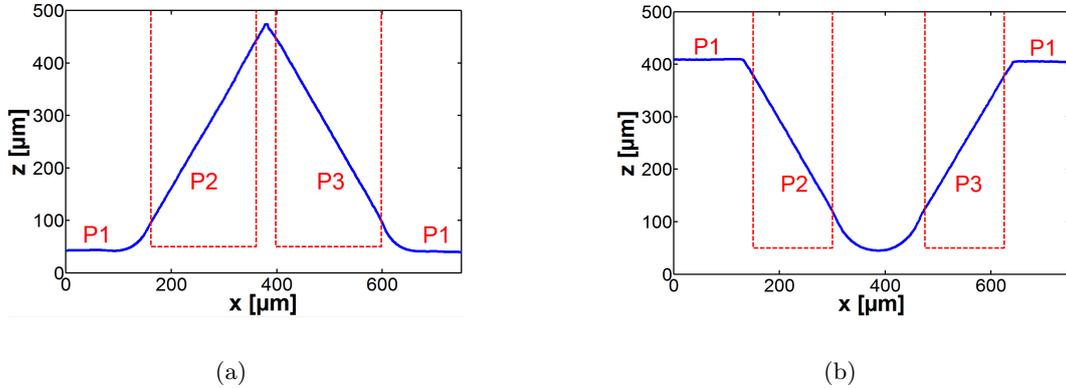


Figure 6.2: Cross-section of reference scans. Profiles of two ss, namely *Wedge* (a) and *Chamfer* (b). Geometric groundtruth exists for the planes P1, P2, P3.

tion, the profiles are calculated with a reconstruction using the high dynamic range SFF method with ten exposure times per image (*HDR10*). The reference profile measurement is computed by taking the mean of more than 900 profiles through the reconstruction.

	α (error) [deg]	β (error) [deg]	d (error) [μm]
<i>W</i>	119.70 (0.53)	59.91 (-0.07)	437.34 (-1.66)
<i>C</i>	120.15 (0.06)	60.52 (0.33)	440.14 (-0.86)

Table 6.1: Comparison of the dense reference profiles and geometric groundtruth for *Wedge* (*W*) and *Chamfer* (*C*).

In Table 6.1, the angles in the groundtruth object $\alpha = P1\angle P2$, $\beta = P2\angle P3$, and the distance d from the intersection of $P2$ and $P3$ to $P1$ are given, and compared to the groundtruth measurement from the measurement protocol. This result determines the geometric correctness of the depth reconstructions.

6.1.2 Predictive Adaption

In this section the illumination adaption in a prediction correction step to increase the compensation speed is evaluated and compared to the iterative adaption. Therefore, we acquired two SFF image stacks. One with an iterative compensation (*ADAPT-IT*), where the compensation is calculated in each step separately, and the predictive compensation (*ADAPT*), where the pattern is tracked and predicted through the stack. To measure the

quality of the compensation, the grey value progress of the images of these two stacks are compared.

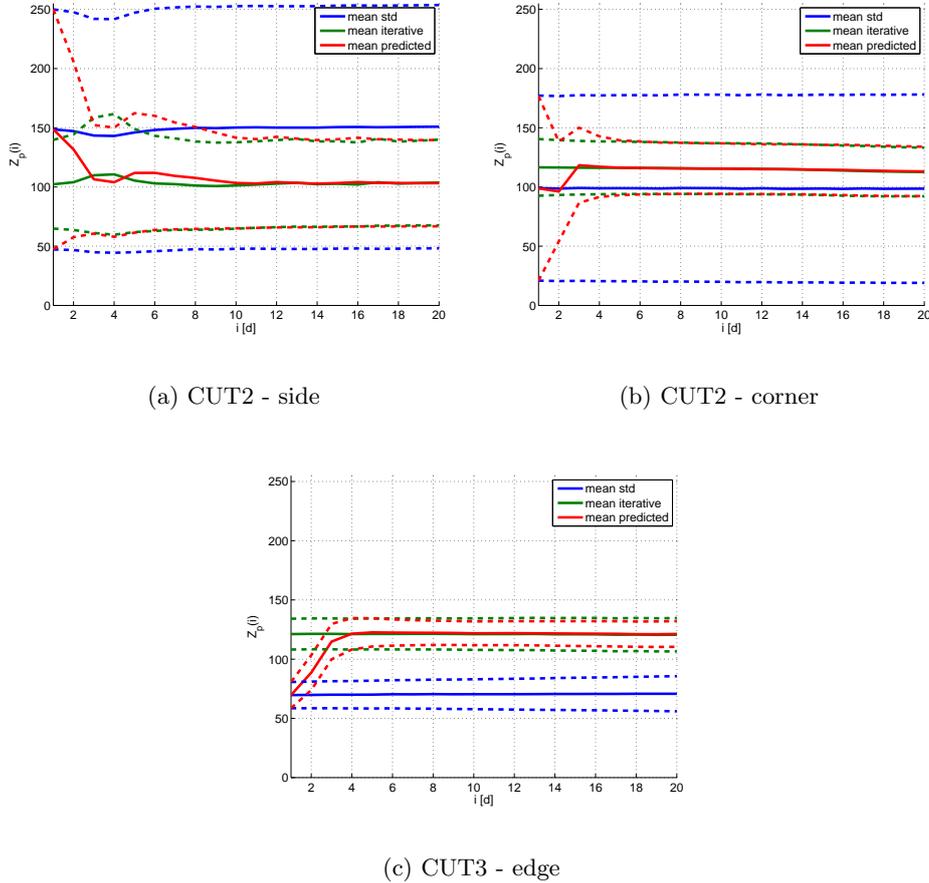


Figure 6.3: Quality evaluation of predictive compensation. Evaluation of mean and standard deviation in the image stack for homogeneous illumination (*STD*), iterative compensation in three iterations (*ADAPT-IT*) and predictive adaption (*ADAPT*) for the reconstruction of three different specimen. Solid lines indicate the mean progress and dashed lines the standard deviation over the first $i = 1 \dots 20$ acquired images.

To determine the best possible compensation result we set the number of iterations to 3 (4 image acquisitions required) for iterative adaption, where the iterative compensation pattern refinement converges to the optimum. We compare mean image grey value and the standard deviation of every acquired image through the image stack of both iterative compensation and predicted compensation. These values indicate the quality of the compensation. Therefore, the mean value of the compensated camera images should converge to the desired camera image \hat{Z} , which is set to a uniform image with grey value

128. Furthermore, the standard deviation of the image grey values should be minimized. In Figure 6.3 the progress of mean and standard deviation of the stack images are illustrated. In this Figure it is shown, that *ADAPT* starts at the mean and standard deviation values of the reconstruction with homogeneous illumination (*STD*) and converges to the *ADAPT-IT* in at least three steps.

6.1.3 Reconstruction with Adaptive Illumination

In this section the relative improvement of the reconstruction results for an adaptive illumination (*ADAPT*) of the specimen is evaluated. These SFF results are evaluated against SFF on an image stack with homogeneous illumination (*STD*), on HDR images constructed from two (*HDR2*), three (*HDR3*), four (*HDR4*) and five (*HDR5*) different exposure times. We followed the method in [14] for selecting suitable exposure times for HDR imaging, but changed the original set of the exposure times $\Delta t = \{\Delta t_0 * 2^n : n = -k, \dots, k\}$ to a narrower stepping, individually adapted to the specimen, because the original exposure stepping did not give satisfactory results, except of the three-exposure case (*HDR3*). Table 6.2 shows the relative exposure times for all HDR evaluations, where Δt_0 is the manually selected exposure time for the single view case (*STD*).

Method	Relative Δt [%]
<i>STD</i>	100
<i>HDR2</i>	50, 150
<i>HDR3</i>	50, 100, 200
<i>HDR4</i>	25, 75, 150, 200
<i>HDR5</i>	25, 50, 100, 200, 400
<i>HDR10</i>	10, 25, 50, 75, 100, 125, 150, 175, 200, 400

Table 6.2: Relative exposure times Δt for image acquisition during constant illumination. *HDR(X)* designates HDR imaging with X exposure times. *HDR10* is only used for the generation of dense reference data.

The initial projector pattern for the illumination adaption is always a homogeneous uniform pattern with gray value 128 (half of the maximum intensity). To ensure the same possibilities also for standard and HDR approaches, the exposure time was selected to result in a mean intensity close to 128 in the camera image.

To calculate the exact difference of every reconstruction to the reference scan (*HDR10*), the result is rigidly registered to the reference scan using a robust variant of the iterative

closest point algorithm, according to [6]. Therefore the degree of error can be expressed through the standard deviation of the surface height difference.

Method	Wedge		Chamfer		CUT1 - front		CUT1 - side	
	σ	#imgs	σ	#imgs	σ	#imgs	σ	#imgs
<i>STD</i>	3.694	73	3.195	65	6.220	40	79.939	61
<i>HDR2</i>	1.255	146	1.429	130	3.661	80	18.133	122
<i>HDR3</i>	1.215	219	1.371	195	1.921	120	18.140	183
<i>HDR4</i>	1.173	292	1.339	260	1.675	160	2.957	244
<i>HDR5</i>	1.178	365	1.325	325	1.478	200	3.175	305
<i>ADAPT</i>	1.219	75	1.345	67	1.664	42	2.443	63
Method	CUT2 - side		CUT2 - corner		CUT2 - edge		CUT3 - edge	
	σ	#imgs	σ	#imgs	σ	#imgs	σ	#imgs
<i>STD</i>	13.100	78	37.960	58	14.30	79	21.6	70
<i>HDR2</i>	7.900	156	6.4294	116	6.1294	158	4.7861	140
<i>HDR3</i>	7.6840	234	4.0276	174	6.0273	237	4.1931	210
<i>HDR4</i>	1.9510	312	3.2380	232	5.5419	316	3.5216	280
<i>HDR5</i>	1.9406	390	2.1125	290	4.8792	395	1.9546	350
<i>ADAPT</i>	2.6046	78	1.9721	58	4.9082	79	2.7244	70

Table 6.3: Reconstruction errors (standard deviation of depth errors) for all approaches over the specimen in $[\mu\text{m}]$. One reconstruction consists of approximately $1.4e6$ depth estimates. The number of image acquisitions for each approach is stated as #imgs. HDR results closest to our approach are highlighted.

Table 6.3 gives the reconstruction errors for all reconstructed specimen and all approaches. The reconstruction results of the reference specimen are illustrated in Figure 6.4 and Figure 6.5, whereas the corresponding focused texture images for standard and adaptive illumination are illustrated in Figure 6.6 and Figure 6.7.

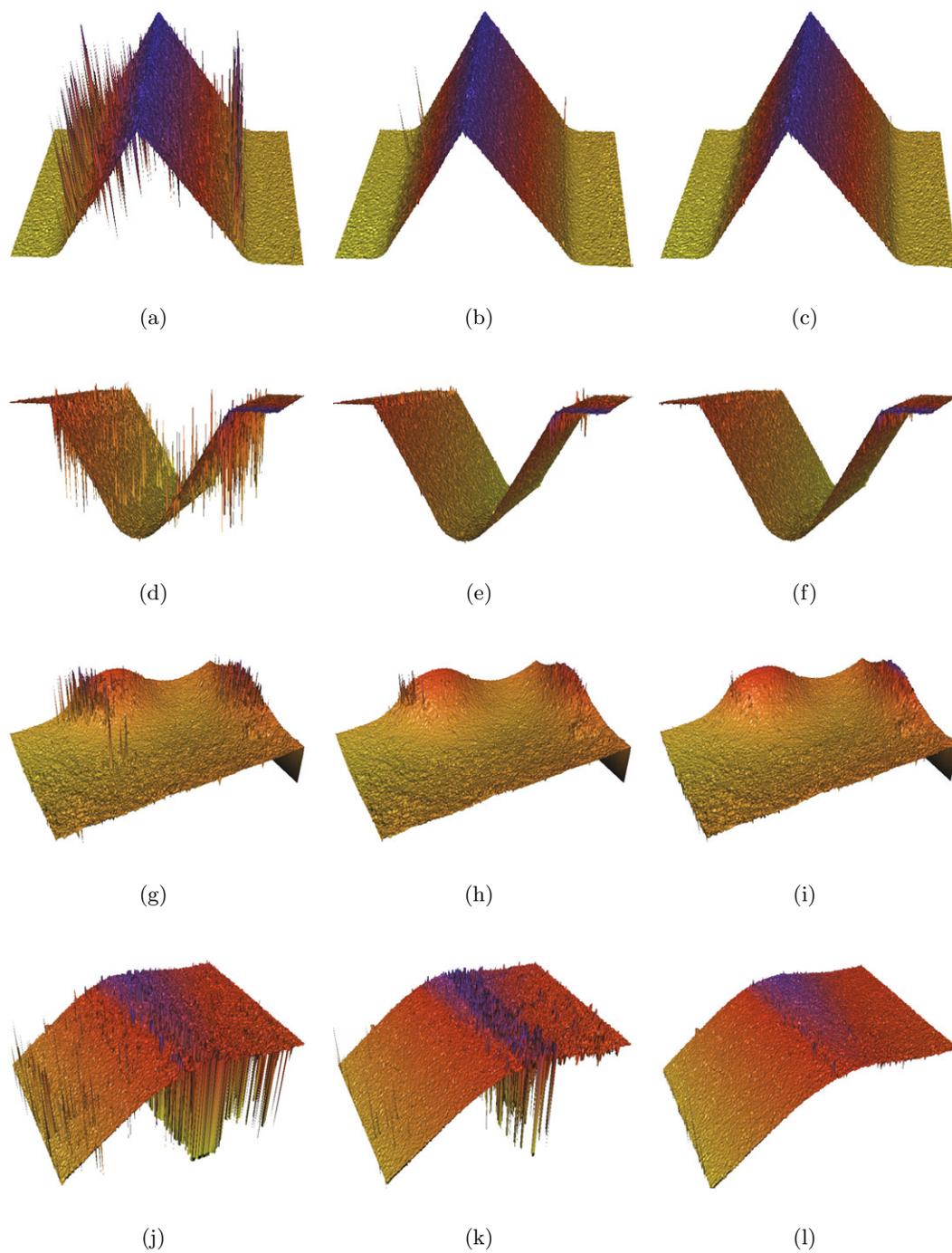


Figure 6.4: Reconstruction results of specimen *Wedge* (a-c), *Chamfer* (d-f), *CUT1 - front* (g-i) and *CUT1 - side* (j-l). SFM results calculated by *STD* (a,d,g,j), *HDR3* (b,e,h,k) and *ADAPT* (c,f,i,l).

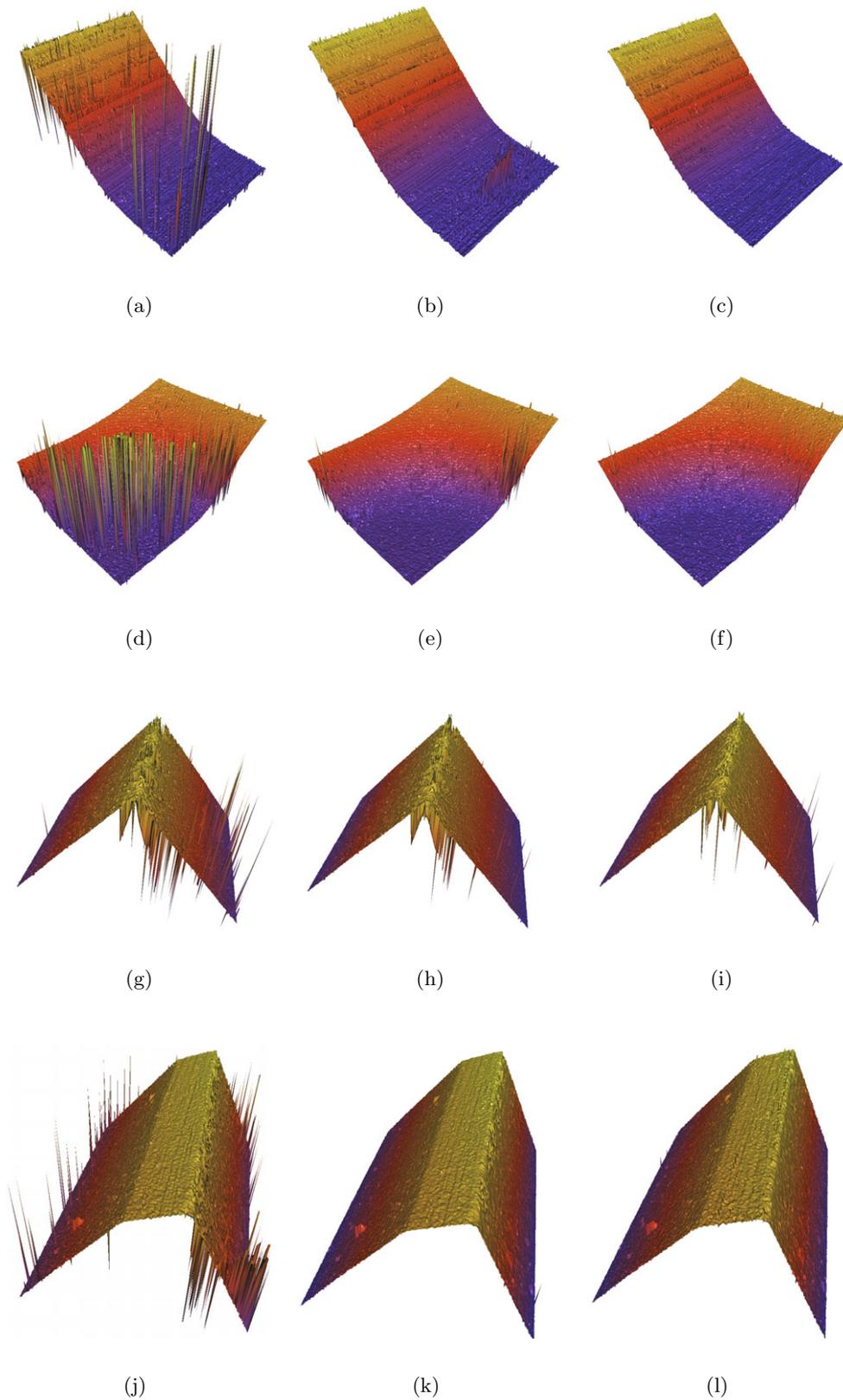


Figure 6.5: Reconstruction results of specimen *CUT2 - side* (a-c), *CUT2 - corner* (d-f), *CUT2 - edge* (g-i) and *CUT3 - edge* (j-l). SFF results calculated by *STD* (a,d,g,j), *HDR3* (b,e,h,k) and *ADAPT* (c,f,i,l).

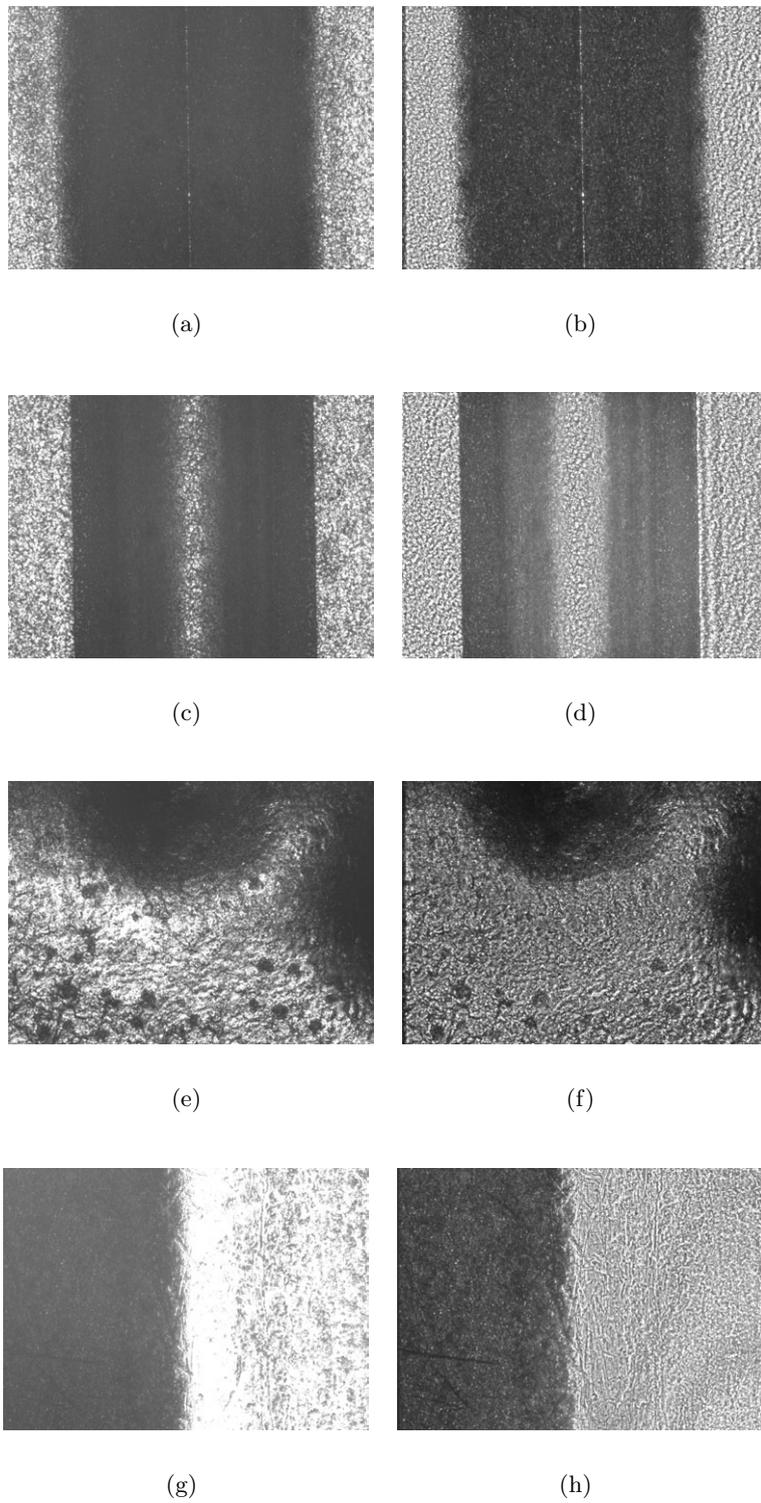


Figure 6.6: All-in-focus images. The focused images are compared between *STD* (a,c,e,g) and *ADAPT* (b,d,f,h) for the specimen *Wedge* (a,b), *Chamfer* (c,d), *CUT1 - front* (e,f) and *CUT1 - side* (g,h).

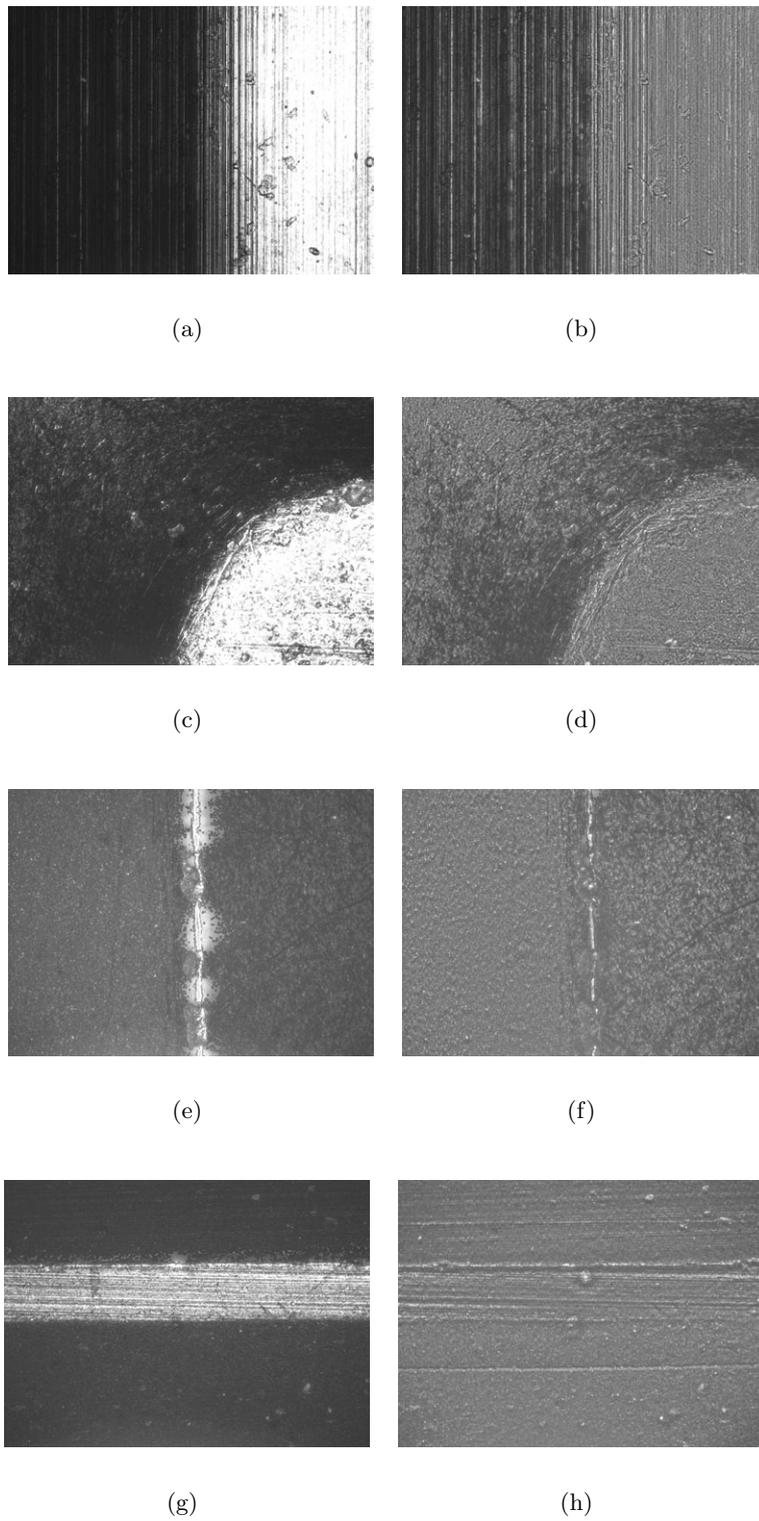


Figure 6.7: All-in-focus texture images. The focused images are compared between *STD* (a,c,e,g) and *ADAPT* for the specimen *CUT2 - side* (a,b), *CUT2 - corner* (c,d), *CUT2 - edge* (e,f) and *CUT3 - edge* (g,h).

6.2 Projected Texture

In this section the HF texture projection, as presented in Section 5.2 (*TEXTURE*), is evaluated. The purpose of this pattern projection is to create additional texture on low textured objects to increase the accuracy and robustness of the SFF reconstruction. Therefore we chose low textured objects to evaluate this method. The reference specimen are presented in Section 6.2.1. The calculated optimal HF pattern for texture enhancement is evaluated in Section 6.2.2. Finally, the reconstruction results of the reference specimen with projected high frequency pattern are compared to traditional SFF results in Section 6.2.3.

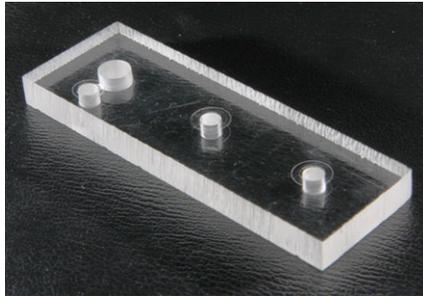
6.2.1 Reference Specimen

In this section the specimen for the reconstruction evaluation of smooth low textured objects are presented.

- Transparent polypropylene plane - *HF-PLANE*
- Planoconvex cylindric optical lens - *HF-LENS*
- Planoconvex diffusor lens - *HF-DIFFLENS*
- Mirror - *HF-MIRROR*
- Planar region of a polished coin - *HF-COIN*

The texture content on these specimen ranges from absolutely no measurable texture on the *HF-DIFFLENS*, *HF-LENS* and *HF-MIRROR* specimen to sufficient measurable texture on the *HF-COIN* specimen. In Figure 6.8 the different specimen for reconstruction are illustrated.

For *HF-PLANE*, *HF-MIRROR* and *HF-COIN* the reconstruction error is defined as the depth standard deviation to a rigidly registered plane and for *HF-LENS* we have built a groundtruth model, according to the lens specifications. For the polypropylene diffusor lens *HF-DIFFLENS* we have no exact groundtruth. Because the surface of this lens is smooth with a spherical shape, the groundtruth for this object is calculated as a filtered HF texture reconstruction result with a focus measure operator twice the size we use for evaluation (15×15).



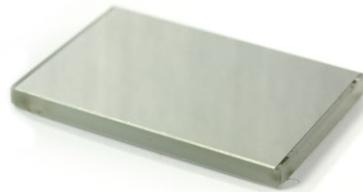
(a) Polypropylene Plane - *HF-PLANE*



(b) Planoconvex Lens - *HF-LENS*



(c) Polypropylene Diffuser Lens - *HF-DIFFLENS*



(d) Mirror - *HF-MIRROR*



(e) Polished Coin - *HF-COIN*

Figure 6.8: Reference specimen with low texture. The specimen (a) plane, (b) planoconvex cylindric lens (c) polypropylen diffusor lens, (d) mirror and (e) coin need additional texture projection for robust reconstruction.

6.2.2 Optimal Pattern Calibration

In this section the frequency of the projected texture pattern, used to enhance the natural object texture, is evaluated. As presented in Section 5.2, the calculated optimal frequency is $2\sigma_p < \frac{1}{2}t_x = \frac{1}{2}t_y < m_{x,y} + 2$, where σ_p is the point spread of one projected pixel and $m_{x,y}$ is the size of the focus measure operator. In this evaluation we want to prove that the calculated optimal frequency delivers also the optimal reconstruction results for real specimen.

$m_{x,y}$	Pattern Frequency $t_{x,y}/2$				
	3.4	6.8	10.2	13.4	20.4
3	41.1660	17.4190	7.9267	8.5767	12.8890
5	55.3710	17.7840	3.9098	6.2270	12.5340
7	56.4720	14.3830	1.6297	4.2421	9.9441
9	59.1490	10.3760	0.7194	3.2119	8.2591
11	57.7680	7.3985	0.4521	1.8536	7.6662

Table 6.4: Evaluation of optimal HF pattern frequency. Reconstruction errors of the *HF-LENS* object for different HF pattern frequencies $t_{x,y}/2$ reconstructed with different size of the focus measure operator $m_{x,y}$ in $[\mu m]$. The calculated optimal HF pattern frequencies are highlighted.

In Table 6.4 the reconstruction errors of the *HF-DIFFLENS* object are shown. The object is illuminated with HF pattern of different frequencies ($t_{x,y}/2 = \{3.4, 6.8, 10.2, 13.4, 20.4\}$) and reconstructed with different sizes of the focus measure operator ($m_{x,y} = \{3, 5, 7, 9, 11\}$). In Figure 6.9 the reconstruction results of the specimen with HF pattern illumination are illustrated.

6.2.3 Reconstruction with HF pattern projection

In this section the evaluation results of SFF reconstructions with projected HF texture pattern (*TEXTURE*) are presented. The focus measure operator has a size of 9×9 , whereas the frequency of the HF pattern is 3×3 projector pixels, which results in approximately 10.2×10.2 camera pixels.

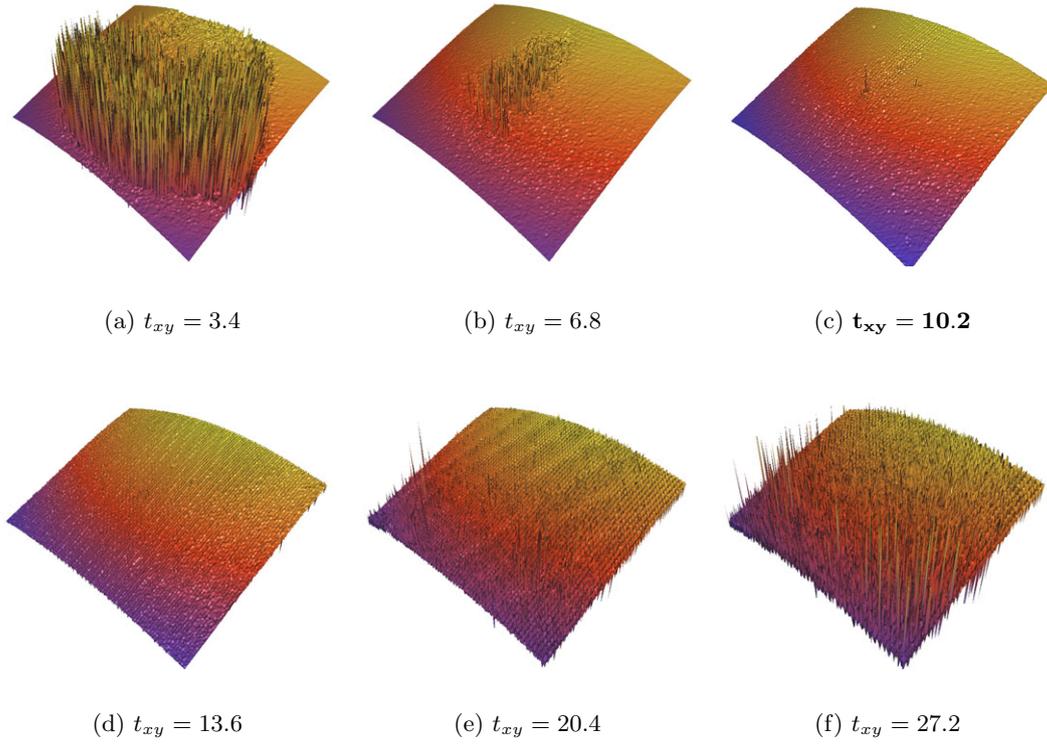


Figure 6.9: Reconstruction of the *HF-DIFFLENS* specimen with HF texture patterns of different frequencies. One projector pixel directly influences $\sim 3.4 \times 3.4$ camera pixels, $2\sigma_p \approx 8.4$ at a Tenengrad window size of 9×9 . The calculated optimal HF pattern frequency is highlighted.

	<i>HF-PLANE</i>	<i>HF-LENS</i>	<i>HF-DIFFLENS</i>	<i>HF-MIRROR</i>	<i>HF-COIN</i>
<i>STD</i>	18.9890	n.m.	n.m.	n.m.	6.5291
<i>HF Pattern</i>	5.7553	2.9322	1.7194	7.4574	2.4865

Table 6.5: Reconstruction errors of texture pattern projection. Reconstruction errors (standard deviation of depth errors) for low textured objects with homogeneous illumination (*STD*) compared to the illumination of a HF pattern (*TEXTURE*) in [μm]. (n.m. not measurable)

In Table 6.5 the reconstruction errors for all reference specimen are shown, whereas the reconstruction results of *HF-LENS*, *HF-DIFFLENS* and *HF-MIRROR* did not lead to a measurable reconstruction result with *STD*, because no texture could be detected. The reconstructed depth maps of the reference specimen are shown in Figure 6.10 and Figure 6.11.

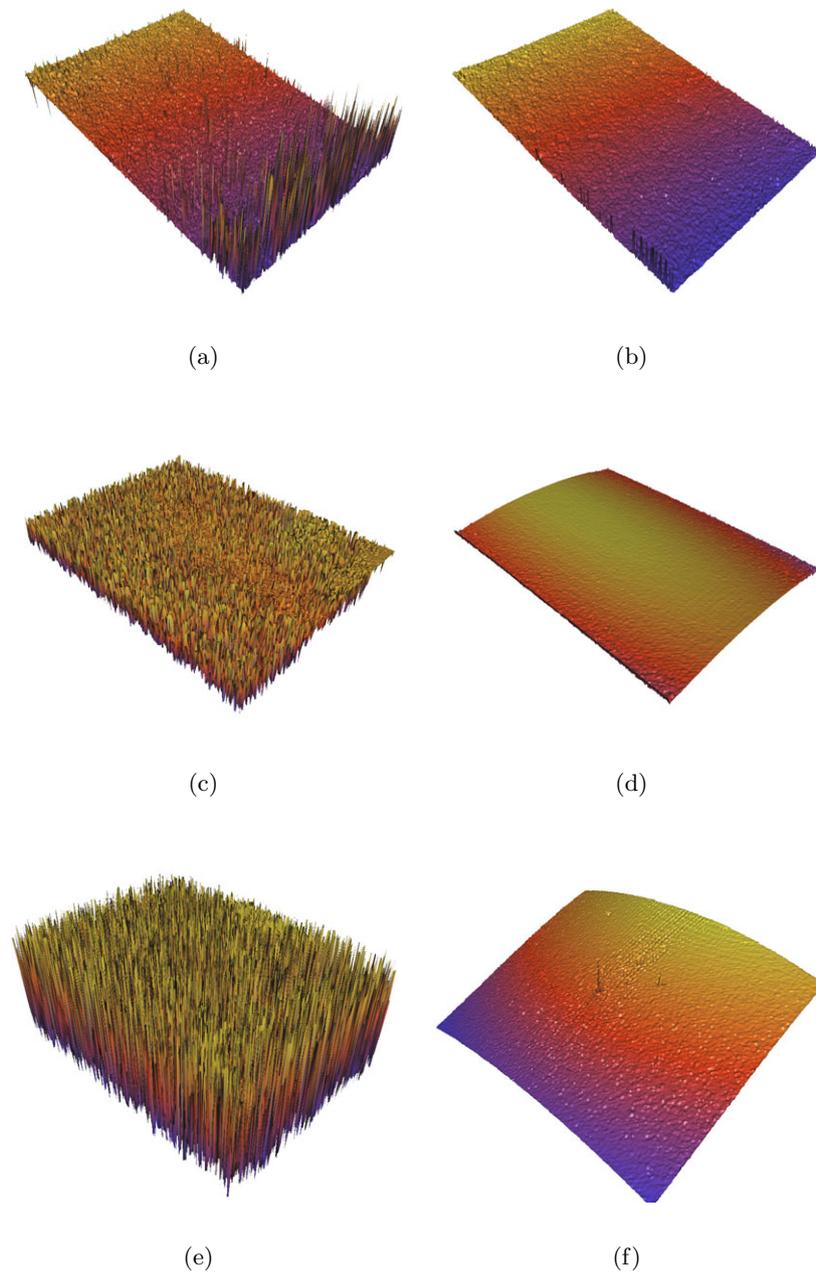


Figure 6.10: Reconstruction results of specimen *HF-PLANE* (a,b), *HF-LENS* (c,d) and *HF-DIFFLENS* (e,f). SFF results calculated by *STD* (a,c,e) and with *TEXTURE* (b,d,f).

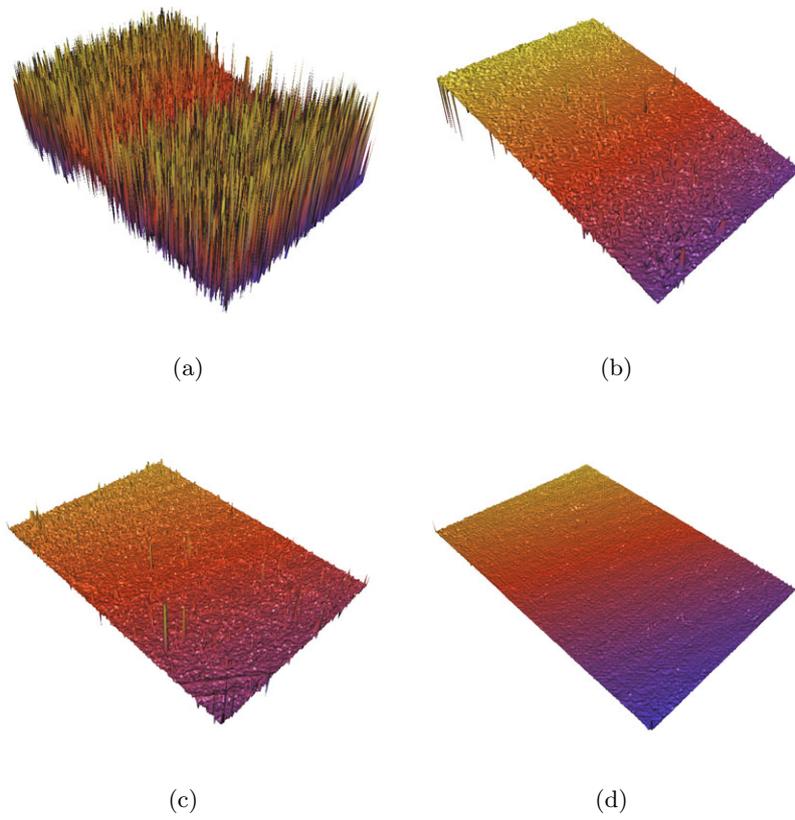


Figure 6.11: Reconstruction results of specimen *HF-MIRROR* (a,b) and *HF-COIN* (c,d). SFF results calculated by *STD* (a,c) and with *TEXTURE* (b,d).

6.3 Combination of HF Pattern with Adapted Illumination

In this section the results for the combination of illumination adaption (*ADAPT*) and HF texture pattern *TEXTURE* according to Section 5.3 are presented. The problem in this case is that we have no usable groundtruth data. This is caused by the operation area of this method. On the one side the reconstruction object has to be smooth enough for the usage of a HF pattern and on the other side it has to contain differences in the dynamic range for adaption. Therefore, both the use of planar objects and the comparison to HDR results are senseless for accurate quality measurement.

To define the quality of this method we evaluate the robustness of the reconstruction results. This is done by the calculation of an outlier percentage in our reconstruction, whereas an outlier is defined as a depth variation higher than $20\mu m$ between consecutive pixels. This outlier percentage is calculated for specimen with both low textured parts and parts with varying reflectance. To compare this illumination to other methods, the results are calculated for traditional SFF with homogeneous illumination (*STD*), HF pattern projection *TEXTURE*, predictive adaption *ADAPT*, predictive adaption with added high frequency pattern *HF+ADAPT* and predictive adaption with a HF pattern used as desired camera image *ADAPTED-HF*. Additionally, to evaluate the robustness of this method to previous reconstructions, the reconstruction of the reference objects from Section 6.1 is compared to the a reconstruction with standard adaption and the groundtruth.

The reference specimen to evaluate the robustness are illustrated in Section 6.3.1, whereas the evaluation results are presented in Section 6.3.2.

6.3.1 Reference Specimen

To test the robustness of this method, the evaluation objects must contain both regions with a high dynamic range and very smooth and low-textured regions. Therefore we selected different regions on a polished coin, namely *COIN1*, *COIN2* and *COIN3*, as illustrated in Figure 6.12. In these regions both, parts with low texture and with reflection differences occur which makes them ideal for the evaluation of this method.

6.3.2 Robustness Evaluation

In this section the combination of HF projection pattern and adapted illumination is evaluated. Therefore different specimen, as illustrated in Section 6.3.1 are reconstructed with the different illumination possibilities. In Table 6.6 the reconstruction error is measured

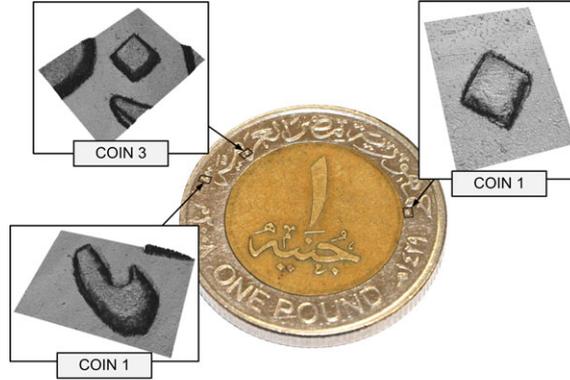


Figure 6.12: Reference specimen for combination of HF pattern with adaption. The regions on the polished coin have both low-textured regions and regions with a high dynamic range in the reflectance.

by the percentage of outliers in the scene.

Method	<i>COIN1</i> outlier [%]	<i>COIN2</i> outlier [%]	<i>COIN3</i> outlier [%]
<i>STD</i>	4.1525	5.3718	6.1728
<i>TEXTURE</i>	2.5003	3.8086	5.7613
<i>ADAPT</i>	3.0499	3.5301	3.7403
<i>HF+ADAPT</i>	1.8322	2.4307	3.2458
<i>ADAPTED-HF</i>	1.6366	2.6554	3.4522

Table 6.6: Outlier for reconstruction of HF pattern combined with adaption. The outlier percentage calculated for the reconstruction of regions on a polished coin for the different illumination methods. The best results are highlighted.

The reconstructed depth maps of the specimen are illustrated in Figure 6.13 (*COIN1*) and Figure 6.14 (*COIN2* and *COIN3*).

To compare the reconstruction of *HF+ADAPT* and *ADAPTED-HF* to the reconstruction results of Section 6.1.3, the reconstruction error is calculated for *CUT2 - side* and *CUT2 - corner* specimen, as listed in Table 6.7.

Method	CUT2 - side		CUT2 - corner	
	σ	#imgs	σ	#imgs
<i>STD</i>	13.100	78	37.960	58
<i>HDR3</i>	7.6840	234	4.0276	174
<i>ADAPT</i>	2.6046	78	1.9721	58
<i>HF+ADAPT</i>	2.2515	78	2.1825	58
<i>ADAPTED-HF</i>	2.4006	78	3.3317	58

Table 6.7: Reconstruction errors (standard deviation of depth errors) for the specimen *CUT2 - side* and *CUT2 - corner* of previous methods compared with the combination of adaption and HF pattern projection in [μm]. The best results are highlighted.

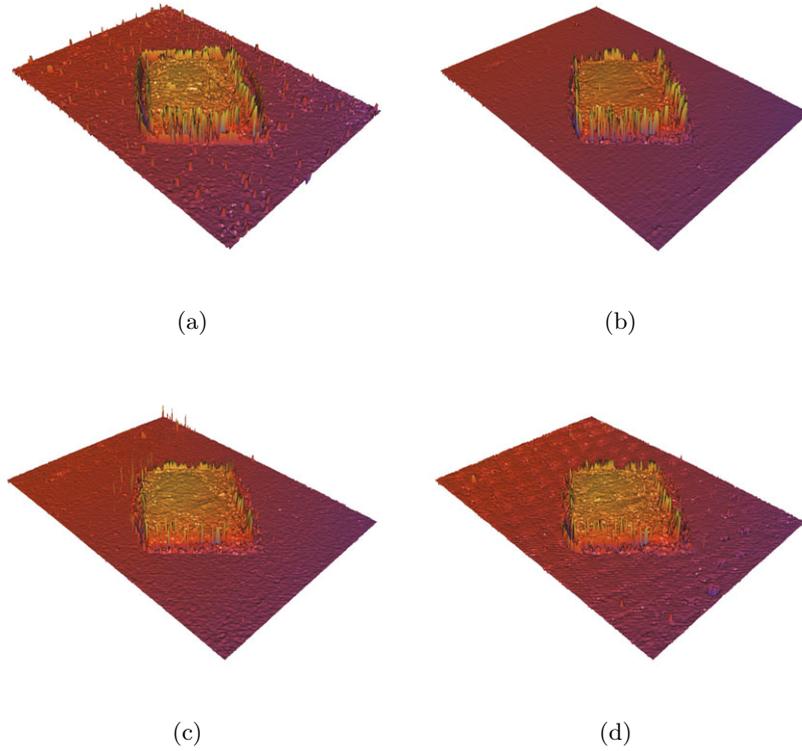


Figure 6.13: Reconstruction result for *COIN1* specimen with *STD* (a), *TEXTURE* (b), *HF+ADAPT* (c) and *ADAPTED-HF* (d).

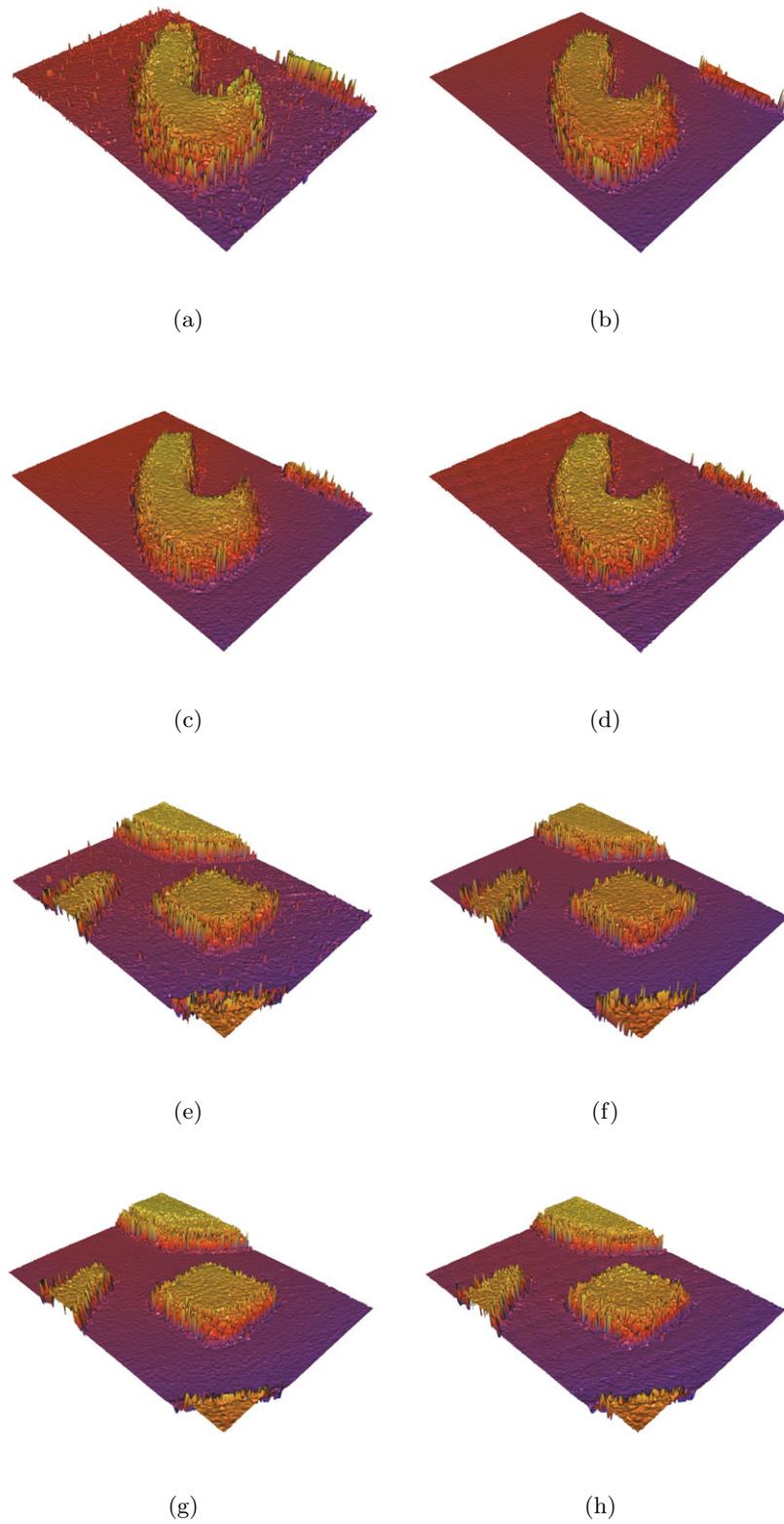


Figure 6.14: Reconstruction result for *COIN2*(a-d) and *COIN3*(e-h) specimen with *STD* (a,e), *TEXTURE* (b,f), *HF+ADAPT* (c,g) and *ADAPTED-HF* (d,h).

6.4 Integral Shape From Focus

In this section the evaluation of the integral shape from focus (ISFF) method is presented. This method calculates an IPSF for the reconstruction of a focused image out of the SFF image stack as presented in Section 5.4. After calculation of the focused image, the focus measure is calculated with a patch-wise correlation of the images from the image stack with the restored focused image. Because the equality value from this correlation is normalized between -1 and 1 , the focus calculation can be stopped at a local maximum above a predefined threshold.

For evaluation we use the IPSF calibrated with the TV-L1 approach on a textured plane (see Fig. 6.15) normal to the optical axis (*PLANE-NORM-GT*). The NCC is calculated in a correlation window of 7×7 .

In Section 6.4.2 the influence of noise on the IPSF focused image reconstruction is evaluated. The depth reconstruction using the ISFF method is compared to traditional SFF method (*STD*) in Section 6.4.3. The speed improvements of the thresholded focus calculation (*ISFF-FAST*) are evaluated in Section 6.4.4. The reference specimen for ISFF evaluation are presented in Section 6.4.1.

6.4.1 Reference Specimen

To evaluate the ISFF method on real objects, we chose regions on a defined measurement object, as illustrated in Figure 6.15. It consists of two planes with a defined height difference of $1\mu m$. The surface has enough measurable texture to get a depth map with the traditional SFF. To evaluate the ISFF method, three different regions are chosen on



Figure 6.15: Planar Specimen. Object with two defined planes in $1\mu m$ height difference.

this plane and are acquired where the object is placed in front of the acquisition system with different angles to the optical axis, namely *PLANE-NORM* (normal to the opti-

ical axis), *PLANE-SKEW5* (skew with rotation of ~ 5 degrees about the y axis) and *PLANE-SKEW10* (skew with rotation of ~ 10 degrees about the y axis).

To measure the influence of noise on the IPSF focused image convolution, we chose a synthetic object (*CHESS*). For this object an exact texture image and an exact groundtruth is available. The synthetic SFF image stack is generated by applying a Gaussian blur according to the level of focus in each image. Therewith a stack of 25 images is generated. In Figure 6.16 the synthetic specimen and sample stack images are shown.

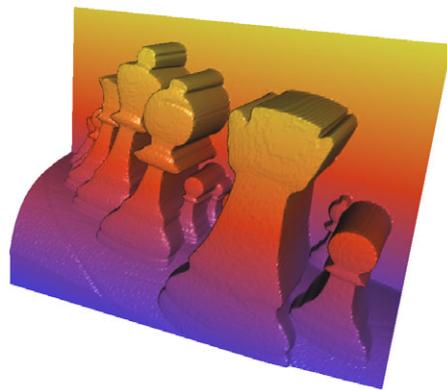
6.4.2 Noise Influence on the Focused Image Reconstruction

The reconstruction of a focused image by deconvolution of the IPSF with the integral image out of the image stack is noise sensitive. Therefore we evaluate the reconstruction of the focused image under the influence of defined noise. To get a defined noise on the acquired images we use a synthetic scene (*CHESS*), where the degree of noise can be specified exactly.

For evaluation, white noise with a standard deviation σ_w ranging from 0 to 0.02 is applied on the stack images. The integral image is calculated by the mean of the sum of the image stack. To calculate the restored focused image, the integrated image stack is convolved with the IPSF estimated from the synthetic groundtruth without noise and with the IPSF from the textured reference plane (*PLANE-NORM-GT*). In Figure 6.17 the error of the restored images with two different IPSF estimations is shown. The error is calculated as the difference in mean and standard deviation to the groundtruth texture image. It can be seen that the standard deviation error is very low when no noise was applied and steeply rises with increasing noise. The resulting restored focus images are illustrated in Figure 6.18.

6.4.3 ISFF Depth Reconstructions

In this section the SFF depth reconstruction calculated with the correlation based focus measure is evaluated. The depth map is reconstructed for the normal plane *PLANE-NORM*, the skew planes *PLANE-SKEW5* and *PLANE-SKEW10* specimen. The correlation focus measure is calculated with both a known focused image from the groundtruth (*ISFF-GT*) and restored focused images with the IPSF from *PLANE-NORM* (*ISFF*), whereas the results are compared with standard SFF reconstructions (*STD*). The reconstruction errors are measured as a standard deviation of depth errors to a rigidly registered



(a) Scene depth map



(b) Depth map with texture



(c) Synthetic stack, image 5



(d) Synthetic stack, image 12



(e) Synthetic stack, image 17



(f) Synthetic stack, image 21

Figure 6.16: Synthetic specimen. For robust evaluation of the influence of noise in the reconstruction of the focused image a reference scene is generated. The SFF stack for the synthetic specimen is calculated by applying Gaussian noise to the texture image according to the level of focus.

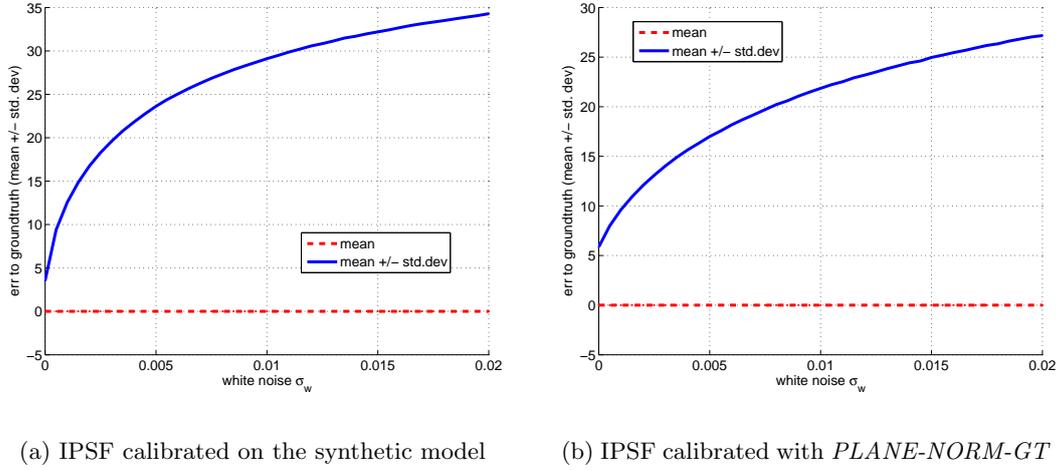


Figure 6.17: Influence of Noise on IPFS convolution. The difference between the all-in-focus texture and the restored focused texture under the influence of noise. The error is measured with the difference of mean and standard deviation between the images. In (a) the focused image is restored with an IPFS calibrated at the noise-free synthetic model (*CHESS*) and (b) is restored with the standard IPFS from *PLANE-NORM-GT*.

plane in [μm]. One reconstruction consists of approximately $1.4e6$ depth estimates. In Table 6.8 the reconstruction error results are listed and the reconstructed depth maps are illustrated in Figure 6.19. The corresponding focused image restorations by convolution through the IPFS are illustrated in Figure 6.20.

Specimen	<i>STD</i>	<i>ISFF-GT</i>	<i>ISFF</i>
	σ	σ	σ
<i>PLANE-NORM</i>	1.6577	1.8820	2.6573
<i>PLANE-SKEW5</i>	1.7505	3.9738	3.6144
<i>PLANE-SKEW10</i>	2.3648	3.9862	16.675

Table 6.8: ISFF reconstruction results. The reconstruction error of the specimen is calculated by the standard deviation σ of the difference to a rigidly registered reference plane in [μm].



(a) Groundtruth texture

(b) Restored $\sigma_w = 0$ (c) Restored $\sigma_w = 0.005$ (d) Restored $\sigma_w = 0.01$ (e) Restored $\sigma_w = 0.015$ (f) Restored $\sigma_w = 0.2$

Figure 6.18: Restored all-in-focus images. The texture is restored by convolution of the integral image with different noise levels in the image stack and the IPSF from *PLANE-NORM-GT*.

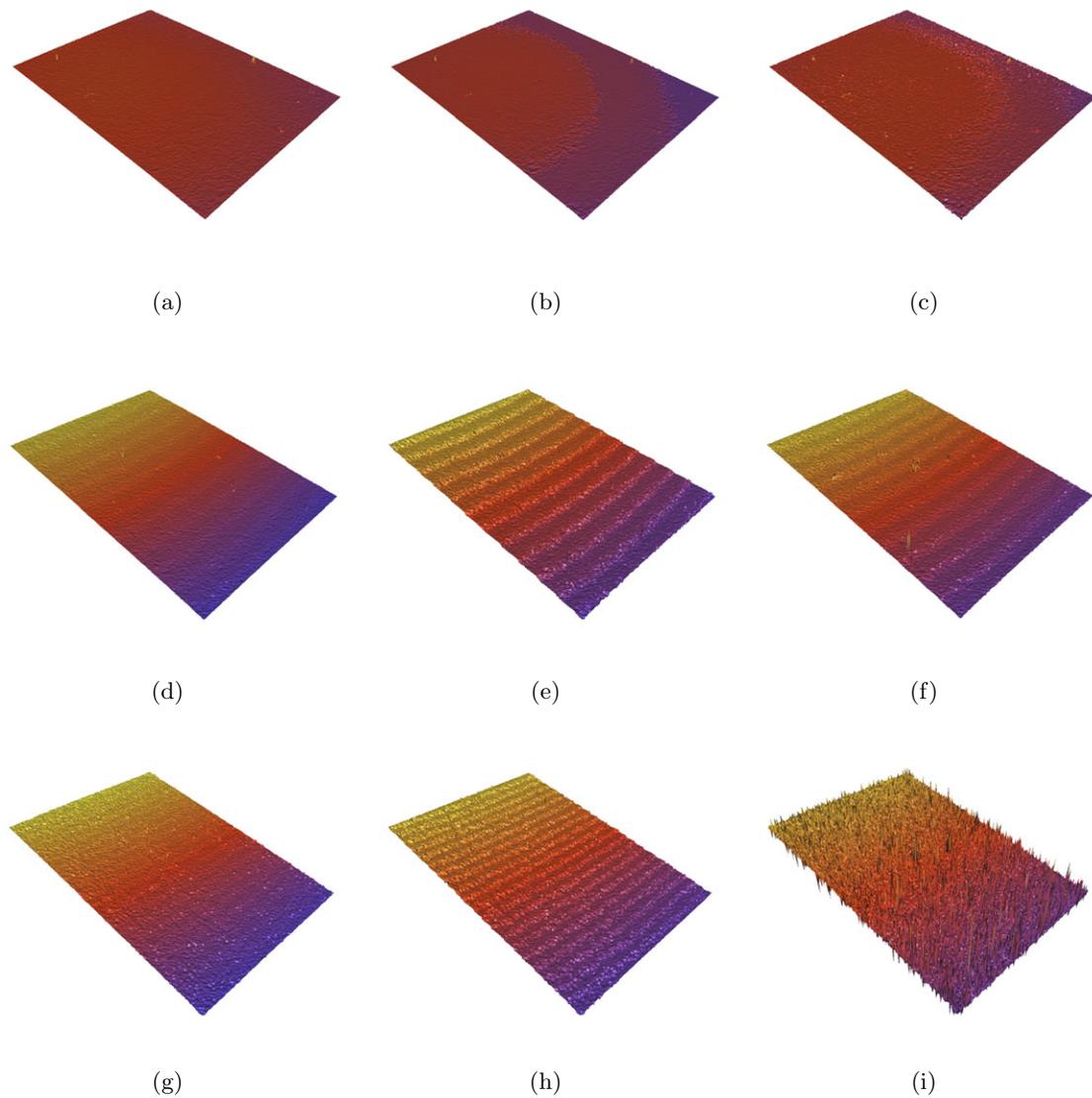


Figure 6.19: ISFF reconstruction results of specimen *PLANE-NORM* (a-c), *PLANE-SKEW5* (d-f) and *PLANE-SKEW10* (g-i). The depth maps are calculated with *STD* (a,d,g), correlation based focus measure with groundtruth texture as NCC template (*ISFF-GT*) (b,e,h) and correlation based focus measure restored all-in-focus image using the IPSF (*ISFF*) (c,f,i).

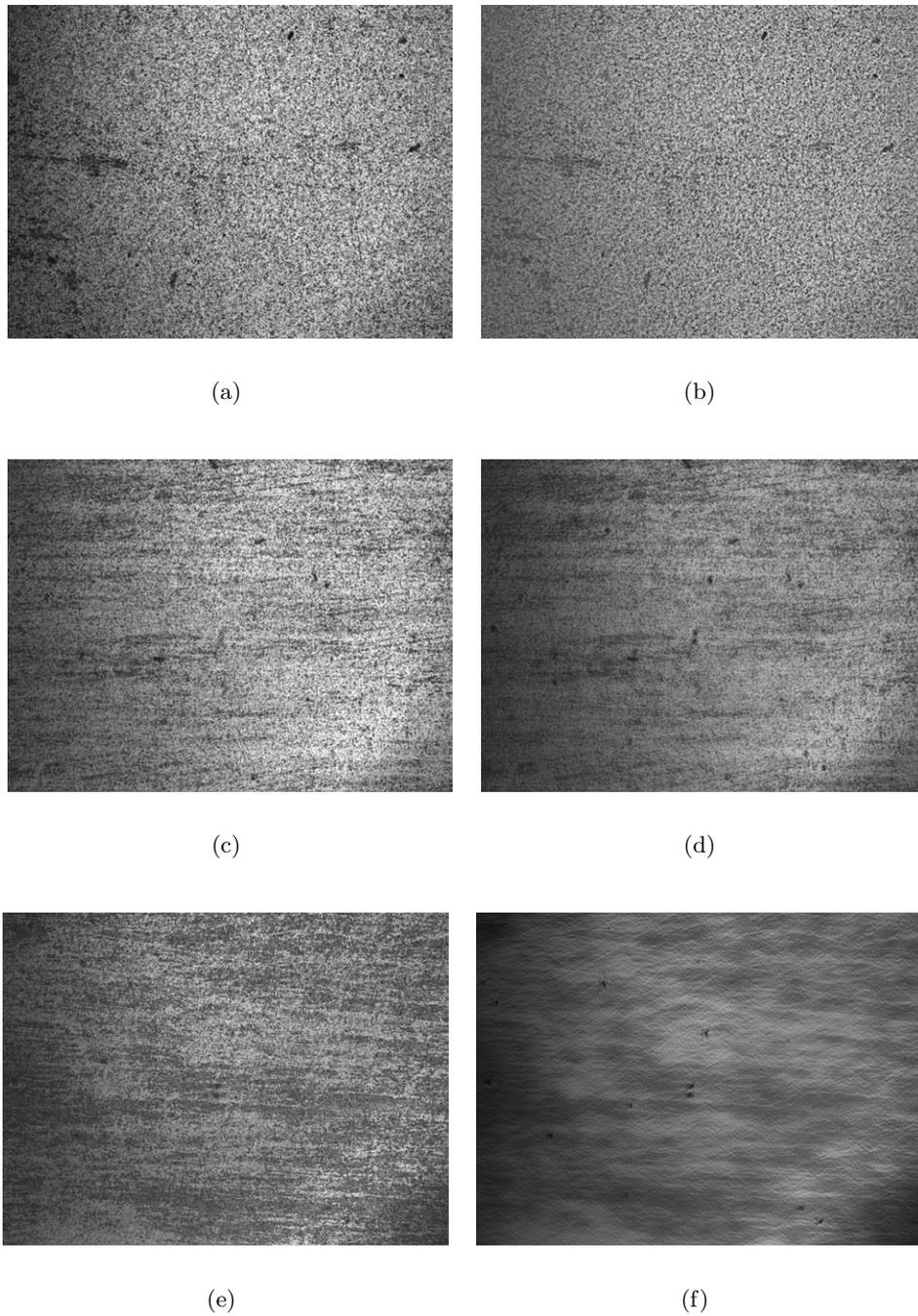


Figure 6.20: IPSF all-in-focus image recovery. The original texture (a,c,e) is compared with the restored texture using the IPSF (b,d,f) for the specimen *PLANE-NORM* (a,b), *PLANE-SKEW5* (c,d) and *PLANE-SKEW10* (e,f).

6.4.4 Speed Improvements of ISFF

The speed of the focus calculation for the ISFF can be increased by an additional constraint to stop the maximum search, when a local maximum in the focus vector is found above a predefined threshold (*ISFF-FAST*). In this section the speed improvements and the reconstruction results with different thresholds are evaluated on the *PLANE-SKEW5* specimen.

First, the reconstruction result of the ISFF with the speed improvements at different threshold values is compared to the reconstruction of the specimen without speed improvements. To evaluate if a focus maximum is found with *ISFF-FAST*, the percentage of wrong focus maxima is recorded. To evaluate the speed improvement, the relative time differences for each threshold level is compared to the depth estimation of the focus maximum estimation of traditional SFF. Additionally, the reconstruction error is calculated for the *ISFF-FAST* result calculated through the standard deviation of the reconstruction difference to a rigidly registered groundtruth plane. In Table 6.9 and Figure 6.21 the evaluation results are illustrated. To measure the relative time improvement the mean of the computation time for ten iterations is calculated. The reconstruction results are shown in Figure 6.22. The best reconstruction result is calculated at a threshold of 0.9 with 1.63 percent outliers compared to *ISFF* and a standard deviation of depth errors of $8.5687\mu\text{m}$ with a time improvement of 31.9 percent.

<i>ISFF-FAST</i>			
Threshold	wrong maxima [%]	time improvement [%]	σ [μm]
<i>0.750</i>	25.7359	43.927	27.7370
<i>0.775</i>	18.9226	43.280	23.3090
<i>0.800</i>	13.2364	42.575	18.8750
<i>0.825</i>	8.6745	38.517	14.8740
<i>0.850</i>	5.2509	35.869	11.4490
<i>0.875</i>	2.8909	35.692	8.8679
<i>0.900</i>	1.5799	31.906	8.5687
<i>0.925</i>	1.6264	29.881	13.4190
<i>0.950</i>	6.0163	25.041	28.8700
<i>0.975</i>	38.5443	18.435	63.9810

Table 6.9: Evaluation of speed improvements in ISFF. To evaluate the reconstruction results of *ISFF-FAST*, the percentage of wrong focus maxima compared to *ISFF* reconstruction is recorded. Furthermore the relative computation time differences are included. The best results are highlighted.

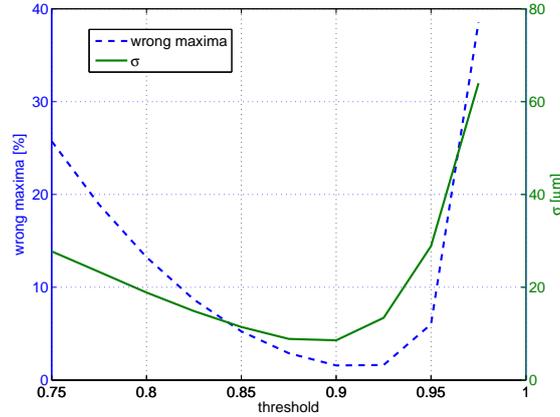


Figure 6.21: Error performance of *ISFF-FAST*. The wrong maxima percentage of *ISFF-FAST* compared to *ISFF* and the standard deviation of depth errors (σ) to a rigidly registered plane is calculated at different thresholds.

6.5 Summary

In this chapter the different methods to increase speed, robustness and accuracy of the SFF reconstruction were evaluated. To test the performance of adaptive illumination the SFF reconstruction results of homogeneous illumination, HDR image acquisition and the adaptive illumination was compared to groundtruth data. To evaluate the texture projection method the SFF reconstructions with and without additional texture of different textured specimen were compared to a known groundtruth. This delivered exact results of the performance of this method compared to traditional SFF. The combination of adaptive illumination and texture projection was tested at different specimen with both local differences in the object reflection and low textured regions. Since groundtruth was not available for this specimen, we measured the number of outliers in the reconstruction result, whereas an outlier was defined as a significant change in depth between consecutive pixels of the reconstruction result. To evaluate the ISFF method we first tested the influence of noise on the all-in-focus texture image creation on a synthetic model. The quality of the ISFF reconstruction was tested with a planar specimen captured at different angles to the optical axis. The error was calculated with the difference to planarity and was compared to reconstruction errors of traditional SFF. Furthermore we evaluated the speed improvements and the error influence at different thresholds.

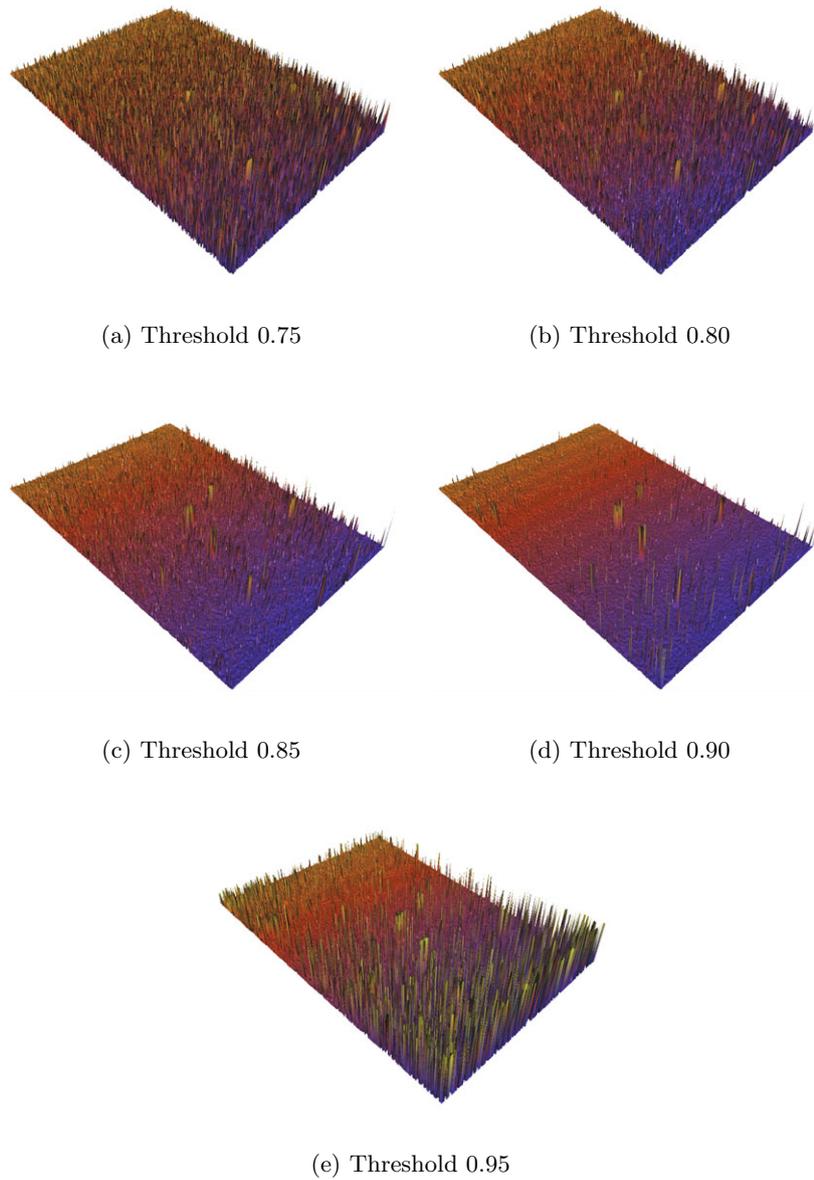


Figure 6.22: Reconstruction results of *ISFF-FAST*. The depth map was reconstructed with the correlation based focus measurement where the maximum search through the correlation stack stops if a local maximum is found beyond a predefined threshold.

Chapter 7

Discussion and Outlook

In this work a SFF acquisition system with a microscopic projector-camera system was realized. The optical system was constructed to use the projector as coaxial and locally adaptive light source to illuminate the specimen. Traditional microscopic SFF reconstruction systems with a homogeneous light source have problems to detect the degree of focus when the dynamics in the specimen reflectance are higher than the radiometric sensor resolution or when the specimen lacks in detectable texture, since it measures the focus through pixel intensity variations. The projector-camera system was designed to overcome these problems. On the one side, with the assistance of a projector as illumination source, it is possible to locally adjust the illumination intensity according to the local reflectance. Therewith, the high dynamic differences can be compensated. On the other side, when reconstructing low textured objects, the projector can be used to project additional texture on the object, which has the same characteristics as the natural texture.

A different problem in SFF focus reconstruction is its computation time. In the algorithm, multiple images have to be acquired with varying distances from the camera to the specimen. To get a dense reconstruction, each point on the specimen has to be in focus on at least one image. A focus measure operator calculates the level of focus for every pixel in every acquired image. The depth is estimated by searching through the focus levels of every pixel in every image. In this work we invented a new focus measure which enables a reduction of the computational cost for depth reconstruction. In this novel method an all-in-focus texture image was restored out of the image stack. Every pixel in this focused image texture was compared to the corresponding pixels in the image stack through NCC. Because the NCC correspondence values are normalized between -1 and 1 the search for a focus maximum can be stopped, when a local maximum beyond a predefined threshold

is found.

In Chapter 6 these methods were analysed concerning their performance, speed and accuracy through different experiments and evaluations. In this section, the results of this analysis and the advantages and disadvantages of the different methods presented in Chapter 5 are discussed.

To compensate a variety of object reflections in the adaption process, we calculated a compensation pattern for the projector in a prediction-correction manner. This reduces the number of required acquisitions to a single compensation. These results were compared to the iterative adaption, where the projection pattern was iteratively calculated in every depth step separately, where four camera acquisitions were necessary in every step. In the experiments it was shown, that the acquisition of an entire stack requires at least three additional depth steps for initialisation of the predictive adaption, which is negligible on a typical stack of sixty images.

The reconstruction results with an adaptive scene illumination were compared to the classical, single image approach, and with reconstruction from HDR imaging. According to our experiments, we obtained a robust reconstruction for all specimen while the standard SFF approach inevitably fails. When comparing the reconstruction errors, HDR imaging requires at least three measurements per image to reach the accuracy level of adapted illumination. Furthermore, for more difficult scenes with higher dynamic ranges even reconstructions with five acquisitions did not reach our level. Therewith we can say that the reconstruction result of our method really competed with methods like HDR imaging while the measurement time of HDR was at least three times higher. Of course, our approach was limited by the dynamic ranges of camera and projector. For a scene dynamic which exceeds the product of both ranges, a full compensation of the scene reflectance can not be guaranteed with a single camera exposure time.

We have also proposed the projection of texture on objects to increase the reconstruction accuracy. With the assistance of a projected texture the accuracy and the robustness was increased for low textured objects compared to standard illumination. Even objects without any measurable texture could be geometrically reconstructed. Furthermore, on objects with natural texture, the additional texture projection increased the robustness and accuracy of the SFF algorithm. The main disadvantage of this method was the strict dynamic range which was created with this method. Because the texture pattern only projects projector values of 0 and 255 it was very vulnerable to differences in the object reflectance. At high reflectance and a projected value of 255, the pixel region easily became

over exposed.

Therefore we invented a method which combines the properties of these methods. It merged the creation of texture through projection with the radiometric compensation of the scene reflectance. This enabled the projector-camera system to reconstruct low-textured objects with varying reflection properties. Although, this method did not have the same performance as the single methods in their field of application, it was generally applicable on different objects without any previous knowledge of object texture and reflectance.

For acceleration of the depth map reconstruction a new focus measurement was presented in this work. The focus was calculated through a gray value comparison of the acquired SFF images and an all-in-focus image, which was calculated through deconvolution of the integrated image stack with a previously calibrated IPSF. The experiments had shown that the reconstruction time for this method could be reduced by more than 30 percent compared to the traditional SFF algorithm with a degradation in accuracy of 1.58 percent. One major problem of the ISFF method was that the convolution of the all-in-focus image was very noise sensitive. On just a slight amount of noise, the all-in-focus image could not be recovered accurate enough to calculate an acceptable depth map. In our experiments we found out, that especially on high defocus of the surface points in the images stack, the reconstruction of the all-in-focus image gets inaccurate. Another problem in the IPSF calibration were the magnification shifts of surface points during the defocusing. Because, when the distance from the object to the specimen varied, also the magnification changed. Although, this led to a change of a surface point in the camera image of a few pixels, Nagahara et al. [40] mentioned that this effect is imperceptible for image reconstruction.

For the texture projection, future work will be to increase the projector to camera pixel ratio to increase the projector resolution on the acquired specimen. This will lead do a smaller PSF of one projector pixel. Therewith, also the texture can be projected with a higher resolution, where the focus measure operators detect more intensity variations, which increases the sensitivity and the spatial resolution of the focus measurement.

For the ISFF method future work will be to include a robust noise estimation in the IPSF convolution. With improvement toward noise sensitivity, the ISFF will deliver better all-in-focus images, which lead to better correlation results with the NCC. Therewith, three dimensional (3D) reconstructions results with the same performance as traditional SFF are possible in less computation time. Furthermore, if a perfect all-in-focus texture

reconstruction is possible, a correlation can be estimated with more simple approaches e.g. the sum of absolute differences (SAD) algorithm. It only calculates the absolute pixel differences instead of using the NCC. With this very fast algorithm the computation time of ISFF can be further decreased.

Appendix A

Acronyms and Symbols

List of Acronyms

3D	three dimensional
AFM	atomic force microscope
CCD	charge coupled device
<i>crf</i>	camera response function
DLP	direct light processing
DOF	depth of field
DP	dynamic programming
FFT	fast Fourier transform
FIS	focused image surface
HDR	high dynamic range
HF	high frequency
IPSF	integral point spread function
ISFF	integral shape from focus
LCD	liquid crystal display
LCoS	liquid crystal on silicon
LTM	light transport matrix
LUT	look-up table
ML	modified Laplacian
MP	megapixel
NCC	normalized cross correlation
OTF	optical transfer function

PDF	probability density function
<i>prf</i>	projector response function
PSF	point spread function
RC	radiometric compensation
SEM	scanning electron microscope
SLM	spatial light modulator
SFF	shape from focus
SFD	shape from defocus
SFSh	shape from shading
SFS	shape from stereo
SML	sum of modified Laplacian
SAD	sum of absolute differences
WGN	white Gaussian noise

Bibliography

- [1] Ahmad, M. and Choi, T.-S. (2005). A heuristic approach for finding best focused shape. *IEEE Transactions on Circuits and Systems for Video Technology*, 15(4):566–574.
- [2] Ashdown, M., Okabe, T., Sato, I., and Sato, Y. (2006). Robust content-dependent photometric projector compensation. In *IEEE International Workshop on Projector-Camera Systems (CVPR)*, page 6, New York City, NY.
- [3] Asif, M. and Choi, T.-S. (2001). Shape from focus using multilayer feedforward neural networks. *IEEE Transactions on Image Processing*, 10(11):1670–1675.
- [4] Banyal, R. and Raghavendra Prasad, B. (2010). Nonlinear response studies and corrections for a liquid crystal spatial light modulator. *Pramana Journal of Physics*, 74:961–971.
- [5] Beil, W. and Carlsen, I. C. (1991). Surface reconstruction from stereoscopy and "shape from shading" in sem images. *Machine Vision and Applications*, 4:271–285.
- [6] Besl, P. J. and McKay, H. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256.
- [7] Bimber, O., Emmerling, A., and Klemmer, T. (2005). Embedded entertainment with smart projectors. In *Proceedings 32th Conference on Computer Graphics and Interactive Techniques*, volume 38, pages 48–55.
- [8] Brown, M., Majumder, A., and Yang, R. (2005). Camera-based calibration techniques for seamless multiprojector displays. *IEEE Transactions on Visualization and Computer Graphics*, 11(2):193–206.
- [9] Chan, T. and Chiu-Kwong, W. (1998). Total variation blind deconvolution. *IEEE Transactions on Image Processing*, 7(3):370–375.
- [10] Collett, E. (2005). *Field Guide to Polarization*, volume FG05. SPIE Publicato.
- [11] Cook, R. L. and Torrance, K. E. (1981). A reflectance model for computer graphics. In *Proceedings 8th Conference on Computer Graphics and Interactive Techniques*, pages 307–316, New York City, NY.
- [12] Danuser, G. (1999). Photogrammetric calibration of a stereo light microscope. *Journal of Microscopy*, 193(1):62–83.

- [13] Darrell, T. and Wohn, K. (1988). Pyramid based depth from focus. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 504–509, Ann Arbor, MI.
- [14] Debevec, P. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. pages 369–378, New York City, NY.
- [15] Dixon, A. J., Doe, N., and Pang, T.-M. (1987). *Industrial applications of confocal scanning optical microscopy*, volume 809. Society for Photo-Optical Instrumentation Engineers.
- [16] Fujii, K., Grossberg, M. D., and Nayar, S. K. (2005). A projector-camera system with real-time photometric adaptation for dynamic environments. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 814–821.
- [17] Greenberg, G. and Boyde, A. (1993). Novel method for stereo imaging in light microscopy at high magnifications. *NeuroImage*, 1(2):121–128.
- [18] Grossberg, M. D., Peri, H., Nayar, S. K., and Belhumeur, P. N. (2004). Making one object look like another: Controlling appearance using a projector-camera system. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 452–459, Washington, DC.
- [19] Grossmann, P. (1987). Depth from focus. *Pattern Recognition Letters*, 5(1):63–69.
- [20] Grundhoefer, A. and Bimber, O. (2008). Real-time adaptive radiometric compensation. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):97–108.
- [21] Gu, J. W., Nayar, S. K., Grinspun, E., Belhumeur, P. N., and Ramamoorthi, R. (2008). Compressive structured light for recovering inhomogeneous participating media. In *Proceedings 10th European Conference on Computer Vision*, pages 845–858, Copenhagen, DK.
- [22] Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521540518.
- [23] Helmlí, F. S. and Scherer, S. (2001). Adaptive shape from focus with an error estimation in light microscopy. *Proceedings IEEE of 2nd International Symposium on Image and Signal Processing and Analysis*, pages 188–193.

- [24] Holoeys Photonics, A. (2009). Spatial light modulators.
- [25] Hopkins, H. H. (1955). The frequency response of a defocused optical system. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 231(1184):91–103.
- [26] Horn, B. (1968). Focusing. Technical Report 160, Massachusetts Institute of Technology - Artificial Intelligence Laboratory.
- [27] Humphreys, G. and Hanrahan, P. (1999). A distributed graphics system for large tiled displays. In *Proceedings IEEE Conference on Visualization*, pages 215–223.
- [28] Jordan, H.-J., Wegner, M., and Tiziani, H. (1998). Highly accurate non-contact characterization of engineering surfaces using confocal microscopy. *Measurement Science and Technology*, 9(7):1142–1151.
- [29] Kim, N., Bovik, A., and Aggarwal, S. (1990). Shape description of biological objects via stereo light microscopy. *IEEE Transactions on Systems, Man and Cybernetics*, 20(2):475–489.
- [30] Koninckx, T. P., Peers, P., Dutre, P., and Van Gool, L. (2005). Scene-adapted structured light. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 611–618, San Diego, CA.
- [31] Koninckx, T. P. and Van Gool, L. (2006). Real-time range acquisition by adaptive structured light. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):432–445.
- [32] Lacey, A. J., Thacker, N. A., Crossley, S., and Yates, R. B. (1998). A multi-stage approach to the dense estimation of disparity from stereo sem images. *Image and Vision Computing*, 16(5):373–383.
- [33] Lewis, J. P. (1995). Fast normalized cross-correlation. In *Vision Interface*, pages 120–123. Canadian Image Processing and Pattern Recognition Society.
- [34] Li, J. and Agathoklis, P. (1995). Measurement and analysis of defocused point spread functions and optical transfer functions of a microscope. *Proceedings IEEE Conference on Communications, Computers and Signal Processing*, pages 407–410.
- [35] Li, K. and Chen, Y. (1999). Optical blending for multi-projector display wall systems. In *Proceedings of IEEE Lasers and Electro-Optics Society*, 1:281–282.

- [36] Malik, A. S. and Choi, T. S. (2008a). Depth estimation by finding best focused points using line fitting. In *Proceedings IEEE International Conference on Image Processing*, pages 120–127.
- [37] Malik, A. S. and Choi, T. S. (2008b). A novel algorithm for estimation of depth map using image focus for 3D shape recovery in the presence of noise. *Pattern Recognition*, 41(7):2200–2225.
- [38] Mann, s. and Picard, R. W. (1995). Being 'undigital' with digital cameras: Extending dynamic range by combining different exposed pictures. In *Proceedings 48th Annual Conference on Imaging Science and Technology*, pages 442–448, Washington, DC.
- [39] Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441.
- [40] Nagahara, H., Kuthirummal, S., Changyin, Z., and Nayar, S. (2011). Flexible depth of field photography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(1):58–71.
- [41] Nayar, S. K. (1992). Shape from focus system. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 302–308, Champaign, IL.
- [42] Nayar, S. K., Harish, P., Grossberg, M. D., and Belhumeur, P. N. (2003a). A projection system with radiometric compensation for screen imperfections. In *IEEE International Workshop on Projector- Camera Systems (ICCV)*, Nice, FR.
- [43] Nayar, S. K. and Mitsunaga, T. (2000). High dynamic range imaging: Spatially varying pixel exposures. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 472–479.
- [44] Nayar, S. K. and Nakagawa, Y. (1990). Shape from focus: An effective approach for rough surfaces. In *Proceedings IEEE International Conference on Robotics and Automation*, volume 1, pages 218–225.
- [45] Nayar, S. K. and Nakagawa, Y. (1994). Shape from focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):824–831.
- [46] Nayar, S. K., Peri, H., Grossberg, M. D., and Belhumeur, P. N. (2003b). A projector with automatic radiometric screen compensation. Technical report, Department of Computer Science, Columbia University, New York.

- [47] Neil, M. A. A., Juskaitis, R., and Wilson, T. (1997). Method of obtaining optical sectioning by using structured light in a conventional microscope. *Optical Letters*, 22(24):1905–1907.
- [48] Noguchi, M. and Nayar, S. K. (1994). Microscopic shape from focus using active illumination. *Proceedings of 12th International Conference on Pattern Recognition*, pages 147–152.
- [49] Oren, M. and Nayar, S. K. (1994). Generalization of lambert’s reflectance model. *Proceedings 21st Conference on Computer Graphics and Interactive Techniques*, pages 239–246.
- [50] Paluszynski, J. and Slówko, W. (2005). Surface reconstruction with the photometric method in sem. *Vacuum*, 78(2-4):533–537. Proceedings of the fifth International Conference on Ion Implantation and other Applications of Ions and Electrons (ION 2004).
- [51] Pentland, A. (1987). A new sense for depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523–531.
- [52] Raskar, R., Welch, G., Cutts, M., Lake, A., Stesin, L., and Fuchs, H. (1998). The office of the future: a unified approach to image-based modeling and spatially immersive displays. Proceedings 25th Conference on Computer Graphics and Interactive Techniques, pages 179–188, New York, NY, USA.
- [53] Rosenfeld, A. and Kak, A. C. (1982). *Digital Picture Processing*. Academic Press, Inc., Orlando, FL, USA.
- [54] Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259 – 268.
- [55] Schlag, J., Sanderson, A. C., P., N. C., and Wimberly, F. C. (1983). Implementation of automatic focusing algorithms for a computer vision system with camera control. Technical report, Carnegie-Mellon University - Department of Electrical Engineering.
- [56] Seitz, S. M., Matsushita, Y., and Kutulakos, K. N. (2005). A theory of inverse light transport. 2:1440–1447.
- [57] Sen, P., Chen, B., Garg, G., Marschner, S. R., Horowitz, M., Levoy, M., and Lensch, H. P. A. (2005). Dual photography. In *Proceedings 32th Conference on Computer Graphics and Interactive Techniques*, pages 745–755, New York, NY, USA.

- [58] Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. Technical report, Pittsburgh, PA, USA.
- [59] Smith, W. J. (2000). *Modern optical engineering - The design of optical systems*. McGraw-Hill.
- [60] Stokseth, P. (1969). Properties of a defocused optical system. *Journal of the Optical Society of America*, 59(10):1314–1321.
- [61] Subbarao, M. (1987). Direct recovery of depth-map. In *Proceedings of the IEEE Computer Society Workshop on Computer Vision*, pages 58–65.
- [62] Subbarao, M. (1989). Efficient depth recovery through inverse optics. In *Machine Vision for Inspection and Measurement*, pages 101–126.
- [63] Subbarao, M. and Choi, T. (1995). Accurate recovery of three-dimensional shape from image focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(3):1769–1786.
- [64] Sungdo-Cha, Lin, P. C., Zhu, L., Botvinick, E. L., Sun, P.-C., and Fainman, Y. (1999). 3d profilometry using a dynamically configurable confocal microscope. In *Three-Dimensional Image Capture and Applications'99*, pages 246–253.
- [65] Tenenbaum, J. M. (1970). *Accommodation in Computer Vision*. PhD thesis, Stanford University CA, Department of Computer Science.
- [66] Wang, D., Sato, I., Okabe, T., and Sato, Y. (2005). Radiometric compensation in a projector-camera system based properties of human vision system. *IEEE International Workshop on Projector- Camera Systems (CVPR)*.
- [67] Wetzstein, G. (2007). Radiometric compensation of global illumination effects with projector-camera systems. Master's thesis, Bauhaus-Universität Weimar.
- [68] Wetzstein, G. and Bimber, O. (2006). Radiometric compensation of global illumination effects with projector-camera systems. In *Proceedings 33th Conference on Computer Graphics and Interactive Techniques*, New York City, NY.
- [69] Wetzstein, G. and Bimber, O. (2007). Radiometric compensation through inverse light transport. In *Proceedings IEEE Conference on Computer Graphics and Applications*, pages 391–399.

-
- [70] Xiong, F. S. and Shafer, S. A. (1993). Depth from focusing and defocusing. *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, pages 68–73.
- [71] Yap, P. T. and Raveendran, P. (2004). Image focus measure based on chebyshev moments. *Proceedings IEEE Vision, Image and Signal Processing*, 151(2):128–136.
- [72] Yun, J. and Choi, T. S. (1999). Accurate 3-d shape recovery using curved window focus measure. *Proceedings IEEE International Conference on Image Processing*, 3:910–914.