

INAUGURAL - DISSERTATION  
zur  
Erlangung der Doktorwürde  
der  
Naturwissenschaftlich-Mathematischen  
Gesamtfakultät  
der  
Ruprecht-Karls-Universität  
Heidelberg

vorgelegt von  
Dipl.-Phys. Tobias Dierig  
aus Heidelberg

Tag der mündl. Prüfung: 03.07.2002



# Gewinnung von Tiefenkarten aus Fokusserien

Gutachter: Prof. Dr. Bernd Jähne  
Prof. Dr. Josef Bille



## **Zusammenfassung**

In der vorliegenden Arbeit werden Verfahren zur Tiefenschätzung aus Fokussereien einer genauen Analyse unterzogen. Mögliche Fehlerquellen werden ausgehend vom optischen Abbildungsprozess bis hin zu den digitalen Bildverarbeitungsalgorithmen eingehend untersucht. Anhand der zu Grunde liegenden physikalischen Prinzipien werden die Fokus-Verfahren mit anderen optischen 3D-Meßtechniken verglichen und entsprechend eingeordnet. Allen bisherigen Verfahren zur Tiefenschätzung aus Fokussereien ist gemein, daß der optische Aufbau aus einer speziellen skalierungsfreien, der sogenannten telezentrischen Optik bestehen muß. Andernfalls sind die sonst entstehenden Skalierungseffekte in einem zusätzlichen, fehlerbehafteten Schritt zu korrigieren. Ein im Rahmen dieser Arbeit entwickeltes neuartiges Verfahren zur simultanen Tiefen- und Bewegungsschätzung räumt diese gravierenden Limitierungen erfolgreich aus. Dazu wird die Fokusserei als Diffusionsprozess mit einer linearen, partiellen Differentialgleichung beschrieben. Die diesen Prozess beschreibenden Parameter werden mittels einer Total Least Squares Methode direkt aus der Fokusserei geschätzt. Die Tiefeninformation der Szene wird aus der gemessenen Diffusionskonstante errechnet. Abschließend werden alle vorgestellten Verfahren der Tiefenbestimmung an synthetischen und realen Fokussereien erprobt.

## **Abstract**

In this thesis different techniques for estimating depth from focal series are quantitatively analyzed. Possible sources of error are scrutinized from the optical imaging process to the digital image processing algorithms. The focus techniques are compared to other optical 3D measurement techniques and classified according to the underlying physical principles. In common to all existing methods of computing depth from focal series is their reliance on a special, non-scaling optical set-up, also termed telecentric. Otherwise inevitable effects due to the scaling have to be corrected for in an additional, erroneous process. A novel technique presented in this thesis for simultaneous estimation of depth and motion successfully eliminates this serious limitation. In this method the focal series is modelled by a diffusion process, described by a linear partial differential equation. The parameters characterizing this process are established by a total least squares estimator directly from the focal series. The information of depth can then be deducted from the measured constant of diffusion. This thesis concludes with an accuracy examination of all presented techniques both on synthetic and real world data.



# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
1.1	Aufbau der Arbeit . . . . .	3
<b>I</b>	<b>Grundlagen</b>	<b>5</b>
<b>2</b>	<b>Optik</b>	<b>7</b>
2.1	Geometrische Optik . . . . .	7
2.1.1	Lochkamera . . . . .	7
2.1.2	Kamera mit endlicher Blende . . . . .	8
2.1.3	Schärfentiefe . . . . .	11
2.1.4	3D-PSF . . . . .	12
2.2	PSF in der Wellenoptik . . . . .	16
2.3	Telezentrische Optik . . . . .	18
<b>3</b>	<b>3D-Meßtechniken</b>	<b>19</b>
3.1	Triangulations-Verfahren . . . . .	19
3.1.1	Stereo . . . . .	19
3.1.2	Depth from Motion . . . . .	22
3.1.3	Depth from Focus . . . . .	23
3.2	Laufzeitverfahren . . . . .	28

3.2.1	Laufzeitverfahren mit moduliertem Licht . . . . .	29
3.2.2	Interferometrie . . . . .	29
<b>4</b>	<b>Depth from Focus</b>	<b>31</b>
4.1	Ein-Bild-Verfahren . . . . .	31
4.2	Zwei-Bild-Verfahren / Depth-from-Defocus . . . . .	32
4.3	Mehr-Bild-Verfahren / Depth-from-Focus . . . . .	38
<b>5</b>	<b>Lokale Parameterschätzung</b>	<b>39</b>
5.1	Optischer Fluß . . . . .	39
5.2	Strukturtensor . . . . .	41
5.3	Erweiterter Strukturtensor . . . . .	42
5.3.1	Diffusions-Schätzung . . . . .	43
<b>II</b>	<b>Algorithmik</b>	<b>47</b>
<b>6</b>	<b>Klassisches „Depth from focus“</b>	<b>49</b>
6.1	Verlauf der Unschärfeparameters . . . . .	49
6.2	Lokale Schärfemaße . . . . .	53
6.2.1	Statistischer Fehler . . . . .	53
6.2.2	Fehler durch bestimmte Bildstrukturen . . . . .	56
6.2.3	Lokale Varianz als Schärfemaß . . . . .	59
6.3	Bestimmung des Maximums . . . . .	63
6.3.1	Einfache Maximumsuche . . . . .	63
6.3.2	Schwerpunktbestimmung . . . . .	65
6.3.3	Lokale Polynominterpolation . . . . .	67
6.3.4	Globaler Polynomfit . . . . .	67
6.4	Synthese eines tiefenscharfen Bildes . . . . .	68



<b>7</b>	<b>Tiefe aus Diffusion</b>	<b>69</b>
7.1	Fokussiererei als Diffusionsprozeß . . . . .	69
7.2	Schätzung der Tiefe aus der Diffusionskonstanten . . . . .	71
<b>III</b>	<b>Messungen und Auswertung</b>	<b>75</b>
<b>8</b>	<b>Synthetische Sequenzen</b>	<b>77</b>
8.1	Berechnung der Sequenzen . . . . .	77
8.1.1	Geometrie . . . . .	77
8.1.2	2D-Punktantwort . . . . .	78
8.2	Test an synthetischen Sequenzen ohne Skalierung . . . . .	82
8.2.1	Bildanzahl / -abstand . . . . .	83
8.2.2	Rauschabhängigkeit . . . . .	83
8.2.3	Numerische Apertur . . . . .	83
8.2.4	Wellenlänge der Bildstrukturen . . . . .	84
8.3	Test an synthetischen Sequenzen mit Skalierung . . . . .	85
8.3.1	Rauschen und Bildabstand . . . . .	85
8.3.2	Szenengeometrie . . . . .	87
8.3.3	Oberflächentextur . . . . .	87
8.3.4	PSF . . . . .	89
<b>9</b>	<b>Reale Sequenzen</b>	<b>91</b>
9.1	Telezentrische Aufnahmen (Handmikroskop) . . . . .	91
9.1.1	Der Aufbau . . . . .	91
9.1.2	Auflösung des Handmikroskops . . . . .	92
9.1.3	Die verwendeten Testkörper . . . . .	92
9.1.4	Aufnahmen des IMS-Testkörpers . . . . .	93

9.2	Nicht telezentrische Aufnahmen . . . . .	94
9.2.1	Der Aufbau . . . . .	95
9.2.2	Testkörper . . . . .	96
9.2.3	Testszenen . . . . .	97
<b>10</b>	<b>Zusammenfassung und Ausblick</b>	<b>99</b>
<b>A</b>	<b>Fehlerfortpflanzung</b>	<b>103</b>
A.1	Erwartungswert und Varianz linearer verschiebungsinvarianter Filter . . .	103
A.2	Erwartungswert und Varianz des quadrierten Bildes . . . . .	104
<b>B</b>	<b>Beschreibung der Software</b>	<b>107</b>
B.1	Zusätzliche Heurisko-Operatoren . . . . .	107
B.1.1	focus.dll . . . . .	107
B.2	Workspaces . . . . .	108
B.2.1	handmikroskop aufnahme.ws . . . . .	108
B.2.2	newFokus.ws . . . . .	109
B.2.3	newFokusInc.ws . . . . .	109
	<b>Literaturverzeichnis</b>	<b>110</b>

# Kapitel 1

## Einleitung

Diese Arbeit entstand im Rahmen des *OpenEye*-Projekts, eines vom Land Baden-Württemberg geförderten Kooperationsprojekts zwischen Universitäten und der Industrie zur Entwicklung intelligenter Bildsensorik. Das Projekt bestand aus mehreren Teilprojekten, von denen sich eine Gruppe mit den Grundlagen, wie z.B. der CMOS-Sensor- und Kamera-Entwicklung am IMS<sup>1</sup>, der Entwicklung von Bildverarbeitungsalgorithmik am IWR<sup>2</sup> der Universität Heidelberg oder der Implementierung von ausgewählten Algorithmen in spezielle FPGA-Hardware<sup>3</sup> an der Universität Mannheim beschäftigte. Die andere Gruppe von Teilprojekten beschäftigte sich in Kooperation mit den jeweiligen Industriepartnern damit, anwendungsspezifische Probleme, basierend auf den geschaffenen Grundlagen, anzugehen. So wurden z.B. mit den Firmen Bosch bzw. Trumpf Schweißnähte mit den hochdynamischen CMOS-Sensoren des IMS beobachtet und in Echtzeit mit entsprechender FPGA-Hardware ausgewertet, während mit EM Gerätebau ein monookulares Tracking-System zur Verwendung in Trainingsgeräten für Augenoperationen entwickelt wurde.

In dem hier relevanten Teilprojekt sollten in Zusammenarbeit mit der Firma A-Tec die Möglichkeiten untersucht werden, den Einsatzbereich des Handmikroskop der Firma dahingehend zu erweitern, daß es in weiten Bereichen tiefenscharfe Bilder und Tiefenkarten, d.h. die Objektentfernung an jedem Bildpunkt, für entsprechende Meßaufgaben liefert. Bei den dafür entwickelten und analysierten Algorithmen sollte neben Genauigkeit und Geschwindigkeit auch auf Eigenschaften wie Einfachheit, Speicherbedarf usw. geachtet

---

<sup>1</sup>Institut für Mikroelektronik Stuttgart

<sup>2</sup>Interdisziplinäres Zentrum für wissenschaftliches Rechnen

<sup>3</sup>Ein FPGA (Field Programmable Gate Array) ist ein Chip mit einer Matrix von einfachen Rechenknoten, deren Verbindung untereinander und deren Funktion frei programmiert werden kann. Da all diese Rechenknoten unabhängig von einander und parallel arbeiten können, eignen sie sich auch für einfachere Bildverarbeitungsoperationen, die sie dann aber extrem schnell ausführen können.



*Abbildung 1.1: Das Handmikroskop der Firma A-Tec.*

werden. Diese Eigenschaften ermöglichen es, den entsprechenden Algorithmus in Spezial-Hardware wie z.B. einem FPGA ablaufen zu lassen und damit den sonst nötigen PC einzusparen.

Im Rahmen dieser Arbeit wurden daher verschiedene *Depth-from-focus*-Verfahren entwickelt und analysiert, die aus einer *Fokusserie*, einer Serie von Bildern mit unterschiedlichen Kameraparametern wie z.B. unterschiedlichen Fokuseinstellungen, eine Tiefenkarte errechnen. Aus dieser Tiefenkarte und der Fokusserie kann dann ein tiefscharfes Bild berechnet werden. Zudem wurde mit einem neuartigen Verfahren ein Problem von Fokusserien angegangen, die nicht mit einer speziellen telezentrischen Optik aufgenommen wurden. Mit der Variation der Kameraparameter ändert sich neben der Schärfe der Abbildung teilweise auch die Bildgröße der Objekte. Mit dem neuen Verfahren ist es möglich, simultan die Objektentfernung und die Skalierungseffekte durch die sich ändernden Bildgrößen zu bestimmen.



*Abbildung 1.2: Drei Bilder einer mit dem Handmikroskop aufgenommenen Fokusserie mit der daraus berechneten Tiefenkarte und dem tiefscharfen Bild.*

## 1.1 Aufbau der Arbeit

Die Arbeit ist in drei große Teile gegliedert. Im ersten Teil sind die Grundlagen zusammengefasst. Kapitel 2 befaßt sich mit den Eigenschaften der optischen Abbildung, insbesondere mit der dabei entstehende Unschärfe. Dazu werden die Punktantworten und Transferfunktionen der optischen Abbildung diskutiert. In Kapitel 3 wird ein Überblick über die unterschiedlichen optischen 3D-Meßtechniken gegeben, in denen die Depth-from-focus-Verfahren eingeordnet werden. Dabei wird herausgestellt, daß Stereo- und Depth-from-focus-Verfahren auf dem gleichen physikalischen Grundprinzip basieren. Anschließend werden in Kapitel 4 unterschiedliche DFF-Verfahren aus der Literatur genauer betrachtet. In Kapitel 5.2 folgt eine Beschreibung des Strukturtensor-Verfahrens zur Bestimmung von Bewegung bzw. anderer, den Grauwertänderungen zugrunde liegender Parameter. Dieses Verfahren wird später die Grundlage eines neuartigen Verfahrens zur Tiefenschätzung sein.

Der zweite Teil beschäftigt sich konkreter mit der Algorithmik zur Auswertung der Fokusserien. Kapitel 6 erläutert ausführlich den Ablauf der Auswertung und mögliche Fehlerquellen. Dabei wird auf die relevanten Bereiche der Bildentstehung eingegangen und daraus der Verlauf der Meßgrößen abgeleitet. Es werden verschiedene Schärfemaße diskutiert sowie unterschiedliche Möglichkeiten zur Detektion des Schärfemaximums beschrieben. In Kapitel 7 wird ein neuartiges Verfahren zur simultanen Bestimmung der Tiefe und der Bewegung bzw. Skalierung auf Basis des Strukturensors vorgestellt.

Im dritten Teil werden die beschriebenen Algorithmen an synthetischen (Kapitel 8) und realen (Kapitel 9) Fokusserien getestet. Abschließend erfolgt eine Zusammenfassung der Ergebnisse und ein Ausblick auf mögliche Erweiterungen der vorgestellten Verfahren.



**Teil I**

**Grundlagen**





# Kapitel 2

## Optik

In diesem Kapitel sollen die für diese Arbeit relevanten Grundlagen der Bildentstehung dargelegt werden, da diese der Ausgangspunkt für ein tiefgreifendes Verständnis der Algorithmen sind.

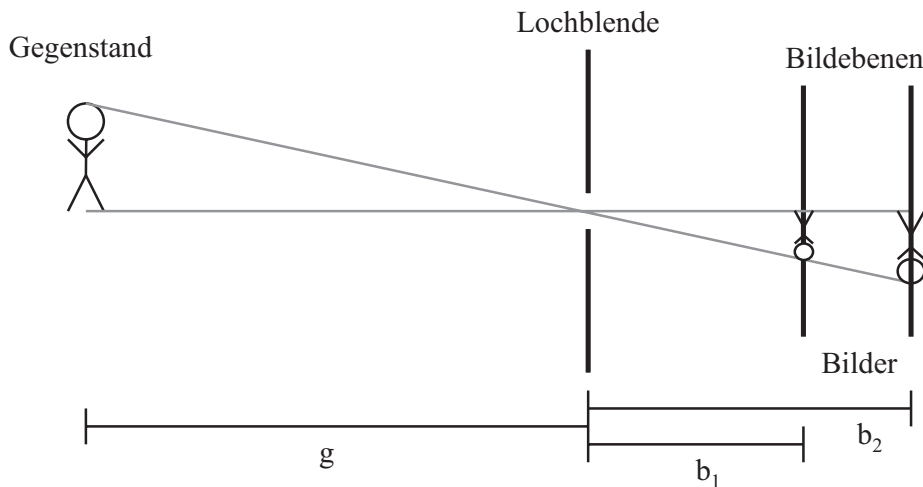
### 2.1 Geometrische Optik

In der geometrischen Optik wird das Licht durch Lichtstrahlen repräsentiert. Diese gehen z.B. von einer punktförmigen Lichtquelle in jede Richtung aus. Treffen sie auf eine Oberfläche, werden sie gebrochen, reflektiert oder absorbiert. Die Wellennatur des Lichtes wird dabei außer Acht gelassen, so daß Effekte wie Beugung oder Interferenz in der geometrischen Optik nicht modelliert werden können. Trotzdem eignet sich dieses Modell für eine Vielzahl von optischen Systemen sehr gut, zumindest für die Gewinnung von Näherungswerten.

#### 2.1.1 Lochkamera

Die Lochkamera ist ein idealisiertes System, bei dem die Lichtstrahlen - vom Objekt ausgehend - durch ein infinitesimal kleines Loch auf einen Schirm treffen. Auf diesem Schirm entsteht dann ein umgekehrtes Bild des Objekts, wie dies in Abbildung 2.1 skizziert ist.

Befindet sich das Objekt der Größe  $G$  dabei im Abstand  $g$  vor der Blende und der Schirm im Abstand  $b$  dahinter, ergibt sich die Bildgröße zu  $B = \frac{b}{g}G$ . Die Größe des Bildes hängt also von dem Objektabstand und dem Schirmabstand ab.



*Abbildung 2.1: Skizze der Lochkamera*

In diesem idealisierten Fall gelangt von jedem Objektpunkt nur genau ein Lichtstrahl durch die Blende. Der Schnittpunkt zwischen diesem Strahl und dem Schirm ergibt dann den jeweiligen Bildpunkt. Damit hängt die Position des Bildpunkts zwar von der Position des Schirms und vom Objektabstand ab, aber ein Punkt in der Welt wird immer auf einen Punkt auf dem Schirm abgebildet, so daß das Bild immer scharf ist.

In der Realität trifft dies aber nicht zu. Zum einen ist es physikalisch nicht möglich, das Loch unendlich klein zu machen. Macht man andererseits das Loch immer kleiner, so nehmen die Beugungseffekte zu, man verläßt den durch die geometrische Optik beschreibaren Bereich und das gesamte Bild wird unscharf.

### 2.1.2 Kamera mit endlicher Blende

Beim Übergang zu einem System mit endlicher kreisförmiger Blende wird zusätzlich ein optisches Element benötigt, das die vom Objektpunkt ausgehenden, divergierenden Lichtstrahlen wieder fokussiert. Wie in Abbildung 2.2 dargestellt, kann dies zum Beispiel durch eine dünne Linse bewerkstelligt werden. Bildseitig schneiden sich die so fokussierten Strahlen im Bildpunkt. Hieraus wird sofort klar, daß im Gegensatz zur Lochkamera Objektpunkte nur dann scharf abgebildet werden, wenn der Schirm den richtigen Abstand von der Linse hat. Liegt der Schirm näher an der Linse oder ist weiter von ihr entfernt, wird der Objektpunkt auf ein Kreisscheibchen abgebildet.

Der Mittelpunkt dieses Kreisscheibchens liegt auf dem Hauptstrahl, dem Strahl vom Ob-

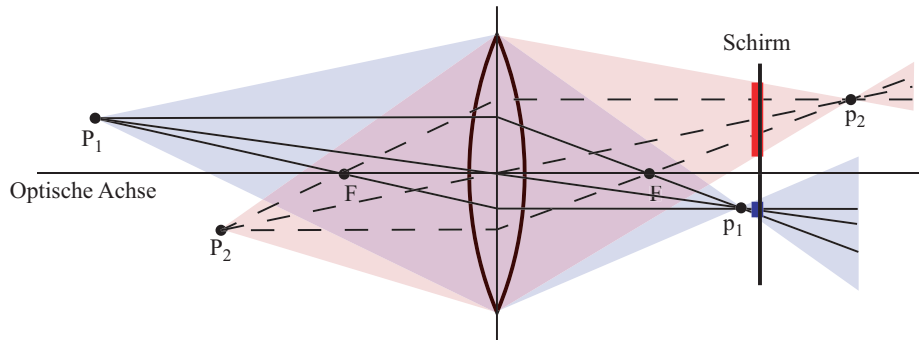


Abbildung 2.2: Kamera

jektspunkt durch den Mittelpunkt der Blende/Linse, und lässt sich damit wie bei der Lochkamera mit dem Strahlensatz bestimmen. Legt man dazu den Ursprung des Koordinatensystems in die Hauptebene der Linse und die Z-Achse entlang der Optischen Achse, ergibt sich für einen Weltspunkt  $P_1 = (X_1, Y_1, Z_1)^T$  der Mittelpunkt  $m_1$  des Kreisscheibchens auf dem um  $s$  hinter der Linse liegenden Schirm zu:

$$\vec{m}_1 = \begin{pmatrix} -X_1 \frac{s}{Z_1} \\ -Y_1 \frac{s}{Z_1} \\ -s \end{pmatrix} \quad (2.1)$$

Zu beachten ist hierbei, daß dies zunächst völlig unabhängig von der Brennweite  $f$  der Linse ist. Im Allgemeinen wird man allerdings den Abstand des Schirms  $s$  so einstellen, daß das Bild scharf abgebildet wird und damit der Bildpunkt  $p_1 = (x_1, y_1, z_1)$  auf dem Schirm zu liegen kommt. Der Zusammenhang zwischen  $Z_1$  und  $z_1$  ergibt sich aus der Gleichung für dünne Linsen

$$\frac{1}{g} + \frac{1}{b} = \frac{1}{f} \quad (2.2)$$

mit der Gegenstandsweite  $g = Z_1$ , der Bildweite  $b = -z_1$  und der Brennweite  $f$  der Linse zu:

$$z_1 = -\frac{Z_1 f}{Z_1 - f} = \frac{Z_1 f}{f - Z_1} \quad (2.3)$$

Die restlichen Koordinaten des Bildpunkts  $p_1$  erhält man - wie oben - mit dem Strahlensatz, so daß insgesamt folgt:

$$\vec{p}_1 = \begin{pmatrix} \frac{Z_1 f}{f - Z_1} \frac{X_1}{Z_1} \\ \frac{Z_1 f}{f - Z_1} \frac{Y_1}{Z_1} \\ \frac{Z_1 f}{f - Z_1} \end{pmatrix} = \begin{pmatrix} X_1 \frac{f}{f - Z_1} \\ Y_1 \frac{f}{f - Z_1} \\ Z_1 \frac{f}{f - Z_1} \end{pmatrix} = \frac{f}{f - Z_1} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} \quad (2.4)$$

Enthält die abzubildende Szene Objekte oder Objektteile mit unterschiedlichem Abstand zur Linse, werden diese auch in unterschiedlichem Abstand hinter der Linse scharf abge-

bildet. Auf einem ebenen Schirm kann also nie die gesamte 3D-Szene, sondern immer nur Teile davon scharf abgebildet werden, es sei denn die Szene selbst ist flach.

Um nun zu bestimmen „wie unscharf“ ein Weltpunkt abgebildet wird, benötigt man den Radius  $r$  des Kreisscheibchens in Abhängigkeit von der Schirmposition  $s$ . Für den Weltpunkt  $P_1$ , der wie oben gezeigt, im Abstand  $z_1$  hinter der Linse auf  $p_1$  scharf abgebildet wird, ergibt sich mit dem Strahlensatz

$$r(s) = r_{\text{Blende}} \frac{|(-z_1) - s|}{(-z_1)} \quad (2.5)$$

wobei  $r_{\text{Blende}}$  der effektive Blendenradius des Optischen Systems ist. Ist das betrachtete Objekt weiter als die Brennweite von der Linse entfernt, d.h.  $Z_1 > f$  und damit auch  $|z_1| > f$  und ist der Schirm in der Nähe der Brennweite hinter der Linse  $s \approx f$ , dann ist  $|z_1| > s$  und wir können schreiben:

$$r(s) = r_{\text{Blende}} \frac{(-z_1) - s}{(-z_1)} = \frac{r_{\text{Blende}}}{z_1} s + r_{\text{Blende}} \quad (2.6)$$

Setzt man für  $z_1$  ein (siehe Gleichung (2.3)) und benutzt statt des Blendenradius die Blendenzahl

$$n_{\text{Blende}} = \frac{f}{2r_{\text{Blende}}} \quad (2.7)$$

erhält man:

$$\begin{aligned} r(s) &= \frac{f - Z_1}{Z_1 2n_{\text{Blende}}} s + \frac{f}{2n_{\text{Blende}}} \\ &= -\frac{1}{2n_{\text{Blende}}} \left[ 1 - \frac{f}{Z_1} \right] s + \frac{f}{2n_{\text{Blende}}} \end{aligned} \quad (2.8)$$

Beschreibt man die Position des Schirms relativ zum scharf abgebildeten Bildpunkt  $p_1$ , also  $\tilde{s} = s - z_1 = s - Z_1 f / (Z_1 - f)$ , fällt der zweite Summand weg.

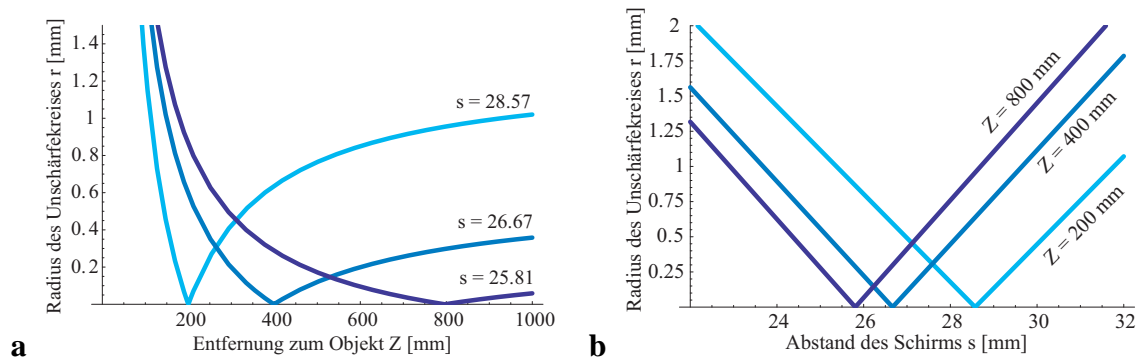
$$r(\tilde{s}) = -\frac{1}{2n_{\text{Blende}}} \left[ 1 - \frac{f}{Z_1} \right] \tilde{s} \quad (2.9)$$

Der Radius des Unschärfekreises steigt also linear mit dem Abstand des Schirms vom schärfsten Punkt aus an. Dieser Anstieg hängt von der Brennweite und dem Objektabstand ab und ist antiproportional zur Blendenzahl.

Interessant ist auch die Abhängigkeit des Radius des Unschärfekreises vom Objektabstand bei festem Schirm. Dazu schreibt man Gleichung (2.8) um:

$$r(Z_1) = \frac{fs}{2n_{\text{Blende}}} \frac{1}{Z_1} + \frac{f - s}{2n_{\text{Blende}}} \quad (2.10)$$

In Abbildung 2.3a ist der Verlauf des Radius  $r$  in Abhängigkeit von der Entfernung des Objektpunktes für unterschiedliche Schirmabstände aufgetragen. Der Schirmabstand beträgt dabei  $s_1 = 28.57$  mm,  $s_2 = 26.67$  mm bzw.  $s_3 = 25.81$  mm, so daß die Objekte in der Entfernung 200, 400 bzw. 800 mm scharf abgebildet werden. Die Brennweite der Linse beträgt  $f = 25$  mm und Blendenzahl  $n_{\text{Blende}} = 1.4$ . In Teil b ist der lineare Verlauf des Radius in Abhängigkeit vom Schirmabstand für die selben Objektentfernungen dargestellt.



**Abbildung 2.3:** Radius  $r$  des Unschärfekreises **a** in Abhängigkeit von der Entfernung des Objektpunktes  $Z$  für verschiedene Schirmabstände **b** in Abhängigkeit vom Schirmabstand  $s$  für unterschiedliche Entfernungen des Objektpunktes.

### 2.1.3 Schärfentiefe

Wie im vorherigen Abschnitt erläutert, werden bei Kameras mit endlicher Blende nur Gegenstände in einer bestimmten Entfernung scharf abgebildet. In vielen Fällen ist eine gewisse Unschärfe, d.h. Unschärfekreisradien bis zu einem Maximalradius  $\varepsilon$ , unproblematisch. Wird der Schirm z.B. durch einen CCD-Chip ersetzt, ist ein Unschärfekreisradius bis zu einer halben Pixelgröße tolerierbar. Daraus ergibt sich bildseitig ein Bereich, in dem der Schirmabstand variiert werden kann bzw. objektseitig ein Tiefenbereich, innerhalb dessen ein Objekt scharf abgebildet wird. Die Ausdehnung dieses Tiefenbereichs nennt man *Schärfentiefe*. Um sie zu berechnen, nähert man Gleichung (2.10) um den Objektabstand  $Z_s$ , der scharf abgebildet wird, in einer Taylorreihe bis zur ersten Ordnung.

$$\tilde{r}(Z) = r(Z_s) + r'(Z_s)(Z - Z_s) \quad (2.11)$$

$$= 0 - \frac{fs}{2n_f Z_s^2}(Z - Z_s) \quad (2.12)$$

Schreibt man  $\tilde{r}$  nun in Abhängigkeit von  $\Delta Z = Z - Z_s$  erhält man:

$$\tilde{r}(\Delta Z) = -\frac{fs}{2n_f Z_s^2} \Delta Z \quad (2.13)$$

Für den maximal zulässigen Unschärfekreisradius  $\varepsilon$  folgt aus  $\tilde{r}(\Delta Z) = \varepsilon$  in erster Näherung:

$$\Delta Z = \frac{2n_f Z_s^2}{f_s} \varepsilon = \frac{2n_f Z_s (Z_s - f)}{f^2} \quad (2.14)$$

D.h. Objektpunkte, die um maximal  $\Delta Z$  von der scharfgestellten Objektebene  $Z_s$  abweichen, werden auf eine Kreisscheibe mit einem Radius  $r$  kleiner als  $\varepsilon$  abgebildet. Im Fall von weit entfernten Objekten  $Z_s \gg f$  kann weiter genähert werden:

$$\Delta Z = \frac{2n_f Z_s^2}{f^2}, \quad \text{für } Z_s \gg f \quad (2.15)$$

Die Schärfentiefe ist also proportional zur Blendenzahl, zum Quadrat der Objektentfernung und antiproportional zum Quadrat der Brennweite. Mit einer Brennweite von  $f = 25$  mm, einer Blendenzahl von  $n_f = 1.4$ , einer Objektentfernung von  $Z_s = 800$  mm und eine Pixelgröße von  $10\mu\text{m}$  also einem maximalen Unschärfekreisradius von  $\varepsilon = 5\mu\text{m}$  erhält man zum Beispiel eine Schärfentiefe von  $\Delta Z \approx 14$  mm.

### 2.1.4 3D-PSF

Wie im vorherigen Abschnitt ausgeführt wird ein Punkt aus dem Objektraum im Bildraum nicht auf einen Punkt, sondern je nach Schirmposition auf Kreisscheibchen mit unterschiedlichem Radius abgebildet. Der Radius steigt linear mit dem Abstand des Schirms vom schärfsten Punkt an. Betrachtet man nun den gesamten dreidimensionalen Bildraum, so wird ein Objektpunkt auf einen Doppelkegel, d.h. zwei Kegel, die sich mit ihren Spitzen im schärfsten Punkt berühren, abgebildet, wie dies in Abbildung 2.4 veranschaulicht ist. Außerhalb dieses Doppelkegels bleiben die Punkte des Bildraums dunkel, während sich innerhalb die Intensität aus der Erhaltung der Strahlungsenergie ergibt. Verteilt man diese homogen innerhalb der jeweiligen Fläche der Kreisscheibchen, so fällt die Intensität mit dem Quadrat des Radius ab.

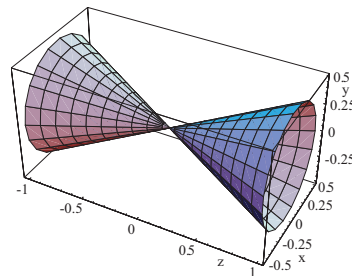


Abbildung 2.4: Doppelkegel der 3D-Punktantwort im Rahmen der geometrischen Optik

Man kann die optische Abbildung dann als linearen, verschiebungsinvarianten Filter beschreiben, wenn dieser Doppelkegel invariant gegenüber Verschiebungen ist, d.h. für jeden aus dem Objektraum abgebildeten Punkt bis auf eine Verschiebung gleich ist. Näherungsweise ist dies der Fall und man bezeichnet ihn somit als *Punktantwort* des Optischen Systems, da er beschreibt, wie ein einzelner abgebildeter Punkt durch das Optische System im Bildraum verändert wird. Die gesamte dreidimensionale Szene, wird durch Überlagerung der Doppelkegel für jeden einzelnen Punkt erzeugt. Dazu zerlegt man die Szene in einzelne Punkte  $\mathbf{X}$  mit ihrer jeweiligen Intensität  $g_o(\mathbf{X})$ , bildet dann jeden Punkt in den Bildraum ab und erhält dort die Intensität  $g'_i(\mathbf{x})$  am entsprechenden Bildpunkt  $\mathbf{x}$ . Den Einfluß der Optischen Abbildung kann man jetzt als Faltung mit der Punktantwort  $h$  im Bildraum schreiben und erhält die endgültige Intensitätsverteilung folgendermaßen:

$$g_i(\mathbf{x}) = \int_{-\infty}^{\infty} g'_i(\mathbf{x} - \mathbf{x}') h(\mathbf{x}') d^3 x' = (g'_i * h)(\mathbf{x}) \quad (2.16)$$

Benutzt man den Radius des Unschärfekreises, wie er in Gleichung (2.11) angegeben ist, kann man die dreidimensionale Punktantwort  $h$  leicht herleiten. Der Doppelkegel wird durch  $r(z)$  festgelegt, außerhalb davon ist die Punktantwort Null und innerhalb verteilt sich, wie oben erwähnt, die Intensität homogen über die Kreisscheibe, so daß man schreiben kann

$$h(\mathbf{x}) = \frac{1}{\pi r(z)^2} \Pi\left(\frac{(x^2 + y^2)^{1/2}}{2 r(z)}\right) \quad (2.17)$$

wobei die Rechteckfunktion  $\Pi(x)$  wie folgt definiert ist:

$$\Pi(x) = \begin{cases} 1 & |x| \leq 1/2 \\ 0 & \text{sonst} \end{cases} \quad (2.18)$$

Setzt man nun  $r(z)$  ein ergibt sich:

$$h(\mathbf{x}) = \frac{1}{\pi \left(\frac{1}{2n_f} \frac{f-Z_p}{Z_p} z\right)^2} \Pi\left(\frac{(x^2 + y^2)^{1/2}}{\frac{1}{n_f} \frac{f-Z_p}{Z_p} z}\right) \quad (2.19)$$

Wie man sieht, ist  $h$  nicht vollständig verschiebungsinvariant, denn der Term  $\frac{f-Z_p}{Z_p}$  ist von der Z-Koordinate  $Z_p$  des abgebildeten Objektpunkts  $P$  abhängig. Diese Abhängigkeit ist für  $f \ll Z_p$  gering und daher vernachlässigbar. Bei einer Brennweite von  $f = 25\text{mm}$ , einem Szenenabstand von  $Z_p = 1000\text{mm}$  und einer Szenentiefe von  $\Delta Z_p = 400\text{mm}$  variiert  $\frac{f-Z_p}{Z_p}$  gerade um 1%.

Lineare verschiebungsinvariante Filter können, wie oben beschrieben, durch ihre Punktantwort im Ortsraum dargestellt werden. Ihre Wirkung auf Intensitätsverteilungen wird

dann durch die Faltung vermittelt. Daneben besteht aber auch die Möglichkeit, sie im Frequenzraum anhand ihrer fouriertransformierten Punktantwort, der komplexwertigen *Transferfunktion*, zu beschreiben. An die Stelle der Faltung tritt dann eine einfache komplexe Multiplikation:

$$\hat{g}_i(\mathbf{k}) = \hat{g}'_i(\mathbf{k}) \hat{h}(\mathbf{k}) \quad (2.20)$$

Dabei sind  $\hat{g}_i(\mathbf{k})$  und  $\hat{g}'_i(\mathbf{k})$  die Fouriertransformierten der Intensitätsverteilungen  $g_i(\mathbf{x})$  und  $g'_i(\mathbf{x})$ . Die Fouriertransformierte der Punktantwort  $h(\mathbf{x})$ , die Transferfunktion  $\hat{h}(\mathbf{k})$ , wird für optische Systeme als *optische Transferfunktion* oder kurz *OTF* bezeichnet.

Der große Vorteil dieser Beschreibung im Frequenzraum liegt in der einfachen Darstellung der Filteranwendung als komplexe Multiplikation. Dadurch läßt sich deren Wirkungsweise ebenso einfach verstehen. Während die Basisfunktionen im Ortsraum Deltapeaks am jeweiligen Raumpunkt  $\mathbf{x}$  sind, wird der Frequenzraum von ebenen Wellen mit dem jeweiligen Wellenvektor  $\mathbf{k}$  aufgebaut. Jede dieser ebenen Wellen erfährt nun durch die Anwendung des Filters  $\hat{h}$  eine Skalierung ihrer Amplitude mit dem Faktor  $|\hat{h}(\mathbf{k})|$  und eine Phasenverschiebung um  $Phase(\hat{h}(\mathbf{k}))$ . Ist die Transferfunktion realwertig, tritt keine Phasenverschiebung auf. Anhand der Transferfunktion läßt sich also recht schnell die Wirkung eines Filters einschätzen.

Im Folgenden soll daher die Transferfunktion  $\hat{h}(\mathbf{k})$  berechnet werden. Dazu kann man die Separierbarkeit der Fouriertransformation ausnutzen und zuerst nur die 2D-Fouriertransformation in der XY-Ebene der Punktantwort durchführen. In dieser Ebene ist die Punktantwort laut Gleichung (2.17) eine Kreisscheibe, deren 2D-Fouriertransformierte die Airy-Funktion ist:

$$\frac{1}{\pi r^2} \Pi\left(\frac{|\mathbf{x}|}{2r}\right) \xleftrightarrow{\mathcal{F}} \frac{2J_1(|\mathbf{k}|r)}{|\mathbf{k}|r} \quad (2.21)$$

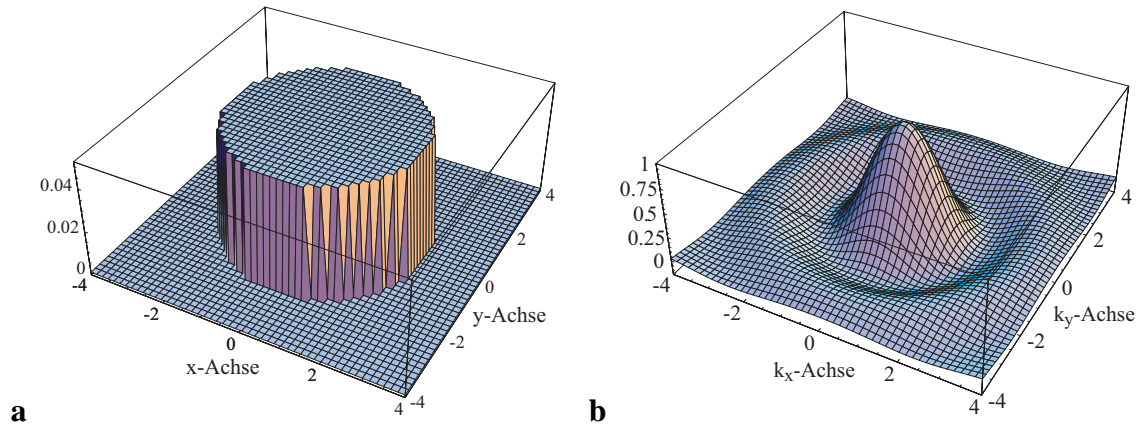
Dabei kennzeichnet das Symbol  $\xleftrightarrow{\mathcal{F}}$  Fouriertransformationspaare, d.h. eine Funktion und ihre Fouriertransformierte und  $J_1$  ist die *Besselfunktion* erster Ordnung.

Mit der Abkürzung  $\alpha = \frac{1}{2n_f} \frac{f-Z_p}{Z_p}$  und unter Verwendung von Gleichung (2.19) und (2.21) ergibt sich für die Fouriertransformation in X- und Y-Richtung:

$$\check{h}(k_x, k_y, z) = \frac{2J_1((k_x^2 + k_y^2)^{1/2} \alpha z)}{(k_x^2 + k_y^2)^{1/2} \alpha z}, \quad \alpha = \frac{1}{2n_f} \frac{f - Z_p}{Z_p} \quad (2.22)$$

In Abbildung 2.5 ist der Verlauf der Kreisscheibe und der Airy-Funktion dargestellt. Die Airy-Funktion fällt zunächst mit steigendem Abstand vom Ursprung schnell ab, so daß die Bildstrukturen mit steigender Wellenzahl zunehmend gedämpft werden. Da sie dann aber das Vorzeichen wechselt und ihr Betrag erst einmal wieder ansteigt, werden noch höher frequente Strukturen weniger stark gedämpft und ihre Phase wird um 180 Grad verschoben.





**Abbildung 2.5:** Die Kreisscheibe (a) und ihre Fouriertransformierte, die Airy-Funktion (b).

Da ihr Betrag insgesamt gesehen mit  $1/(k_x^2 + k_y^2)^{1/2}$  fällt, werden sehr hohe Frequenzen trotz sich wiederholender Vorzeichenwechsel weitgehend weggedämpft.

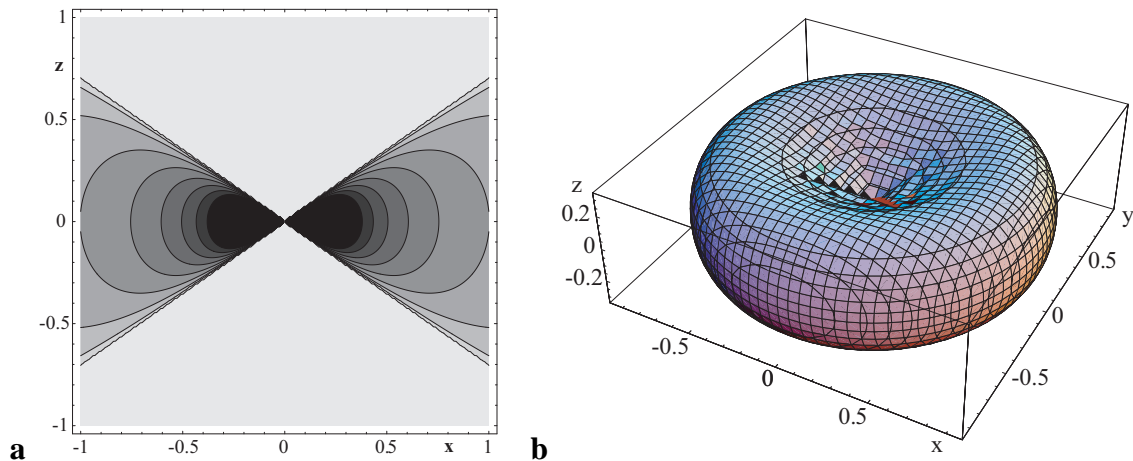
Um schließlich die dreidimensionale OTF zu erhalten, muß noch die verbleibende 1D-Fouriertransformation in z-Richtung durchgeführt werden. Dafür benutzt man folgendes Fouriertransformationspaar

$$\frac{J_1(2\pi x)}{x} \xleftrightarrow{\mathcal{F}} 2 \left(1 - \left(\frac{k}{2\pi}\right)^2\right)^{1/2} \Pi\left(\frac{k}{4\pi}\right) \quad (2.23)$$

und das Ähnlichkeitstheorem der Fouriertransformation, welches besagt, daß aus  $f(x) \xleftrightarrow{\mathcal{F}} \hat{f}(k) \Rightarrow f(ax) \xleftrightarrow{\mathcal{F}} \frac{1}{|a|} \hat{f}(k/a)$  folgt und erhalten:

$$\hat{h}(\mathbf{k}) = \frac{4}{|(k_x^2 + k_y^2)^{1/2} \alpha|} \left(1 - \frac{k_z^2}{(k_x^2 + k_y^2) \alpha^2}\right)^{1/2} \Pi\left(\frac{k_z}{2(k_x^2 + k_y^2)^{1/2} \alpha}\right) \quad (2.24)$$

In Abbildung 2.6 ist OTF  $\hat{h}(\mathbf{k})$  zur Veranschaulichung zwei- und dreidimensional dargestellt. In der 3D-Darstellung ist eine Fläche mit konstantem Funktionswert  $\hat{h}(\mathbf{k}) = c$  gezeigt, während in der 2D-Darstellung ein XZ-Schnitt zu sehen ist. Wie auch an der Rechteckfunktion  $\Pi$  in Gleichung (2.24) leicht zu erkennen, ist die OTF in einem Doppelkegel in z-Richtung gleich Null, während sie im verbleibenden Volumen antiproportional zu dem Abstand vom Ursprung abnimmt. Daher werden Strukturen, die in stark z-Richtung orientiert sind, vollständig unterdrückt, haben sie dagegen eine ausreichende Komponente in der xy-Ebene, werden sie nur gedämpft.

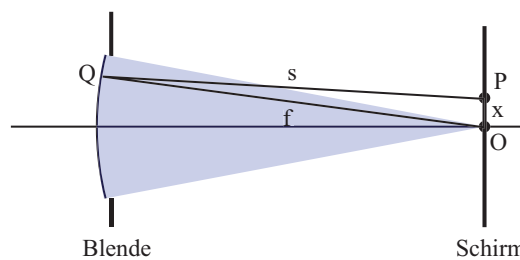


**Abbildung 2.6:** Die dreidimensionale Optische Transferfunktion  $\hat{h}(\mathbf{k})$  als a) 2D-Kontur-Plot und als b) 3D-Oberfläche.

## 2.2 PSF in der Wellenoptik

Die Berechnung der zwei- und besonders der dreidimensionalen Punktantwort unter Berücksichtigung der Wellennatur des Lichts ist recht aufwendig. Daher werden hier nur kurz die für diese Arbeit relevanten Ergebnisse aufgeführt. Für eine detaillierte Abhandlung sei z.B. auf Born and Wolf [1980] oder Hecht [1989] verwiesen.

Zur Berechnung wird das Prinzip von Huygens verwandt, welches besagt, daß jeder Punkt einer Wellenfront als Ausgangspunkt einer sekundären Kugelwelle betrachtet werden kann. Die Einhüllende der Kugelwellen bildet dann die fortgeschrittene neue Wellenfront. Die Intensität an einem Punkt kann damit durch die Überlagerung all dieser Kugelwellen berechnet werden.



**Abbildung 2.7:** Skizze zur Berechnung der 2D-Punktantwort in der Wellenoptik

Um die 2D-Punktantwort zu berechnen, geht man davon aus, daß die Wellenfront in der Blende durch das optische System schon so verändert wurde, daß sie auf einen Punkt  $O$  auf

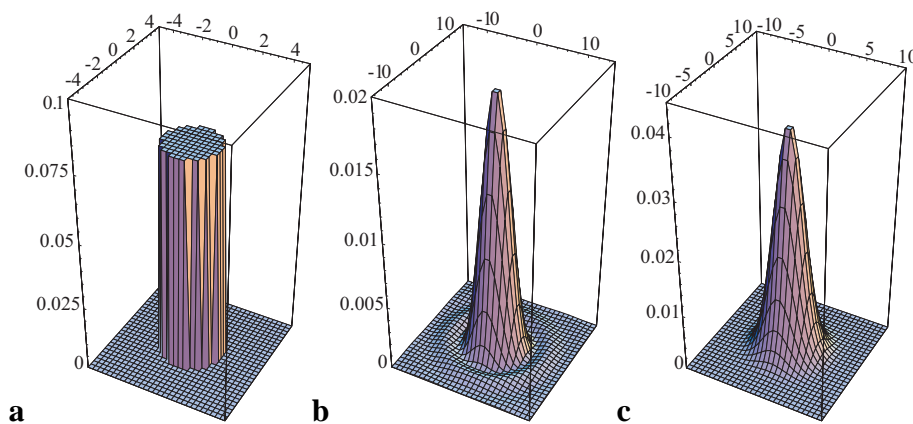
dem Schirm zuläuft und in ihm konvergiert, wie in Abbildung 2.7 skizziert. D.h. der Punkt  $O$  ist im Sinne der geometrischen Optik der scharf abgebildete Bildpunkt. Die Intensität an einem Punkt  $P$  in der Umgebung von  $O$  ergibt sich durch Integration über alle Teile der Wellenfront unter Berücksichtigung der Phasenverschiebung durch den variierenden Abstand.

Damit ergibt sich für eine kreisförmige Blende mit monochromatischem Licht als zweidimensionale Punktantwort die Airy-Funktion:

$$h(\mathbf{x}) = \frac{1}{4\pi\sigma^2} \left( \frac{2J_1(|\mathbf{x}|/\sigma)}{|\mathbf{x}|/\sigma} \right)^2, \quad \sigma = \frac{\lambda}{2\pi r_{\text{Blende}}} \quad (2.25)$$

Da sich die Position der Minima und Maxima der Airy-Funktion schnell mit der Wellenlänge ändert, führt eine Überlagerung vieler Airy-Funktionen mit unterschiedlichen Wellenlängen zu einer Gaußfunktion. Die 2D-PSF nimmt also bei Verwendung von polychromatischem bzw. weißem Licht die Form einer Gaußfunktion an.

$$h(\mathbf{x}) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\mathbf{x}^2}{2\sigma^2}\right) \quad (2.26)$$

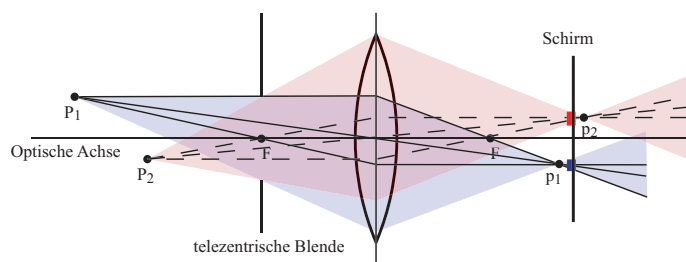


**Abbildung 2.8:** Zweidimensionale Punktantworten im Vergleich: **a** Die Kreisscheibe (Pillbox) aus der geometrischen Optik, **b** Die Airy-Funktion für monochromatische Beleuchtung und Wellenoptik sowie **c** Die Gaußfunktion bei polychromatischer Beleuchtung und Wellenoptik.

In der Abbildung 2.8 sind die drei unterschiedlichen 2D-Punktantworten nebeneinander skizziert. Für kreisförmige Blenden ergibt sich im Rahmen der geometrischen Optik unter Vernachlässigung von Abbildungsfehlern eine Kreisscheibe (Teilbild a). Bei Berücksichtigung der Wellennatur erhält man bei monochromatischer Beleuchtung die Airy-Funktion (Teilbild b) und bei polychromatischer bzw. weißer Beleuchtung eine Gaußfunktion (Teilbild c).

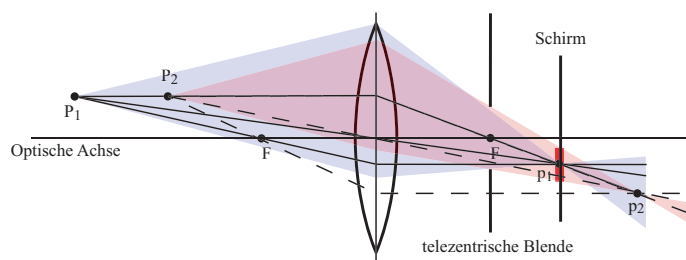
## 2.3 Telezentrische Optik

Bei der telezentrischen Optik handelt es sich um ein spezielles im Rahmen dieser Arbeit verwendetes optisches Abbildungssystem, bei dem Variationen des Objektabstandes (objektseitige Telezentrie) und/oder des Schirmabstandes (bildseitige Telezentrie) nicht, wie bei den in Abschnitt 2.1.1 und 2.1.2 beschriebenen optischen Systemen, zu Veränderungen der Bildgröße führen. Um dies zu erreichen fügt man in den Strahlengang am bild- bzw. objektseitigen Brennpunkt eine bzw. zwei zusätzliche Blenden ein.



*Abbildung 2.9: Skizze einer bildseitig telezentrischen Optik.*

In Abbildung 2.9 ist ein bildseitig telezentrisches System skizziert. Die zusätzliche Blende im objektseitigen Brennpunkt schränkt das Strahlenbündel so ein, daß die Mittelachse des im Bildraum entstehenden Doppelkegels parallel zur optischen Achse liegt. Dadurch ist der Mittelpunkt des Unschärfekreises unabhängig von der Schirmposition und es ergibt sich keine Änderung in der Bildgröße, während die Konstruktion bzw. Berechnung des scharf abgebildeten Bildpunkts  $p_i$  von der zusätzlichen Blende unbeeinflusst bleibt.



*Abbildung 2.10: Skizze einer objektseitig telezentrischen Optik.*

Bei dem objektseitig telezentrischen Aufbau, wie in Abbildung 2.10 zu sehen, liegen die Mittelachsen der Doppelkegel zwar schief im Bildraum, allerdings sind sie für Objektpunkte in unterschiedlicher Entfernung gleich. D.h. mit dem Objektabstand ändert sich der Radius des Unschärfekreises, aber nicht sein Mittelpunkt.

# Kapitel 3

## 3D-Meßtechniken

In diesem Kapitel soll ein Überblick über einige bildbasierte 3D-Meßtechniken, ihre Gemeinsamkeiten und Unterschiede gegeben werden. Im Wesentlichen lassen sich die Verfahren nach dem ihnen zugrundeliegenden Prinzip in zwei Hauptgruppen unterteilen. Die eine Gruppe sind die Triangulations-Verfahren, zu denen z.B. Stereo, Depth from Motion und Depth from Focus gehören. Die andere Gruppe sind die Laufzeitverfahren, zu denen z.B. die Interferometrie gehört. Die einzelnen Verfahrensgruppen, wie z.B. Stereo, lassen sich weiter unterteilen, z.B. in aktive und passive Verfahren. Während die aktiven Verfahren eine spezielle Beleuchtung benutzen, um die 3D-Information zu gewinnen, sind die passiven unabhängig von einer solchen Beleuchtung. Eine ausführliche Übersicht ist bei [Schwarte et al., 1999] zu finden.

### 3.1 Triangulations-Verfahren

Allen Triangulations-Verfahren ist gemein, daß sie ein Objekt aus zwei oder mehr Blickwinkeln betrachten und den sich daraus ergebenden Unterschied auswerten, um die Tiefeninformation zu rekonstruieren. Daß dies auf sehr unterschiedliche Weise geschehen kann, ist an den drei im Folgenden vorgestellten Verfahren zu erkennen. Auf den ersten Blick ist es dabei nicht selbstverständlich, daß ihnen dasselbe Prinzip zugrunde liegt.

#### 3.1.1 Stereo

Ein einfaches Stereosystem ist in Abbildung 3.1 gezeigt. Es besteht aus zwei Kameras mit parallelen optischen Achsen im Abstand  $b$ , der als *stereo Basis* bezeichnet wird. Die

Bildebene liegt jeweils im Abstand  $s$  hinter der Linse und der Ursprung der Kamerakoordinatensysteme befindet sich in der Linse, wobei die  $z$ -Achse auf der Optischen Achse zu liegen kommt. Das Weltkoordinatensystem befindet sich in gleicher Orientierung in der Linsenebene mittig zwischen der linken und rechten Kamera.

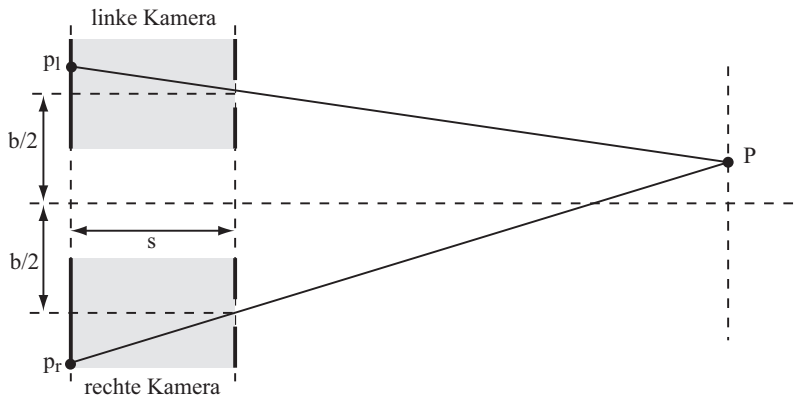
Ein Weltpunkt  $P$  wird in den beiden Kameras auf die unterschiedlichen Bildpunkte  $p_l$  und  $p_r$  abgebildet. Ihre Koordinaten ergeben sich, wie in Kapitel 2.1.1 dargelegt, durch Zentralprojektion:

$$\mathbf{p}_l = \begin{bmatrix} x_l \\ y_l \\ z_l \end{bmatrix} = -\frac{s}{Z} \begin{bmatrix} X - b/2 \\ Y \\ Z \end{bmatrix} \quad (3.1)$$

$$\mathbf{p}_r = \begin{bmatrix} x_r \\ y_r \\ z_r \end{bmatrix} = -\frac{s}{Z} \begin{bmatrix} X + b/2 \\ Y \\ Z \end{bmatrix}$$

Der Abstand  $p$ , in dem die beiden Bildpunkt in der Bildebene liegen, wird als *Parallaxe* bzw. *Disparität* bezeichnet. Aus Gleichung (3.1) für die Bildkoordinaten folgt:

$$p = |\mathbf{p}_r - \mathbf{p}_l| = \left| -\frac{s}{Z}(X - b/2) + \frac{s}{Z}(X + b/2) \right| = b \frac{s}{Z} \quad (3.2)$$



**Abbildung 3.1:** Skizze eines Stereo-Aufbaus

Aus einer gemessenen Parallaxe  $p$  erhält man also die Entfernung durch

$$Z = \frac{bs}{p}. \quad (3.3)$$

Ist diese Messung mit einem Fehler  $\Delta p$  behaftet, so ergibt sich daraus der Fehler in der Entfernung von:

$$\Delta Z = \frac{\partial Z}{\partial p} \Delta p = -\frac{bs}{p^2} \Delta p = -\frac{Z^2}{bs} \Delta p \quad (3.4)$$

D.h. der Fehler in der Tiefenmessung  $\Delta Z$  ist proportional zum Quadrat des Beobachtungsabstands  $Z$  und antiproportional zu Stereobasis  $s$ .

Die Tiefenbestimmung wurde damit auf eine Bestimmung der Parallaxe in den zwei Bildern reduziert. Die generelle Vorgehensweise ist dabei, daß die beiden korrespondierenden Bildpunkte  $p_l$  und  $p_r$  gesucht werden, die ein und derselbe Objektpunkt auf den Schirmen der linken und rechten Kamera erzeugt. Man geht dabei davon aus, daß eine kleine Umgebung von  $P$  auf entsprechende Umgebungen von  $p_l$  und  $p_r$  derart abgebildet werden, daß sie bezüglich bestimmter Merkmale ähnlich bzw. gleich sind. Für jeden Bildpunkt im Bild der linken Kamera wird dann nach dem Bildpunkt im rechten Kamerabild gesucht, der ihm bezüglich dieses Merkmals am ähnlichsten ist. Diese Suche muß nicht im gesamten Bild erfolgen, sondern kann aufgrund der bekannten Kameraanordnung auf eine Linie eingeschränkt werden.

Zwei wichtige grundlegende Probleme, die die Suche nach den korrespondierenden Punkten erschweren, sollen noch erwähnt werden:

- *Fehlende Objektstruktur*: Hat die zu vermessende Szene in Teilbereichen keine bzw. keine ausreichende Struktur oder nur Struktur senkrecht zur und nicht entlang der Stereobasis, so ist es in diesen Bereichen unmöglich die korrespondierenden Bildpunkte eindeutig zu bestimmen, da alle Bildpunkte in diesem Bereich das gleiche Merkmal aufweisen.
- *Verdeckungen*: Hat die Szene starke Tiefenvariationen und ist die Stereobasis groß, so kann es Bereiche geben, die nur in einer Kamera abgebildet werden und in der anderen durch andere Szenenteile verdeckt sind. In diesen Bereichen ist es dann nicht möglich, die korrespondierenden Punkte zu finden, da es keinen Bildpunkt mit dem passenden Merkmal gibt.

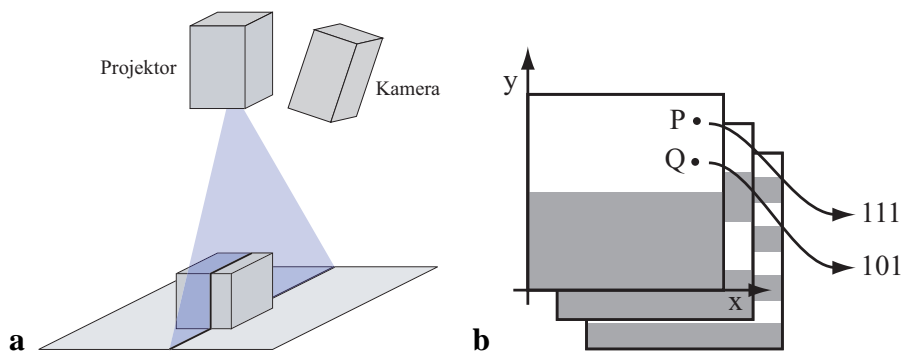
### Aktives Stereo

Das Problem der fehlenden Objektstruktur wird versucht, mit aktiven Stereo-Verfahren anzugehen. Die eine Kamera der bisher besprochenen passiven Verfahren wird dabei durch einen Projektor ersetzt. Im Bild der Kamera wird dann das vom Projektor projizierte Signal ausgewertet und nicht die Grauwertstruktur der Szene. Dadurch wird das Verfahren unabhängig von der jeweiligen Grauwertstruktur, solange das Signal des Projektors nicht durch diese Grauwertstruktur zu stark überlagert wird.

Das Signal des Projektors kann dabei recht unterschiedliche Formen annehmen. Es ist möglich, nur einen Punkt oder eine Linie zu projizieren und dann den Versatz im Kamerabild

auszuwerten, wie dies in Abbildung 3.2a skizziert ist. Um die gesamte Oberfläche zu erfassen, muß allerdings in ein oder zwei Dimensionen gescannt werden, was recht aufwendig ist.

Bei anderen Verfahren werden nacheinander  $n$  binäre Muster projiziert (s. Abb. 3.2 b), so daß die Position jedes Bildpunktes im Projektor in diesem Muster codiert ist. Aus den  $n$  aufgenommenen Kamerabildern kann für jeden Kamerabildpunkt der Code und damit die Position des entsprechenden Projektorbildpunktes bestimmt werden. Daraus ergibt sich dann die Parallaxe und damit die Entfernung des Objektpunktes. Mit  $n$  Bildern können  $n^2$  Positionen codiert werden, so daß mit steigenden Genauigkeitanforderungen immer mehr Muster projiziert und Bilder aufgenommen werden müssen. Will man Subpixel-Genauigkeit erreichen, werden häufig sinusförmige Muster projiziert; aus der Phasenlage läßt sich dann subpixelgenau auf die Position des Projektorbildpunktes schließen. Da dazu nur drei Muster benötigt werden, kann die Projektion auch gleichzeitig in den drei Farbkanälen erfolgen, wodurch diese Technik echtzeitfähig werden kann. (siehe z.B. Häusler and Ritter [1993])



**Abbildung 3.2:** **a** Skizze eines einfachen Streifenlichtprojektors. **b** Veranschaulichung des Gray-Codes: Die  $y$ -Position der Punkt  $P$  und  $Q$  in der Projektorbildebene ist durch den binären Helligkeitscode in den aufeinanderfolgenden Projektionen festgelegt.

All diesen aktiven Verfahren ist gemein, daß sie im Gegensatz zu den passiven Verfahren auch strukturlose Oberflächen vermessen können. Problematisch können allerdings stark strukturierte Oberflächen oder eine sehr helle Umgebung sein, da dann die Helligkeit des Projektors nicht mehr ausreicht, die Muster mit ausreichendem Kontrast zu projizieren.

### 3.1.2 Depth from Motion

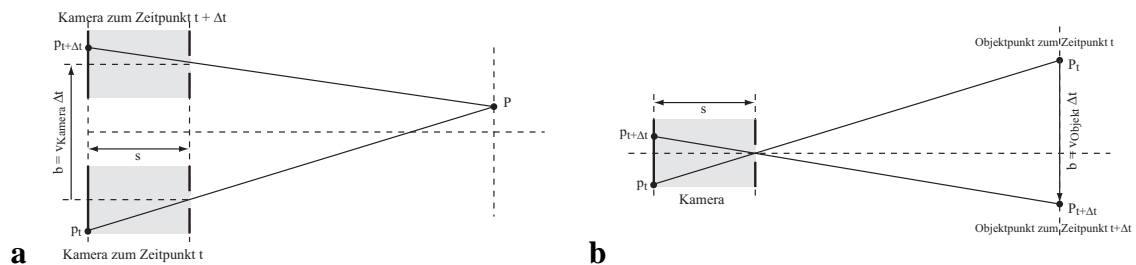
Depth-from-Motion ist den Stereo-Verfahren sehr ähnlich. Prinzipiell wird die zweite Kamera durch eine Aufnahme nach einer bestimmten Zeit  $\Delta t$  ersetzt, wobei sich entweder die Kamera oder aber auch das Objekt selbst bewegt haben können (siehe Abbildung 3.3).



Die Parallaxe  $p$  ergibt sich dann aus den zeitlich nacheinander aufgenommenen Bildern, wobei die Basislänge  $b$  der Entfernung entspricht, um die sich die Kamera bzw. das Objekt in der Zeit  $\Delta t$  bewegt hat. Die Entfernung erhält man wie beim Stereo-Verfahren:

$$Z = s \frac{b}{p} = s \frac{v_{\text{Kamera}} \Delta t}{v_{\text{Bildpunkt}} \Delta t} = s \frac{v_{\text{Kamera}}}{v_{\text{Bildpunkt}}} \quad (3.5)$$

Dabei wurden die Geschwindigkeiten der Kamera bzw. des Objektes  $v_{\text{Kamera}}$  und des Bildpunktes  $v_{\text{Bildpunkt}}$  im Bild und deren Zusammenhang mit der Basislänge  $b = v_{\text{Kamera}} \Delta t$  und der Parallaxe  $p = v_{\text{Bildpunkt}} \Delta t$  benutzt. Man kann also die Entfernung  $Z$  eines Objektpunktes bei bekannter Kamera- bzw. Objektgeschwindigkeit direkt aus der Geschwindigkeit seines Bildpunktes im Bild berechnen.



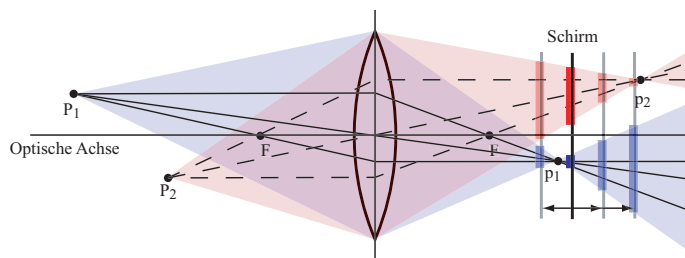
**Abbildung 3.3:** Skizze eines *Depth-from-Motion-Systems* **a** bei bewegter Kamera, **b** bei bewegtem Objekt.

### 3.1.3 Depth from Focus

Alle *Depth-from-Focus*-Verfahren nutzen zur Bestimmung der Tiefeninformation die Tatsache, daß optische Systeme, wie in Kapitel 2 beschrieben, nur Teile einer 3D-Szene scharf abbilden können. Hierzu werden im Allgemeinen ein oder mehrere Bilder der Szene mit unterschiedlichen Kameraeinstellungen aufgenommen. Je nach Anzahl der aufgenommenen Bilder ergeben sich dabei unterschiedliche Vorgehensweisen. Wurde nur ein Bild aufgenommen, wird meist a priori Wissen über die Szene herangezogen, um aus dem Vergleich mit dem Bild auf die Tiefeninformation zu schließen. Bei zwei oder mehreren Bildern kann die Tiefe aus deren Vergleich bzw. dem Verlauf gewisser Bildmerkmale über die gesamte Sequenz errechnet werden. Die verschiedenen Verfahren, die dazu entwickelt wurden, werden in Kapitel 4 genauer vorgestellt. Hier sollen lediglich das grundlegende Prinzip kurz skizziert sowie die Gemeinsamkeiten und Unterschiede zu den anderen Triangulationsverfahren herausgestellt werden.

## Grundprinzip

In Abbildung 3.4 ist ein einfaches System zur Aufnahme einer Bildsequenz mit unterschiedlichen Kameraparametern skizziert. In diesem Beispiel wurde die Position  $s$  des Schirms bzw. CCDs im Laufe der Sequenz verändert. Für die unterschiedlichen Schirmpositionen sind die Unschärfekreise der zwei Objektpunkte  $P_1$  und  $P_2$  durch blaue und rote Balken gekennzeichnet.



**Abbildung 3.4:** Skizze eines optischen Systems zur Tiefenbestimmung mittels Depth-from-Focus. Darin sind die unterschiedlichen Schirmpositionen, mit denen die verschiedenen Aufnahmen gemacht werden, gezeigt.

Könnte man nun die Radien  $r_i$  der Unschärfekreise, die sich in jedem Bild der Sequenz für ein und denselben Objektpunkt ergeben, direkt messen, würden sich mit den aus Abschnitt 2.1.2 bekannten Zusammenhängen zwischen Radius, Schirmposition und Objektpunktentfernung auf die Objektpunktentfernung schließen lassen. Im Allgemeinen besteht die Szene allerdings nicht aus vereinzelt punktförmigen, sondern vielmehr aus ausgedehnten Objekten. Die Unschärfekreise bzw. allgemeiner die Punktantworten überlappen sich, so daß eine direkte Messung ihrer Radien bzw. Halbwertsbreiten nicht möglich ist. Man mißt daher ein lokales Merkmal in den Bildern, das folgender Minimalanforderung genügen muß: Die Extrema der Halbwertsbreite der Punktantwort bzgl. der Schirmposition sollen Extrema in diesem Merkmal entsprechen. Dann kann mit Hilfe dieses Merkmals für jeden Objektpunkt die Schirmposition  $s_z$  gefunden werden, bei der die Halbwertsbreite der Punktantwort minimal ist. Mit der aus Abschnitt 2.1.2 bekannten Abbildungsgleichung für dünne Linsen (2.2) läßt sich daraus die Entfernung  $Z$  des Objektpunkts bestimmen:

$$Z = \frac{f s_z}{s_z - f} \quad (3.6)$$

Um den Fehler in der Tiefenschätzung zu berechnen, nähert man nun die oben stehende Gleichung um den Punkt  $s = s_z$  in einer Taylor-Reihe bis zur ersten Ordnung:

$$\tilde{Z}(s) = Z(s_z) + \frac{\partial Z(s_z)}{\partial s} (s - s_z) \quad (3.7)$$

Mit der Ableitung

$$\frac{\partial Z}{\partial s} = -\frac{f^2}{(s-f)^2} = -\frac{Z^2}{s^2} \quad (3.8)$$

ergibt sich dann:

$$\tilde{Z}(s) = Z(s_Z) - \frac{Z^2}{s_Z^2}(s - s_Z) \quad (3.9)$$

Aus einem Fehler in der Bestimmung der Schirmposition  $\Delta s = s - s_Z$  ergibt sich also ein Tiefenfehler von:

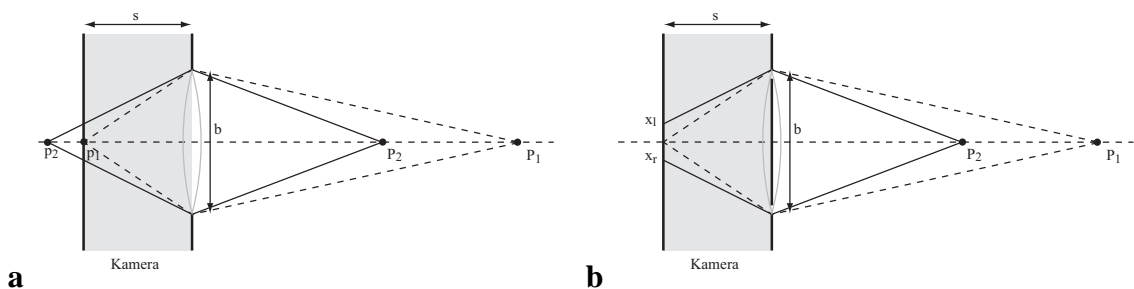
$$\Delta Z = |\tilde{Z}(s) - Z(s_Z)| = \frac{Z^2}{s_Z^2} \Delta s \quad (3.10)$$

Wie auch schon bei den Stereo-Verfahren ergibt sich auch hier ein Anstieg im Fehler mit der Entfernung im Quadrat.

Die Bestimmung der Halbwertsbreiten der Punktantwort bzw. des davon abhängigen Merkmals muß, wie oben schon erwähnt, über die Bildsequenz für denselben Objektpunkt erfolgen. Wie aus Abbildung 3.4 hervorgeht, wandert der Mittelpunkt der Punktantwort mit der Schirmposition. Entweder muß man dies bei der Auswertung der Bilder entsprechend berücksichtigen oder aber man benutzt die in Abschnitt 2.3 beschriebene bildseitig telezentrische Optik. Bei dieser speziellen Optik wird ein Objektpunkt unabhängig von der Position des Schirms immer auf die gleichen Bildkoordinaten abgebildet.

### Depth-from-Focus als Triangulationsverfahren

Schon das gleiche Verhalten des Fehlers mit der Entfernung legt eine Ähnlichkeit zu den Triangulationsverfahren nahe. Im Folgenden soll dies veranschaulicht werden, für einen ausführlichen Vergleich der Verfahren sei auf Schechner and Kiryati [2000] verwiesen.



**Abbildung 3.5:** **a** Skizze eines Fokussystems. **b** Skizze eines Fokussystems bei dem die Blende bis auf zwei entgegengesetzt liegende Löcher abgedeckt wurde.

In Abbildung 3.5a ist ein optisches System mit Blendenradius  $r_f = b/2$  und Brennweite  $f$  gezeigt, das auf den Punkt  $P_1 = [X_1, Y_1, Z_1]^T$  fokussiert ist. D.h. der Schirmabstand

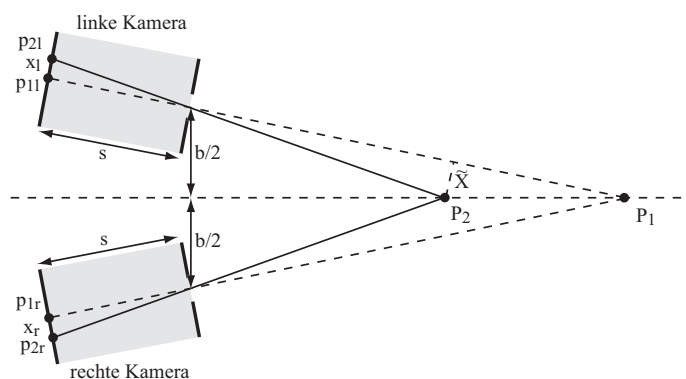
beträgt  $s = z_1 = Z_1 f Z_1 - f$ . Der näher an der Linse liegende Punkt  $P_2 = [X_2, Y_2, Z_2]^T$  wird auf eine Kreisscheibe abgebildet, für deren Radius sich laut Abschnitt 2.1.2 ergibt:

$$r = \frac{b}{2} \left[ \frac{s}{Z_2} + \frac{f-s}{f} \right] \quad (3.11)$$

Deckt man nun, wie in Abbildung 3.5b skizziert, die komplette Blende bis auf zwei gegenüberliegende kleine Löcher ab, bleiben von der PSF, die bisher die Form eines Doppelkegels hatte, nur zwei einzelne Randstrahlen übrig. Für fokussierte Objektpunkte schneiden sich diese Strahlen auf dem Schirm, so daß man weiterhin einen einzelnen Bildpunkt erhält. Nicht fokussierte Objektpunkte wie z.B.  $P_2$  werden dagegen auf zwei getrennte Bildpunkte abgebildet. Diese in Abbildung 3.5b mit  $p_l$  und  $p_r$  bezeichneten Bildpunkte haben einen Abstand von  $2r$ , dem Durchmesser des Unschärfekreises.

$$|p_1^l - p_1^r| = 2r = b \left[ \frac{s}{Z_2} + \frac{f-s}{f} \right] \quad (3.12)$$

Die Tatsache, daß hier ein Objektpunkt auf zwei Bildpunkte mit entfernungsabhängigem Abstand abgebildet wird, kann als weiteres Indiz für die Ähnlichkeit zu Stereoverfahren gesehen werden.



**Abbildung 3.6:** Skizze eines Stereo-Aufbaus bei dem die Kameras so gedreht wurden, daß  $P_1$  parallaxenfrei abgebildet wird.

In Abbildung 3.6 ist ein Stereosystem zu sehen, welches die gleichen Ausdehnungen wie das Fokussystem aus Abbildung 3.5 hat. Der Abstand der linken und rechten Kamera voneinander ist genau der Durchmesser der Blende des Fokussystems, und auch der Schirmabstand ist der gleiche. Gegenüber der bisher betrachteten Stereoanordnung wurden die Kameras so gedreht, daß die Parallaxe für den Punkt  $P_1$  Null ist, genau so wie bei dem Fokussystem aus Teilbild b die Bildpunkte von  $P_1$  zusammenfallen, d.h. den Abstand Null haben. Unter der Annahme, daß die Stereobasis  $b$  viel kleiner als die Abstände  $Z_1$  und  $Z_2$

der Objektpunkte ist, läßt sich die Parallaxe näherungsweise berechnen. Dazu wendet man den näherungsweise geltenden Strahlensatz für die Hilfsgröße  $\tilde{X}$  zweimal an:

$$\frac{x_l}{s} \approx \frac{\tilde{X}}{Z_2} \quad \text{für } b \ll Z_{1,2} \quad (3.13)$$

$$\frac{\tilde{X}}{Z_1 - Z_2} \approx \frac{b/2}{Z_1} \quad \text{für } b \ll Z_{1,2} \quad (3.14)$$

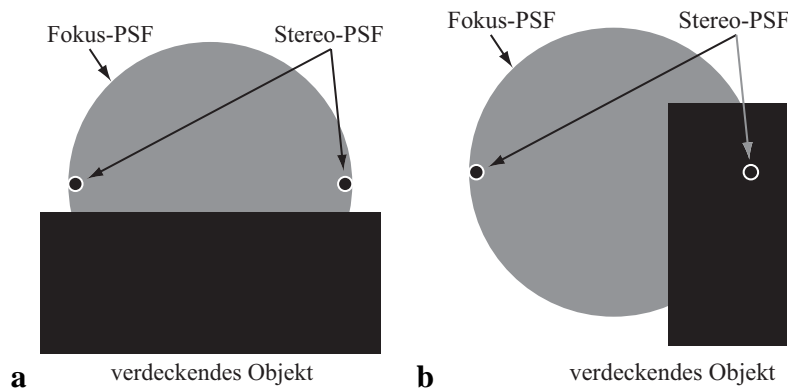
Ineinander eingesetzt ergibt sich für  $x_l$ :

$$x_l = \frac{b}{2} \left[ \frac{s}{Z_2} + \frac{f-s}{f} \right] \quad (3.15)$$

Und für die Parallaxe:

$$p = |x_r - x_l| = b \left[ \frac{s}{Z_2} + \frac{f-s}{f} \right] \quad (3.16)$$

Man erhält also für die Parallaxe  $p$  eines Stereosystems den gleichen Wert, wie für den Durchmesser des Unschärfekreises eines gleich dimensionierten Fokussystems. Als Konsequenz daraus kann man festhalten, daß es keinen grundlegenden Unterschied zwischen Depth-From-Focus-Verfahren und Stereo-Verfahren gibt und damit deren Tiefenauflösung gleich gut ist. Allerdings sind die Linsendurchmesser und damit die Blendendurchmesser im Bereich von 1 cm, während die Basislängen von Stereosystemen im Allgemeinen ein bis zwei Größenordnungen darüber liegen, woraus sich ein entsprechend höhere Auflösung der Tiefeninformation ergibt.



**Abbildung 3.7:** Skizze der 2D-PSF eines Fokus- und eines Stereosystems. **a** Nur die Fokus-PSF wird durch ein verdeckendes Objekt gestört, die Messung des Unschärferadius bleibt aber möglich. **b** Wird die Stereo-PSF verdeckt, ist die Bestimmung der Parallaxe unmöglich.

Häufig wird davon ausgegangen, daß bei Depth-From-Focus-Verfahren, da sie monookular d.h. mit nur einer Optik und Kamera messen, im Gegensatz zu Stereo-Verfahren mit zwei

oder mehreren Kameras, keine Verdeckungen auftreten. Da die Verfahren auf dem gleichen physikalischen Prinzip basieren, kann das so nicht richtig sein. Im Allgemeinen treten bei Depth-From-Focus-Verfahren sogar häufiger Verdeckungen auf, allerdings sind deren Auswirkungen meist nicht so gravierend. Um sich diesen Sachverhalt zu veranschaulichen, kann man die PSF zweier Systeme betrachten. Zum einen ein Stereosystem, bei dem die Kameras so gegeneinander verdreht wurden, daß das weiter vorne liegende, verdeckende Objekt parallaxenfrei abgebildet wird und zum anderen ein Fokussystem gleicher Dimension<sup>1</sup>, welches auf eben dieses verdeckende Objekt fokussiert ist. In Abbildung 3.7a ist die 2D-Punktantwort des Fokussystems und des gesamten Stereosystems zusammen mit dem scharf abgebildeten verdeckenden Objekt skizziert. Es ist klar zu erkennen, daß Teile der ausgedehnten PSF des Fokussystems verdeckt und damit gestört wurden, während die PSF des Stereosystems, die ja nur aus zwei Punkten besteht, noch unverdeckt erscheint. Die flächige Fokus-PSF wird natürlich häufiger von weiter vorne liegenden Objekt verdeckt, allerdings ist es trotzdem noch möglich den Radius der Fokus-PSF, wenn auch fehlerbehaftet, zu messen, solange die Überdeckung weniger als 50 % der Fläche ausmacht. In Teilbild b ist das verdeckende Objekt nun so angeordnet, daß ein Punkt der Stereo-PSF vollständig verschwindet. Die Parallaxe kann hier nicht mehr bestimmt werden.

## 3.2 Laufzeitverfahren

Die Laufzeitverfahren basieren auf einem völlig anderen physikalischen Grundprinzip. Sie messen, wie der Name schon andeutet, die Zeit  $t$ , die ein Signal vom Sensor zum Objektpunkt und zurück benötigt. Die Entfernung  $Z$  zu diesem Objektpunkt ergibt sich dann zu

$$Z = ct/2 \quad (3.17)$$

wobei  $c$  die Geschwindigkeit des benutzten Signals ist. Die Hauptunterschiede zu Triangulationsverfahren sind die Unabhängigkeit des Fehlers in der Tiefenbestimmung von dem Objektabstand

$$\Delta Z = c/2\Delta t \quad (3.18)$$

und die Tatsache, daß wirklich koaxial, d.h. entlang der optischen Achse, gemessen wird. Dadurch ist es z.B. möglich, auch in schmalen Bohrlöchern die Tiefe zu messen.

---

<sup>1</sup>D.h. der Blendendurchmesser entspricht der Basislänge des Stereosystems.

### 3.2.1 Laufzeitverfahren mit moduliertem Licht

Die eine Gruppe der Laufzeitverfahren arbeitet mit moduliertem inkohärentem Licht. Dazu wird von einem Sender moduliertes Licht ausgesandt und der Empfänger mißt die Laufzeit, in dem er das empfangene Signal mit einem entsprechenden Referenzsignal korreliert. Die Signalgeschwindigkeit  $c$  entspricht dabei der Gruppengeschwindigkeit des ausgesandten modulierten Lichts, also  $300 \mu/ps$ . Diese hohe Geschwindigkeit erfordert eine sehr hohe Zeitauflösung der verwandten Meßtechnik.

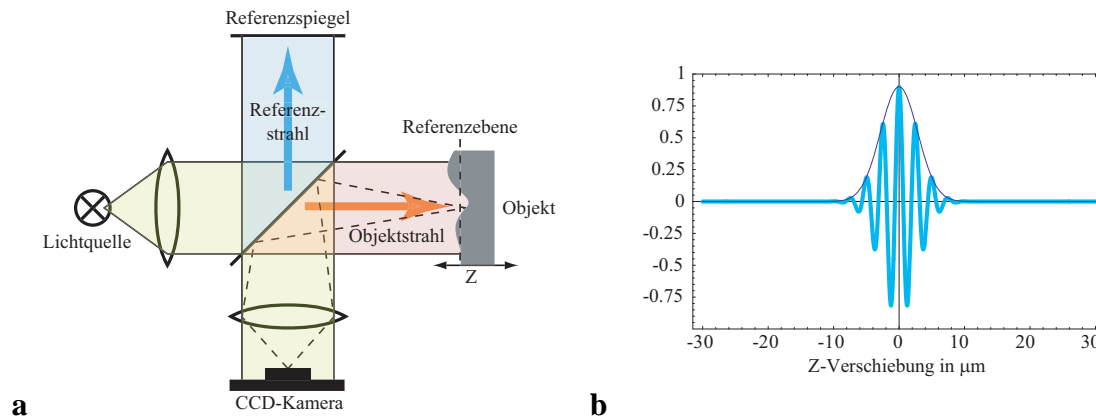
Es gibt nun unterschiedliche Möglichkeiten der Modulation des Lichts. Bei der *Pulsmodulation* wird ein Lichtimpuls ausgesandt und direkt die Zeit bis zu seinem Echo gemessen. Dagegen kann auch ein *kontinuierlich moduliertes Signal* benutzt werden, die Laufzeit wird dann durch Überlagerung mit Signal gleicher oder aber auch unterschiedlicher Frequenz aus der Phasenverschiebung ermittelt. Der meßbare Tiefenbereich ist aufgrund der nur im Bereich von einer halben Wellenlänge eindeutigen Phasenverschiebung recht klein. Er kann allerdings durch die Verwendung mehrerer Messungen mit unterschiedlichen Modulationsfrequenzen erweitert werden. Moduliert man allerdings mit einem *Pseudo-Rauschen*, kann dieses Problem umgangen werden. Der Meßbereich ergibt sich dabei aus der Wiederhollänge des Rauschgenerators.

Die Hauptquellen für Meßungenauigkeiten liegen in den Verstärkern und in der anschließenden Mischung und Korrelation von dem Reflektierten und dem Referenzsignal in der Elektronik. Neuere Entwicklungen, wie z.B. der Photomischdetektor (PMD) [Schwarte et al., 1998], versuchen dies komplett im Empfänger vorzunehmen und eliminieren damit diese Fehlerquellen.

### 3.2.2 Interferometrie

Die Interferometrie arbeitet mit kohärentem Licht, das ähnlich wie in einem Michelson-Interferometer aufgespaltet wird. Der eine Strahl läuft, wie in Abbildung 3.8a zu sehen, zu dem zu messenden Objekt, während der andere als Referenzstrahl dient. Referenz- und Objektstrahl werden überlagert und ihre Phasendifferenz wird gemessen. Da diese nur im Bereich einer halben Wellenlänge eindeutig ist, läßt sich die Objektentfernung nur in einem sehr kleinen Bereich absolut bestimmen, wogegen man die erreichbare Genauigkeit bis zu einer Tausendstel Wellenlänge steigern kann.

Der begrenzte Meßbereich läßt sich durch die Verwendung von mehreren Wellenlängen erweitern. Man überlagert dabei zwei eng benachbarte Wellenlängen, so daß sich eine Schwebung mit einer um Größenordnungen größeren Wellenlänge ergibt, durch die nun



**Abbildung 3.8:** **a** Aufbau zur Entfernungsmessung mittels Interferometrie. In dem einem Arm des Michelson-Interferometers wurde der Spiegel gegen das zu messende Objekt ausgewechselt. Für die Weißlicht-Interferometrie muß, wie mit  $\leftarrow Z \rightarrow$  angedeutet, in Z-Richtung gescannt werden. **b** Für die Weißlicht-Interferometrie ist der Signalverlauf für einen Bildpunkt über die Entfernung Z von der Referenzebene skizziert. In der Nähe der Referenzebene ist das Interferenzmuster zu sehen. Das Maximum der Einhüllenden legt den Objektabstand fest.

der Meßbereich bestimmt wird.

Die Weißlicht-Interferometrie nutzt aus, daß breitbandiges Licht eine sehr kurze Kohärenzlänge hat. Interferenzen treten nur in diesem kleinen Bereich von einigen Mikrometern für weißes Licht auf. Verschiebt man nun das zu messende Objekt in Z-Richtung, treten die Interferenzen immer nur in den Bereichen des Objekts auf, die gerade annähernd die gleiche Entfernung wie der Referenzspiegel haben. Man detektiert im aufgenommenen Bildstapel also die Interferenzerscheinungen, d.h. schnelle hell-dunkel-Wechsel, die mit der Entfernungsänderung auftreten, wie in Abbildung 3.8b skizziert, und mißt die Position ihrer Einhüllenden und damit die Objektentfernung. Eine ausführliche Beschreibung des Verfahrens ist in Dresler et al. [1992] zu finden.



# Kapitel 4

## Depth from Focus

In diesem Kapitel soll ein Überblick über die Vielzahl von in der Literatur beschriebenen Depth-from-Focus-Verfahren gegeben werden. Auch diese lassen sich, wie schon in Kapitel 3 beschrieben, in Untergruppen unterteilen, z.B. anhand der Anzahl der verwandten Bilder; dies hat nämlich einen wesentlichen Einfluß auf die Algorithmik und das Vorwissen, welches hinzugenommen werden muß. Weiterhin gibt es natürlich auch hier wieder aktive und passive Verfahren. Die grundlegenden Ideen zu diesen Verfahren wurden in den 80er Jahren von Jarvis [1983], Pentland [1987], Grossmann [1987] und Krotkov [1987] erarbeitet.

### 4.1 Ein-Bild-Verfahren

Wie in Kapitel 2 ausführlich dargelegt, kann die Bildentstehung in zwei Schritte aufgeteilt werden. Zuerst wird die 3D-Szene mittels Zentralprojektion auf den Schirm abgebildet und anschließend das dort entstandene, ungestörte Bild  $g'_i(\mathbf{x})$  mit der 2D-Punktantwort  $h(\mathbf{x}, z)$  des optischen Systems gefaltet:

$$g_i(\mathbf{x}) = h(\mathbf{x}, z) * g'_i(\mathbf{x}) \quad (4.1)$$

Dabei wurde für die PSF  $h(\mathbf{x}, z)$  geschrieben, um deutlich zu machen, daß die 2D-PSF von der Entfernung  $Z$  des jeweiligen Objektpunkts abhängig ist.

Wie aus obiger Gleichung hervorgeht, können die im resultierenden Bild  $g_i(\mathbf{x})$  enthaltenen Strukturen also gleichermaßen von der Grauwertverteilung des abgebildeten Objekts oder von der PSF herrühren. Das Gleiche gilt für die "Schärfe" bzw. "Unschärfe", die ein Verfahren zur Tiefenrekonstruktion messen muß. Daher ist es klar, daß Verfahren, die nur ein

Bild zur Tiefengewinnung heranziehen, zusätzlich a-priori-Wissen einbeziehen müssen. Pentland schlug in [Pentland, 1987] unter anderem vor, die "Unschärfe" bzw. den Radius des Unschärfekreises in der Nähe von scharfen Kanten zu messen. Da sich in diesen Bildbereichen der Grauwert des Objekts nahezu sprunghaft ändert, rührt die "Unschärfe" allein von der PSF her, die er als gaußförmig annahm. Unter diesen Annahmen konnte er die Standardabweichung der gaußförmigen PSF aus dem Bild berechnen und daraus die Entfernung bestimmen. Grossmann [1987] stellte ein ähnliches Verfahren vor. Dabei wurden zuerst die Kanten im Bild und deren Orientierung bestimmt und dann aus dem an diesen Stellen berechneten Unschärfemaß auf die Tiefeninformation geschlossen. Als Unschärfemaß benutzte er die Halbwertsbreite des Peak der ersten Ableitung der Grauwerte senkrecht zu der Kante.

Beide Verfahren stellten spezielle Anforderungen an die betrachteten Objekte bzw. Szenen: Sie müssen scharfe Kanten enthalten. Dies ist natürlich nicht immer der Fall, so daß der Einsatzbereich dieser Verfahren entsprechend eingeschränkt ist. Allgemein gilt dies für alle Ein-Bild-Verfahren, da sie immer recht genaue Annahmen über die Objekte benötigen, um gemessene "Unschärfe" korrekt der PSF oder dem Objekt selbst zuzuordnen. Trotzdem gibt es gerade bei speziellen Anwendungen mit bekannten Eigenschaften der betrachteten Objekte interessante Einsatzmöglichkeiten für diese Verfahren. So maß z.B. Scholz [1995] die Zellkonzentrationen mit einem speziell angepaßten Ein-Bild-Verfahren und auch Geißler [1998] benutzte ein solches Verfahren zur Bestimmung der Größenverteilung von Blasen.

## 4.2 Zwei-Bild-Verfahren / Depth-from-Defocus

Neben dem oben erwähnten Ein-Bild-Verfahren beschrieb Pentland im gleichen Artikel [Pentland, 1987] auch ein Zwei-Bild-Verfahren, auch *Depth-From-Defocus*-Verfahren genannt. Er ging von zwei Bildern  $g_1(\mathbf{x})$  und  $g_2(\mathbf{x})$  derselben Szene aus, die mit unterschiedlichen Blendenzahlen aufgenommen wurden. Wie schon in Kapitel 2 erläutert, gehen die beiden Bilder  $g_1$  und  $g_2$  aus demselben ungestörten Bild  $g'$  durch Faltung mit der PSF des optischen Systems folgendermaßen hervor:<sup>1</sup>

$$g_1(\mathbf{x}) = h_1(\mathbf{x}, Z) * g'(\mathbf{x}) \quad (4.2)$$

$$g_2(\mathbf{x}) = h_2(\mathbf{x}, Z) * g'(\mathbf{x}) \quad (4.3)$$

Da die Bilder mit unterschiedlichen Blenden aufgenommen wurden, sind auch die PSFs  $h_1$  und  $h_2$  unterschiedlich. Pentland nahm diese PSFs als gaußförmig mit verschiedenen

<sup>1</sup>Der Index  $i$ , der anzeigt, daß es sich um die Grauwertverteilung im Bildraum handelt, wurde aus Gründen der Übersichtlichkeit weggelassen.

Standardabweichungen  $\sigma_1$  und  $\sigma_2$  an:

$$h_{1,2}(\mathbf{x}, Z) = \frac{1}{2\pi\sigma_{1,2}(Z(\mathbf{x}))^2} \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_{1,2}(Z(\mathbf{x}))^2}\right) \quad (4.4)$$

Die Standardabweichungen wurden dabei als Funktion der Entfernung  $Z(\mathbf{x})$  des zum jeweiligen Bildpunkt gehörenden Objektpunkts geschrieben, um zu verdeutlichen, daß sie davon abhängig sind.

Betrachtet man nun eine genügend kleine kreisförmige Umgebung eines Bildpunktes, kann man diese Abhängigkeit vernachlässigen und in Kreiskoordinaten  $r$  und  $\theta$  schreiben:

$$g_1(r, \theta) = \frac{1}{2\pi\sigma_1^2} \exp\left(-\frac{r^2}{2\sigma_1^2}\right) * g'(r, \theta) \quad (4.5)$$

$$g_2(r, \theta) = \frac{1}{2\pi\sigma_2^2} \exp\left(-\frac{r^2}{2\sigma_2^2}\right) * g'(r, \theta) \quad (4.6)$$

In diese Umgebung kann man nun die Faltung als komplexe Multiplikation im Fourierraum darstellen. Mit den Fouriertransformationspaaren

$$g_{1,2}(r, \theta) \xleftrightarrow{\mathcal{F}} \hat{g}_{1,2}(\lambda, \theta) \quad (4.7)$$

$$g'(r, \theta) \xleftrightarrow{\mathcal{F}} \hat{g}'(\lambda, \theta) \quad (4.8)$$

$$\frac{1}{2\pi\sigma_{1,2}^2} \exp\left(-\frac{r^2}{2\sigma_{1,2}^2}\right) \xleftrightarrow{\mathcal{F}} \frac{1}{2\pi\sigma_{1,2}^2} \exp\left(-\frac{1}{2}\lambda^2\sigma_{1,2}^2\right) \quad (4.9)$$

ergibt sich

$$\hat{g}_{1,2}(\lambda, \theta) = \frac{1}{2\pi\sigma_1^2} \exp\left(-\frac{1}{2}\lambda^2\sigma_{1,2}^2\right) \hat{g}'(\lambda, \theta) \quad (4.10)$$

Integriert man über  $\theta$  und bildet den Quotienten aus  $\hat{g}_1$  und  $\hat{g}_2$ , so kann man  $\hat{g}'$  kürzen und erhält:

$$\frac{\hat{g}_1(\lambda)}{\hat{g}_2(\lambda)} = \exp\left(\lambda^2 2\pi^2(\sigma_2^2 - \sigma_1^2)\right) \quad \text{mit} \quad \hat{g}_{1,2}(\lambda) = \int_0^{2\pi} \hat{g}_{1,2}(\lambda, \theta) d\theta \quad (4.11)$$

Bildet man nun noch den natürlichen Logarithmus ergibt sich:

$$\underbrace{\lambda^2 2\pi^2(\sigma_2^2 - \sigma_1^2)}_A = \underbrace{\ln \hat{g}_1(\lambda) - \ln \hat{g}_2(\lambda)}_{B(\lambda)} \quad (4.12)$$

Fittet man  $A\lambda^2$  an die aus dem Bild gewonnenen  $B(\lambda)$  an, läßt sich daraus  $A$  bestimmen. Wird dann noch das eine Bild mit einer sehr großen Blendenzahl bzw. einem kleinen Blendenzahl aufgenommen (Lochkamera), so ist  $\sigma_1 = 0$  und aus  $A$  folgt direkt  $\sigma_2$  und damit

der Objektstand  $Z$ . Es ist aber auch möglich, drei oder mehr Bilder aufzunehmen und das sich daraus ergebende Gleichungssystem

$$A_{ij} = 2\pi^2(\sigma_i^2 - \sigma_j^2) \quad (4.13)$$

für jedes  $\sigma_i$  zu lösen.

Dieses ist die grundlegende Vorgehensweise aller Depth-Form-Focus-Verfahren. Allerdings wurden von verschiedenen Autoren verbesserte Verfahren vorgeschlagen, die vor allem folgende Probleme bzw. Einschränkungen angehen:

- Aufnahme eines Bildes mit einer sehr kleinen Blendenöffnung (**Lochkamera**), wofür man entweder eine starke Beleuchtung der Szene oder eine empfindliche Kamera benötigt.
- Einschränkung auf **gaußförmige Punktantwort**, was nicht immer gegeben ist und daher zu Meßfehlern führen kann.
- **Rauschanfälligkeit** des Verfahrens.
- Fehler können sich auch aus der Wahl der **Größe der Umgebung** ergeben, die zur Berechnung der Entfernung herangezogen wird. Zum einen geht man davon aus, daß die Entfernung innerhalb der Umgebung konstant ist, was mit steigender Umgebungsgröße immer weniger erfüllt sein wird. Zum anderen sollte die Umgebung möglichst groß gewählt werden, um Ungenauigkeiten bei der Fouriertransformation zu vermeiden.

Ens schlug in [Ens and Lawrence, 1993] zwei Verfahren vor, die sowohl mit anderen als auch mit gaußförmigen Punktantworten arbeiteten und auch nicht darauf angewiesen waren, daß ein Bild mit einer "Lochkamera" aufgenommen wurde. Bei dem einen Verfahren handelt es sich um eine regularisierte inverse Filterung. Die beiden Bilder  $g_{1,2}$  gehen aus dem ungestörten Bild  $g'$  durch Faltung mit den Punktantworten  $h_1$  und  $h_2$  hervor, wobei er davon ausging, daß  $h_2$  die größere Unschärfe hervor rufen soll.

$$g_1(\mathbf{x}) = h_1(\mathbf{x}) * g'(\mathbf{x}) \quad (4.14)$$

$$g_2(\mathbf{x}) = h_2(\mathbf{x}) * g'(\mathbf{x}) \quad (4.15)$$

Dann muß es einen weiteren Filter  $h_3$  geben, so daß er angewandt auf das "schärfere" Bild  $g_1$  das "unschärfere" Bild  $g_2$  ergibt:

$$g_2(\mathbf{x}) = h_3(\mathbf{x}) * g_1(\mathbf{x}) \quad (4.16)$$

Daraus folgt mit Gleichung (4.14):

$$h_2(\mathbf{x}) = h_3(\mathbf{x}) * h_1(\mathbf{x}) \quad (4.17)$$

In einer kleinen Umgebung sind diese konstant und man kann daher im Fourierraum schreiben:

$$\frac{\hat{g}_2(\mathbf{k})}{\hat{g}_1(\mathbf{k})} = \frac{\hat{h}_2(\mathbf{k}) \cdot \hat{g}'(\mathbf{k})}{\hat{h}_1(\mathbf{k}) \cdot \hat{g}'(\mathbf{k})} = \frac{\hat{h}_2(\mathbf{k})}{\hat{h}_1(\mathbf{k})} = \hat{h}_3(\mathbf{k}) \quad (4.18)$$

Dies ist die einfache inverse Filterung, wie sie in ähnlicher Form für gaußförmige Punkantworten von Pentland benutzt wurde. Ens nahm für  $\hat{h}_3$  eine näherungsweise quadratische Form an und minimierte das folgende Funktional:

$$\|\hat{h}_3(\mathbf{k}) \cdot \hat{g}_1(\mathbf{k}) - \hat{g}_2(\mathbf{k})\|^2 + \lambda \|C(\hat{h}_3)\|^2 \rightarrow \text{Minimum} \quad (4.19)$$

Dabei ist der erste Term der sogenannte Datenterm, der dafür sorgen soll, daß  $h_3$  Gleichung (4.16) erfüllt, d.h. zu den Daten passt. Die Funktion  $C(\cdot)$  im zweiten Term wird Null, wenn  $\hat{h}_3$  eine quadratische Form annimmt, so daß der zweite Term dafür sorgt, daß  $\hat{h}_3$  eine möglichst quadratische Form hat. Der Faktor  $\lambda$  erlaubt eine Gewichtung zwischen den beiden Termen.

Bei dem zweiten Verfahren handelt es sich um einen matrixbasierten Ansatz, der mit den Bilddaten im Ortsraum arbeitet. Dazu werden wiederum kleine Umgebungen um den jeweils betrachteten Punkt so in Matrix- bzw. Vektorform geschrieben, daß sich Faltungen als Matrix-Vektor-Produkt schreiben lassen. Gleichung (4.16) ergibt in dieser Notation dann:

$$\mathbf{g}_{2S} = \mathbf{g}_{1BT} \mathbf{h}_{3S} \quad (4.20)$$

Dabei steht der Index S für Vektoren der Länge  $N^2$ , die aus den hintereinander gereihten Zeilen des jeweiligen  $N \times N$  Bildes aufgebaut wurden. Der Index BT steht  $N^2 \times N^2$  block Toeplitz Matrizen aus dem entsprechenden  $N \times N$  Bild. Für den genauen Aufbau dieser Matrizen sei z.B. auf [Geißler and Dierig, 1999] verwiesen. Aus (4.20) könnte man direkt  $\mathbf{h}_{3S}$  berechnen:

$$\mathbf{h}_{3S} = \mathbf{g}_{1BT}^{-1} \mathbf{g}_{2S} \quad (4.21)$$

Allerdings wird dieser Ansatz instabil, sobald die Bilddaten verrauscht sind, weshalb Ens und Lawrence einen Regularisierungsansatz ähnlich zu (4.19) benutzten.

$$\|\mathbf{g}_{1BT} \cdot \mathbf{h}_{3S} - \mathbf{g}_{2S}\|^2 + \lambda \|\mathbf{C} \cdot \mathbf{h}_{3S}\|^2 \rightarrow \text{Minimum} \quad (4.22)$$

Wieder sorgt der erste Term dafür, daß (4.20) möglichst gut erfüllt wird, während der zweite Term Abweichungen von  $\mathbf{h}_{3S}$  von einer vorgegebenen Familie von Filtern bestraft. Im Gegensatz zu (4.19) findet die Regularisierung nun im Ortsraum statt. Da es im Allgemeinen schwierig ist die Matrix  $\mathbf{C}$  zu bestimmen, schlugen Ens und Lawrence vor, eine

Lookup-Tabelle von für jede Objektentfernung vorberechneten  $h_3(\mathbf{x})$  anzulegen und anschließend jenes  $\tilde{h}_3(\mathbf{x})$  aus der Tabelle zu suchen, welches

$$\sum_{\mathbf{x}} \left[ \tilde{h}_3(\mathbf{x}) g_1(\mathbf{x}) - g_2(\mathbf{x}) \right]^2 \rightarrow \text{Minimum} \quad (4.23)$$

minimiert.

Alle bisherigen Verfahren gingen davon aus, daß die Defokussierung zumindest in einer kleinen Umgebung um den betrachteten Punkt als linear und verschiebungsinvariant angenommen werden kann. Wie in Kapitel 2 beschrieben, handelt es sich bei optischen Systemen zwar um lineare, aber nicht um verschiebungsinvariante Systeme, denn die PSF hängt von der Entfernung des jeweiligen Objektpunkts ab. Um den dadurch entstehenden Fehler zu vermeiden, modellierten Rajagopalan und Chaudhuri die Unschärfe durch Defokussierung in verschiedenen verschiebungsvarianten Ansätzen. Sie benutzen zur Orts-Frequenz-Darstellung das komplexe Spektrogramm bzw. die Wigner-Verteilung, die sich wie folgt berechnen lassen

$$C_g(k_x, x) = \int_{-\infty}^{\infty} g(x') u(x' - x) e^{-ik_x x'} dx' \quad (4.24)$$

$$W_g(k_x, x) = \int_{-\infty}^{\infty} g\left(x + \frac{x'}{2}\right) g^*\left(x - \frac{x'}{2}\right) e^{-ik_x x'} dx' \quad (4.25)$$

wobei  $u$  eine geeignete Fensterfunktion und  $g^*$  das komplex Konjugierte von  $g$  ist. Um die Schreibweise zu vereinfachen, wurde hier nur eine Dimension angeschrieben. Beide Beschreibungen enthalten zu jedem Ort  $x$  ein komplettes Spektrum. Die Faltung des ungestörten Bildes mit der ortsabhängigen Punktantwort des optischen Systems (siehe z.B. Gleichung (4.14)), kann nun im Orts-Frequenz-Raum als komplexe Multiplikation mit einer ortsabhängigen optischen Transferfunktion  $H_{1,2}$  geschrieben werden und man erhält für die komplexen Spektrogramme der Bilder  $g_1$  und  $g_2$ :

$$C_{g_{1,2}}(k_x, x) = C_{g'}(k_x, x) H_{1,2}(k_x, x) \quad (4.26)$$

Daraus ergibt sich für den direkten Zusammenhang von  $C_{g_1}$  und  $C_{g_2}$ :

$$C_{g_2}(k_x, x) = C_{g_1}(k_x, x) H(k_x, x) \quad \text{mit} \quad H(k_x, x) = \frac{H_2(k_x, x)}{H_1(k_x, x)} \quad (4.27)$$

Für gaußförmige Transferfunktionen  $H_{1,2}(k_x, x) = \exp(-k_x^2 \sigma_{1,2}^2(x))$  erhält man für  $H$ :

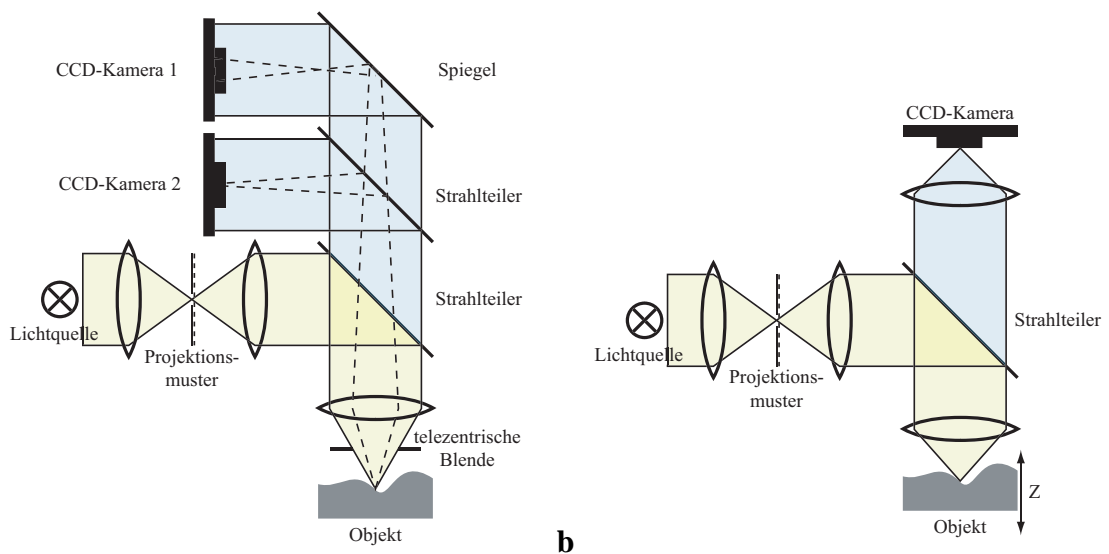
$$H(k_x, x) = \exp(-k_x^2 s(x)) \quad \text{mit} \quad s(x) = \sigma_2^2(x) - \sigma_1^2(x) \quad (4.28)$$

Da bei gegebenem  $C_{g_1}$  und  $H$  nicht unbedingt auch  $C_{g_2}$  eine gültige Orts-Frequenz-Darstellung sein muß, erhält man die folgende näherungsweise Lösung für  $s(x)$ :

$$\int_{-\infty}^{\infty} \left[ |C_{g_2}(k_x, x)|^2 - |C_{g_1}(k_x, x)|^2 \exp(-k_x^2 s(x)) \right]^2 dk_x \rightarrow \min \quad (4.29)$$

Hat man damit  $s(x)$  für jeden Bildpunkt bestimmt, läßt sich die Entfernung am jeweils entsprechenden Objektpunkt berechnen.

Im Rahmen eines Regularisierungsansatzes stellten Rajagopalan and Chaudhuri [1993] räumliche Glattheitsbedingungen an den Unschärfeparameter  $s(x)$  (i. allg. die Halbwertsbreite der PSF), während Ens and Lawrence [1993] Bedingungen an die Form der PSF bzw. OTF stellten. In [Rajagopalan and Chaudhuri, 1999] modellierten sie den ortsabhängigen Unschärfeparameter als Markov-Random-Field. Eine Zusammenfassung ihrer Arbeiten ist in [Chaudhuri and Rajagopalan, 1999] zu finden.



**Abbildung 4.1:** **a** Skizze des von Nayar et al. [1996] vorgestellten aktiven Depth-From-Defocus-Systems zur Echtzeit-Tiefenmessung. In denselben Strahlengang wird die strukturierte Beleuchtung sowie die Kameras mit unterschiedlichen Fokuseinstellungen eingekoppelt. **b** Skizze des von Scheuermann et al. [1995] vorgestellten konfokalen Mikroskops zur Tiefenmessung mittels aktiven Depth-From-Focus.

Neben den oben beschriebenen passiven Verfahren wurden auch aktive Verfahren entwickelt. So stellte Nayar et al. [1996] ein System vor, welches in Video-Echtzeit Tiefeninformationen liefert. Dabei wurde wie bei einem konfokalen Mikroskop durch die gleiche Optik beleuchtet, durch die auch die Bilder aufgenommen wurden, wie in Abbildung 4.1a skizziert. Während konfokale Mikroskope im Allgemeinen zumindest in Z-Richtung abtasten, wurde hier der Strahlengang mit einem Strahlteiler auf zwei Kameras mit unterschiedlichen Fokuseinstellungen geleitet<sup>2</sup>. Die verwendete Kamera-Optiken waren bildseitig telezentrisch (s. Abschnitt 2.3) ausgelegt, so daß die beiden Bilder trotz verschiedener

<sup>2</sup>Dies ist auch daher möglich, da mit Mikroskopen meist echte 3D-Volumen aufgenommen werden, während der hier vorgestellte Sensor Oberflächen in einem 3D-Raum vermessen sollte.

Fokuseinstellungen gleich skaliert und damit registriert waren. Mit einem auf das zur Beleuchtung verwendete schachbrettartige Muster optimierten Laplace-Filter wurde die Unschärfe in den beiden Bildern detektiert und daraus die Objektentfernung berechnet.

Aus der aktiven Beleuchtung der Szene ergeben sich zwei Vorteile: Zum einen ist man durch die Beleuchtung relativ unabhängig von der Textur in der Szene selbst; zum anderen unterliegt die Beleuchtung, da sie durch die gleiche Optik erfolgt, ebenfalls der PSF, d.h. nicht fokussierte Szenenbereiche werden nicht nur unscharf abgebildet, sondern auch noch unscharf beleuchtet. Durch diese doppelte Unschärfe, ergibt sich eine erhöhte Tiefenaufklärung.

### 4.3 Mehr-Bild-Verfahren / Depth-from-Focus

Die Depth-From-Focus-Verfahren benötigen eine Reihe von Bildern mit variierenden Kameraparametern, z.B. Blendenradius oder Schirmabstand. Sie berechnen anschließend in jedem Bild ein Maß für die Unschärfe und suchen für jeden Punkt dasjenige Bild, in dem der Punkt mit der geringsten Unschärfe abgebildet wurde. Aus den zu diesem Bild gehörenden Kameraparametern läßt sich dann die Objektentfernung berechnen. Da diese Verfahren im Allgemeinen deutlich mehr als zwei Bilder benötigen, sind sie recht langsam und daher für bewegte Objekte nicht geeignet. Allerdings müssen keine speziellen Annahmen über die Form der Punktantwort gemacht werden und sie benötigen, außer dem bei passiven Verfahren immer nötigen ausreichenden Kontrast, auch keine speziellen Szeneneigenschaften, wie z.B. scharfe Kanten. Lediglich das benutzte Schärfemaß muß ein Extremum bei den Kameraparametern annehmen, die den jeweiligen Objektpunkt scharf abbilden. Meist wurden diese Verfahren im Zusammenhang mit Autofokussystemen untersucht. So stellt Jarvis [1983] ein solches Verfahren erstmals vor. Während Krotkov [1987] verschiedene Schärfemaße auf deren Eignung für ein Autofokussystem untersuchte, schlugen Darell and Wohn [1990] zur schnellen Berechnung ein Schärfemaß auf Basis von Gauß- und Laplace-Pyramiden vor. Und auch Subbarao and Tyan [1998] untersuchten die Eigenschaften verschiedener Schärfemaße auf deren Verwendung in Autofokussystemen hin.

Aktive Verfahren wurden von Scheuermann et al. [1995] für die Mikroskopie vorgestellt und in [Scheuermann et al., 1999] weiterentwickelt. Wie in Abbildung 4.1b skizziert, handelt es sich dabei um ein konvokales Mikroskop, bei dem zur Beleuchtung ein statistisches Rausch- und später ein Schachbrett-Muster benutzt wurde. Wie bei [Nayar et al., 1996] erfolgte die Beleuchtung durch die gleiche Optik, so daß die Punktantwort einmal für die Beleuchtung und einmal für die Bildaufnahme angewendet werden muß, wodurch die Tiefenaufklärung noch gesteigert wird.



# Kapitel 5

## Lokale Parameterschätzung

In diesem Kapitel sollen die Grundlagen einer lokalen Parameterschätzmethode vorgestellt werden. Ausgehend vom Begriff des Optischen Fußes wird in Abschnitt 5.1 die Kontinuitätsgleichung der Grauwerte vorgestellt und mit der Methode der kleinsten Quadrate gelöst. In Abschnitt 5.2 ergibt sich über einen total-least-squares-Ansatz der Strukturtenor, dessen Eigenvektor zum kleinsten Eigenwert die Lösung der Kontinuitätsgleichung der Grauwerte darstellt. Verschiedene Erweiterungen dieser Kontinuitätsgleichung werden in Abschnitt 5.3 vorgestellt und eine davon, die Diffusionsschätzung, ausführlicher diskutiert.

### 5.1 Optischer Fluß

Betrachtet man eine Sequenz von Bildern, in denen eine Bewegung zu sehen ist, so läßt sich diese sichtbare Bewegung durch den *optischen Fluß*  $\mathbf{f} = [f_x, f_y]^T$  beschreiben. Er bezeichnet für jeden Bildpunkt die Geschwindigkeit mit der sich die Grauwerte bewegen. Auch wenn dieser optische Fluß nicht immer der realen Bewegung in der Szene entspricht, so ist er doch das einzige, was an Bewegung im Bild meßbar ist.

In Anlehnung an Kontinuitätsgleichungen aus der Physik kann man auch für Grauwerte eine Kontinuitätsgleichung aufstellen, welche als *brightness change constraint equation* oder kurz *BCCE* bezeichnet wird.

$$\frac{\partial g}{\partial t} + \mathbf{f} \nabla g = 0 \quad (5.1)$$

Sie besagt, daß zeitliche Änderungen des Grauwerts nur auf eine Bewegung einer räumlichen Grauwertstruktur zurückzuführen sind. Dies ist natürlich nur zum Teil richtig, da z.B.

schon Beleuchtungsänderungen zu Grauwertänderungen führen, denen keine Bewegung zu Grunde liegt. Trotzdem läßt sich mit Gleichung (5.1) unter bestimmten Bedingungen (keine Belichtungsänderung, keine Reflexe oder sonstige Störungen) der optische Fluß berechnen. Im eindimensionalen Bereich ist die Lösung einfach:

$$f = -\frac{\partial g}{\partial t} \bigg/ \frac{\partial g}{\partial x} \quad (5.2)$$

In zwei oder mehr Dimensionen ist Gleichung (5.1) unterbestimmt, da der gesuchte Vektor  $\mathbf{f}$  zwei oder mehr Komponenten hat, die Gleichung aber nur eine skalare Bedingung darstellt. Man kann also nur eine Komponente von  $\mathbf{f}$  berechnen und diese ist parallel zum Grauwertgradienten  $\nabla g$ . Dieses Problem wird als *Blendenproblem* bezeichnet. Um diese Einschränkung zu umgehen, müssen mehrere einschränkende Gleichungen gefunden werden, was z.B. durch die Anwendung von (5.1) auf mehrere Komponenten eines Farbbildes möglich ist. Hier ist aber folgende Lösung von größerem Interesse: Man geht davon aus, daß der optische Fluß in einer kleinen Umgebung konstant ist, so daß man für jeden Bildpunkt aus dieser Umgebung Gleichung (5.1) aufstellen kann. Umfaßt die Umgebung  $N > 2$  Bildpunkte, hat man nun ein überbestimmtes Gleichungssystem zu lösen:

$$\begin{bmatrix} \partial_x g(1) & \partial_y g(1) \\ \partial_x g(2) & \partial_y g(2) \\ \vdots & \vdots \\ \partial_x g(N) & \partial_y g(N) \end{bmatrix} \begin{bmatrix} \tilde{f}_x \\ \tilde{f}_y \end{bmatrix} = - \begin{bmatrix} \partial_t g(1) \\ \partial_t g(2) \\ \vdots \\ \partial_t g(N) \end{bmatrix} \quad (5.3)$$

oder

$$\mathbf{G}\tilde{\mathbf{f}} = -\mathbf{g} \quad (5.4)$$

wobei die partiellen Ableitungen mit  $\partial_x g = \partial g / \partial x$  abgekürzt wurden. Zu beachten ist, daß alle  $N$  Gleichungen linear abhängig sein können, wenn z.B. der Grauwertgradient in der gesamten Umgebung gleich ist. Dann tritt wieder das Blendenproblem auf und man kann, wie oben, nur die Komponenten von  $\mathbf{f}$  bestimmen, die parallel zu dem Gradienten sind.

Die Näherungslösung  $\tilde{\mathbf{f}}$  findet man, wenn man die Abweichung von Gleichung (5.4) minimiert:

$$|\mathbf{G}\tilde{\mathbf{f}} + \mathbf{g}|^2 \rightarrow \text{Minimum} \quad (5.5)$$

Mittels der Methode der kleinsten Quadrate erhält man daraus:

$$\tilde{\mathbf{f}} = -(\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{g} \quad (5.6)$$

Ein solches Least-Squares-Verfahren zur Bestimmung des optischen Flusses wurde von Lucas and Kanade [1981] eingeführt und hat sich aufgrund seiner geringen Komplexität und guter Ergebnisse bewährt [Barron et al., 1994; Haußecker and Spies, 1999].

## 5.2 Strukturtensor

Bei der Methode der kleinsten Quadrate wird nur  $\mathbf{g}$  aber nicht  $\mathbf{G}$  als fehlerbehaftet angenommen. Da  $\mathbf{G}$  die partielle Ableitungen der Bilddaten enthält, ist davon auszugehen, daß auch  $\mathbf{G}$  Fehler aufweist. Daher formuliert man die Kontinuitätsgleichung (5.1) im Rahmen einer total least squares Schätzung [Van Huffel and Vandewalle, 1991] wie folgt:

$$\mathbf{d}^T \bar{\mathbf{f}} = 0 \quad (5.7)$$

Der Datenvektor  $\mathbf{d} = [\partial_x g, \partial_y g, \partial_t g]^T$  enthält nun alle datenabhängigen Terme, während  $\bar{\mathbf{f}} = [f_x, f_y, 1]^T$  die zu bestimmenden Komponenten des optischen Flusses enthält. Um diese unterbestimmte Vektorgleichung lösen zu können, geht man davon aus, daß innerhalb einer kleinen Umgebung  $\bar{\mathbf{f}}$  konstant ist. Gleichung (5.7) ist dann näherungsweise für alle Bildpunkt der Umgebung erfüllt und man kann die Minimierungsaufgabe wie folgt formulieren

$$\sum_{i \in U} w_{i-i_0} [(\mathbf{d}_i^T \bar{\mathbf{f}})^2 + \lambda(1 - \bar{\mathbf{f}}^T \bar{\mathbf{f}})] \rightarrow \text{Minimum}, \quad (5.8)$$

wobei  $w_{i-i_0}$  einen Wichtungsfaktor innerhalb der Umgebung darstellt. Der zweite Term in der Summe stellt sicher, daß die triviale Lösung  $\bar{\mathbf{f}} = 0$  vermieden wird, in dem  $\bar{\mathbf{f}}^T \bar{\mathbf{f}} = 1$  erzwungen wird. Die Komponenten des optischen Flusses erhält man dann mit  $f_x = \bar{f}_1 / \bar{f}_3$  und  $f_y = \bar{f}_2 / \bar{f}_3$ . Als Lösung von (5.8) erhält man folgende Eigenwertgleichung, wie in [Haußecker and Spies, 1999] und [Spies, 2001] ausführlich beschrieben wird:

$$\mathbf{J} \bar{\mathbf{f}} = \lambda \bar{\mathbf{f}} \quad \text{mit} \quad \mathbf{J} = \sum_{i \in U} w_{i-i_0} (\mathbf{d}_i \mathbf{d}_i^T) \quad (5.9)$$

Dabei ist  $\mathbf{J}$  ein symmetrischer reeller Tensor. Er wird als *Strukturtensor* bezeichnet, da er die gesamte Information über die lokale Grauwertstruktur des Bildes beinhaltet. Der Eigen-

vektor zum kleinsten Eigenwert minimiert Gleichung (5.8) und stellt damit die gesuchte Lösung dar.

### 5.3 Erweiterter Strukturtensor

Wie in Abschnitt 5.1 erwähnt, ist die Kontinuitätsgleichung der Grauwerte (5.1) nicht immer erfüllt. Daher wurde sie in unterschiedlicher Weise erweitert, um sie den entsprechenden Verhältnissen anzupassen. Es gibt dabei zwei Gruppen von Erweiterungen: Zum einen die geometrischen Erweiterungen [Haußecker et al., 1999; Haußecker and Fleet, 2001; Spies and Jähne, 2001; Garbe, 2001], die die Einschränkung auf lokal konstante Bewegungen aufheben:

**Affine Bewegung:** Ergänzt man die reine Translation mit einer lokale konstanten affinen Transformation [Fleet, 1992], so kann man für  $\mathbf{f}$  schreiben:

$$\bar{\mathbf{f}} = \mathbf{t} + \mathbf{A}\mathbf{x} = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} a_3 & a_4 \\ a_5 & a_5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (5.10)$$

Für den Daten- und den Parametervektor erhält man dann:

$$\mathbf{d} = [\partial_x g \quad \partial_y g \quad x \partial_x g \quad x \partial_y g \quad x \partial_y g \quad y \partial_y g \quad \partial_t g]^T \quad (5.11)$$

$$\bar{\mathbf{f}} = [a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_5 \quad 1]^T \quad (5.12)$$

Zum Anderen wurden zusätzlich Helligkeitsänderungen eingeführt, so zum Beispiel:

**Quell-Terme:** Bei konstanten Helligkeitsänderungen tritt ein zusätzlicher Quell-Term der Stärke  $q$  auf:

$$f_x \partial_x g + f_y \partial_y g + \partial_t g = q \quad \rightarrow \quad \mathbf{d} = [\partial_x g \quad \partial_y g \quad 1 \quad \partial_t g]^T \quad \text{und} \quad \bar{\mathbf{f}} = [f_x \quad f_y \quad -q \quad 1]^T \quad (5.13)$$

Dies ist bei globalen Beleuchtungsänderungen der Szene der Fall, aber auch lokale Quellen und Senken lassen sich z.B. in Infrarot-Bilder beobachten, wenn Objekte erhitzt oder abgekühlt werden, wie von Garbe [2001] untersucht wurde.

**Exponentieller Abfall:** Hierbei ist die Helligkeitsänderung proportional zum Grauwert und zur Zerfallskonstante  $\kappa$ . Dies ist z.B. beim Abklingen von Fluoreszenzen zu beobachten.

$$f_x \partial_x g + f_y \partial_y g + \partial_t g = -\kappa g \quad \rightarrow \quad \mathbf{d} = [\partial_x g \quad \partial_y g \quad g \quad \partial_t g]^T \quad \text{und} \quad \bar{\mathbf{f}} = [f_x \quad f_y \quad \kappa \quad 1]^T \quad (5.14)$$

**Diffusion:** Bei Transportprozessen ist neben der reinen Verschiebung Diffusion zu beobachten. Nach dem 2. Fickschen Gesetz läßt sich mit der Diffusionskonstanten  $D$  dann schreiben:

$$f_x \partial_x g + f_y \partial_y g + \partial_t g = D \Delta g \quad \rightarrow \quad \mathbf{d} = [\partial_x g \quad \partial_y g \quad \Delta g \quad \partial_t g]^T \quad \text{und} \quad \bar{\mathbf{f}} = [f_x \quad f_y \quad -D \quad 1]^T \quad (5.15)$$

Alle diese Erweiterungen können im Rahmen des in Abschnitt 5.2 dargelegten Konzepts berechnet werden, wenn man die Daten- und Parametervektoren entsprechend anpaßt. Hierbei soll aber noch darauf hingewiesen werden, daß es in höheren Dimensionen auch zu einem entsprechend höherdimensionalen Blendenproblem kommt. Ein Teil der Gleichungen wird dabei linear abhängig und die Lösung kann dann nur noch in einem Unterraum durch Singulärwertzerlegung bestimmt werden. Für eine genaue Beschreibung sei auf [Spies, 2001] verwiesen.

### 5.3.1 Diffusions-Schätzung

Im Rahmen dieser Arbeit ist die Erweiterung des Strukturtenors um die Diffusion und die damit verbundene Schätzung der Diffusionskonstanten  $D$  am wichtigsten. Dazu soll zunächst eine Bildsequenz betrachtet werden, die eine gaußförmige Grauwertverteilung enthält, die sich im Laufe der Bildsequenz verbreitert.

$$g(\mathbf{x}, t) = \frac{1}{2\pi\sigma^2(t)} \exp\left(-\frac{(\mathbf{x} - \mathbf{x}_0(t))^2}{2\sigma^2(t)}\right) \quad (5.16)$$

Dabei sind sowohl der Mittelpunkt  $\mathbf{x}_0(t) = \mathbf{v}t$  als auch die Varianz  $\sigma^2(t) = dt$  zeitabhängig. Für die Diffusionsschätzung erhält man, wie oben beschrieben, folgende Kontinuitätsgleichung:

$$f_x \partial_x g + f_y \partial_y g + \partial_t g = D \Delta g \quad (5.17)$$

Nimmt man nun der Einfachheit halber an, daß jegliche Grauwertänderung auf die Diffusion zurückzuführen ist, d.h.  $\mathbf{f} = 0$ , so erhält man für die Diffusionskonstante  $D$ :

$$D = \frac{\partial_t g}{\Delta g} = \frac{\partial_t g}{\partial_x \partial_x g + \partial_y \partial_y g} \quad (5.18)$$

Für die Grauwertverteilung aus (5.16) und mit  $\mathbf{v} = 0$  erhält man folgende partielle Ablei-

tungen:

$$\partial_t g(\mathbf{x}, t) = \left[ -\frac{1}{t} + \frac{x^2 + y^2}{2dt^2} \right] g(\mathbf{x}, t) = \frac{x^2 + y^2 - 2dt}{2dt^2} g(\mathbf{x}, t) \quad (5.19)$$

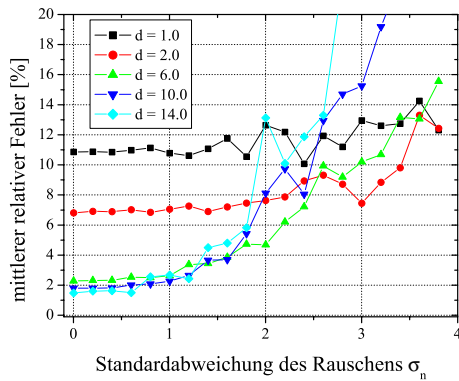
$$\partial_x \partial_x g(x, y, t) = \left[ -\frac{1}{dt} + \left( -\frac{x}{dt} \right)^2 \right] g(\mathbf{x}, t) = \frac{x^2 - dt}{d^2 t^2} g(\mathbf{x}, t) \quad (5.20)$$

$$\partial_y \partial_y g(x, y, t) = \left[ -\frac{1}{dt} + \left( -\frac{y}{dt} \right)^2 \right] g(\mathbf{x}, t) = \frac{y^2 - dt}{d^2 t^2} g(\mathbf{x}, t) \quad (5.21)$$

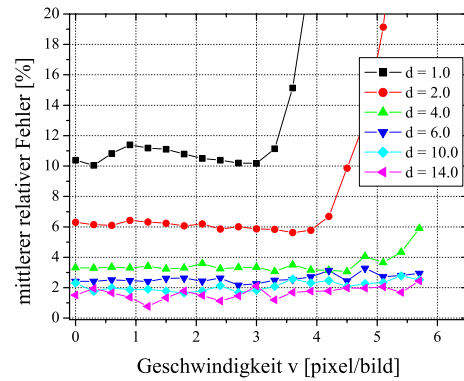
Und damit:

$$D = \frac{\partial_t g}{\partial_x \partial_x g + \partial_y \partial_y g} = \frac{d}{2} \quad (5.22)$$

In Abbildung 5.1a ist der mittlere relative Fehler der geschätzten Diffusionskonstante  $D$  für  $d = 2.0$ ,  $d = 6.0$ ,  $d = 10.0$  und  $d = 14.0$  in Abhängigkeit von der Standardabweichung des Rauschens dargestellt. Der relative Fehler steigt, wie erwartet, mit zunehmendem Rauschen an; für größere Diffusionskonstanten allerdings stärker als für kleine, was darauf zurückzuführen sein dürfte, daß das Signal bei größerem  $D$  schneller abfällt. Das Signal ist daher im Mittel kleiner und damit ist das Signal-zu-Rausch-Verhältnis schlechter als bei kleinerem  $D$ .



**a**



**b**

**Abbildung 5.1:** Mittlerer relativer Fehler der mit dem erweiterten Strukturtensor geschätzten Diffusionskonstanten  $D$ : Aufgetragen gegen **a** die Standardabweichung des Rauschens und **b** die Geschwindigkeit  $v$  mit der sich die gaußförmige Grauwert-Struktur bewegt. Die verschiedenen Kurven ergeben sich aus Bildsequenzen, die mit unterschiedlichem  $d$  synthetisiert wurden.

In Abbildung 5.1b ist der mittlere relative Fehler der Diffusionskonstanten  $D$  für verschiedene  $d$  über der Geschwindigkeit  $v = |v|$  bei einer Standardabweichung des Rauschens von  $\sigma_n = 0.5$  aufgetragen. Zwei Trends sind zu erkennen: Zum einen nimmt der Fehler mit steigender Geschwindigkeit zu, von einem bestimmten Punkt an sogar sprunghaft.

Dieses Verhalten ist damit zu erklären, daß die lokal berechneten zeitlichen Ableitungen bei zu großen Verschiebungen zwischen zwei aufeinanderfolgenden Bildern der Sequenz nicht mehr korrekt bestimmt werden können. Das zeitliche Abtasttheorem ist dann nicht mehr erfüllt. Zum anderen sind die Fehler bei größeren  $d$  generell kleiner und die Grenzgeschwindigkeit, ab der der Fehler sprunghaft ansteigt, verschiebt sich zu größeren Geschwindigkeiten hin. Dies liegt daran, daß höhere Diffusionskonstanten  $d$  die Grauwertverteilung schneller abflachen lassen; diese flacheren Strukturen lassen sich auch bei höheren Geschwindigkeiten noch zuordnen. Anders ausgedrückt: Die höheren Diffusionskonstanten unterdrücken hohe Wellenzahlen stärker und daher ist das zeitliche Abtasttheorem auch bei höheren Geschwindigkeiten noch erfüllt.





**Teil II**

**Algorithmik**



# Kapitel 6

## Klassisches „Depth from focus“

In diesem Kapitel sollen die klassischen Depth-From-Fokus-Verfahren genauer betrachtet und analysiert werden. Es wird hier davon ausgegangen, daß die zu einem Objektpunkt gehörenden und korrespondierenden Bildpunkte schon bekannt sind, d.h. daß eine gegebenenfalls durch die Variation der Kameraparameter auftretende Skalierung schon korrigiert wurde bzw. gar nicht auftritt, da eine in Kapitel 2.3 beschriebene telezentrische Optik verwendet wurde. Für die Auswertung von Fokussequenzen, die mit nicht telezentrischen Optiken aufgenommen worden sind, wird in Kapitel 7 ein neuartiges Verfahren vorgestellt.

In Abschnitt 6.1 wird zunächst der Verlauf des Unschärfeparameters (Unschärfekreisradius im Falle der Pillbox-PSF) sowie das Frequenzverhalten der optischen Transferfunktion über den Verlauf der Fokuserie genauer beschrieben. Anschließend wird in Abschnitt 6.2 die lokale Varianz als ein Schärfemaß vorgestellt, analysiert und ihr Fehlerverhalten untersucht. Abschnitt 6.3 befaßt sich dann mit Verfahren zur schnellen und robusten Bestimmung des Maximums des Schärfemaßes, während in Abschnitt 6.4 abschließend die Rekonstruktion eines in weiten Bereichen tiefenscharfen Bildes beschrieben wird.

### 6.1 Verlauf der Unschärfeparameters

Um eine Fokuserie aufzunehmen, kann man entweder fokussieren, d.h. den Abstand Linse-Schirm/CCD variieren oder aber man verändert den Abstand Objekt-Kamera. Der Verlauf

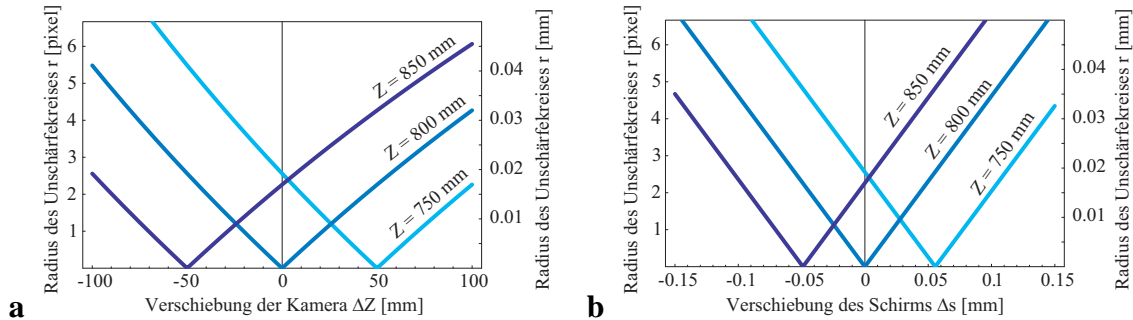
des Unschärfekreisradius wurde für diese beiden Fälle in 2.1.2 hergeleitet.

$$r(s) = \left| -\frac{1}{2n_f} \left[ 1 - \frac{f}{Z} \right] (s - s_z) \right| \quad \text{mit} \quad s_z = \frac{Zf}{Z - f} \quad (6.1)$$

$$r(Z) = \left| \frac{fs}{2n_f Z} - \frac{s - f}{2n_f} \right| \quad (6.2)$$

In Abbildung 6.1 sind beide Fälle dargestellt. Die Kamera mit Brennweite  $f = 25$  mm und Blende  $n_f = 1.4$  war dabei auf eine Entfernung von 800 mm scharfgestellt, d.h. der Schirmabstand ist  $s_z = 25.81$  mm. Man erkennt den linearen Verlauf von (6.1) in Teilbild b. In Teilbild a ist im Gegensatz zu Abbildung 2.3 aus Abschnitt 2.1.2 nun auch ein nahezu linearer Verlauf von Gleichung (6.2) zu sehen. Dies liegt an dem hier dargestellten Ausschnitt, der für eine Szene mit einer Tiefenvariation von ca. 200 mm und einen Beobachtungsabstand von 800 mm völlig ausreichend ist. Nähert man  $r(Z)$  in einer Taylorreihe bis zur ersten Ordnung<sup>1</sup> um den Objektabstand  $Z_s$ , der scharf abgebildet wird, ergibt sich dieser lineare Zusammenhang. Wie in Abschnitt 2.1.3 beschrieben, erhält man mit  $\Delta Z = Z - Z_s$ :

$$\tilde{r}(\Delta Z) = \left| -\frac{fs}{2n_f Z_s^2} \Delta Z \right| \quad (6.3)$$



**Abbildung 6.1:** Radius  $r$  des Unschärfekreises für drei unterschiedlich weit entfernte Objektpunkte. **a** in Abhängigkeit von der relativen  $Z$ -Verschiebung der Kamera  $\Delta Z$ . **b** in Abhängigkeit der relativen Schirmposition  $\Delta s$ . Die Kamera mit Brennweite  $f = 25$  mm und Blende  $n_f = 1.4$  war jeweils auf eine Entfernung von 800 mm scharfgestellt, d.h.  $s_z = 25.81$  mm.

In beiden Fällen erhält man für den Radius des Unschärfekreises den Betrag einer Geraden, die man folgendermaßen charakterisieren kann:

$$r(i) = |A(Z) \cdot (i - i_0(Z))| \quad (6.4)$$

<sup>1</sup>Terme höhere Ordnung sind jeweils um den Faktor  $1/Z_s$  kleiner, d.h. in diesem Fall fast drei Größenordnungen, so daß man sie vernachlässigen kann.

Der Betrag dieser Geraden ist bei  $i = i_0$  minimal und sowohl dieses Minimum  $i_0(Z)$ , als auch die Geradensteigung  $A(Z)$  können vom Objektabstand abhängen. Ziel der Depth-From-Fokus-Verfahren ist es, das Minimum  $i_0(Z)$  an jedem Bildpunkt zu bestimmen und daraus die Objektentfernung zu berechnen. Die Genauigkeit, mit der dies möglich ist, hängt neben der Genauigkeit der Bestimmung der Unschärferadien von der Steigung  $A(Z)$  ab.

Im Allgemeinen kann der Unschärferadius nicht direkt bestimmt werden, da sich die Unschärfekreise der benachbarten Objektpunkte überlagern. Es muß also ein Maß gefunden werden, daß proportional zum Unschärferadius ist, oder das zumindest monoton mit ihm steigt bzw. fällt und ein ausgeprägtes Extremum an der Stelle  $i = i_0$  hat.

Dazu muß man sich die Veränderungen ansehen, welche die Unschärfe in den Bildern verursacht. In Kapitel 2 wurden drei 2D-Punktantworten beschrieben, von denen die Gauß- und die Kreisscheiben-Funktion genauer betrachtet werden sollen. Da hier keine spezielle monochromatische Beleuchtung zum Einsatz kommt, ist die dritte Funktion, die Airy-Funktion, von nicht so großer Bedeutung. Während der Unschärfekreisradius bei der Kreisscheiben-Funktion direkt den Radius angibt, soll er, wie in der Literatur üblich, bei der Gaußfunktion mit der Standardabweichung  $\sigma$  gleichgesetzt werden, womit man

$$h_K(\mathbf{x}) = \frac{1}{\pi r^2} \Pi\left(\frac{|\mathbf{x}|}{2r}\right) \quad (6.5)$$

$$h_G(\mathbf{x}) = \frac{1}{2\pi r^2} \exp\left(-\frac{\mathbf{x}^2}{2r^2}\right) \quad (6.6)$$

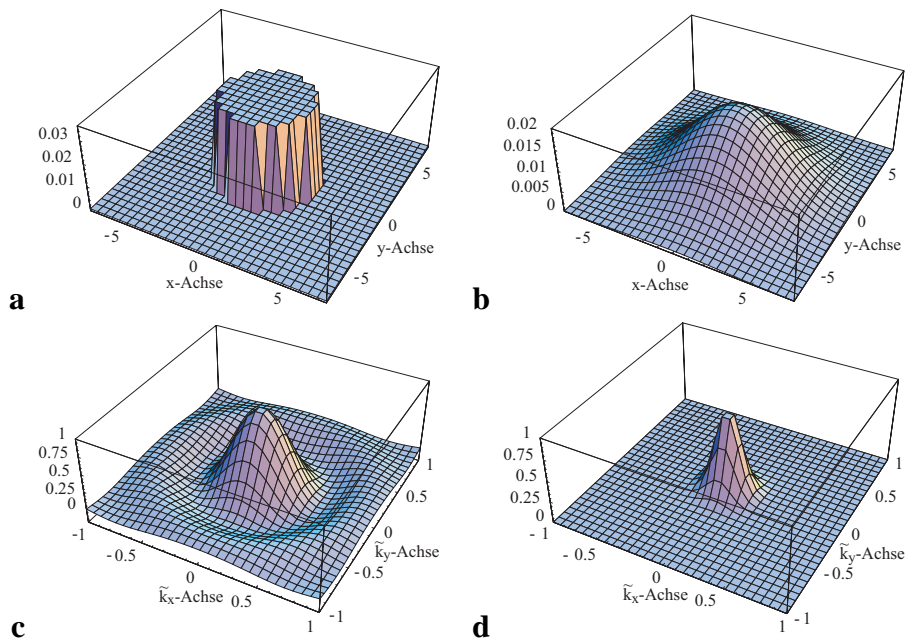
erhält, wobei  $\Pi(x)$  die Rechteckfunktion aus Gleichung (2.18) ist und die Indizes  $G$  und  $K$  für Gauß- bzw. Kreisscheiben-Funktion stehen. Die zugehörigen Transferfunktionen ergeben sich, wie in Abschnitt 2.1.4 beschrieben, mittels Fouriertransformation zu:

$$\hat{h}_K(\mathbf{k}) = \frac{2J_1(|\mathbf{k}|r)}{|\mathbf{k}|r} \quad (6.7)$$

$$\hat{h}_G(\mathbf{k}) = \exp\left(-\frac{\mathbf{k}^2 r^2}{2}\right) \quad (6.8)$$

In Abbildung 6.2 sind die Punktantworten mit den dazu gehörenden Transferfunktionen nochmals dargestellt, wobei die Wellenzahl  $k$  durch die auf die maximal mögliche Wellenzahl normierte Wellenzahl  $\tilde{k}$  ersetzt wurde.

Von Interesse ist nun, wie einzelne Wellenzahlen  $\mathbf{k}$  mit wachsendem Unschärfekreisradius bzw. Unschärfeparameter  $r$  gedämpft werden. Da die Transferfunktionen rotationssymmetrisch sind, wird im Folgenden nur noch der radiale Anteil betrachtet werden. In Abbildung 6.3 ist die Dämpfung von Bildstrukturen mit den normierten Wellenzahlen  $\tilde{k} = 1, 0.5, 0.25$  und  $0.125$  in Abhängigkeit vom Unschärfeparameter  $r$  dargestellt. Qualitativ würde man diesen Verlauf in einer Fokuserie beobachten, wenn entweder nur diese eine Wellenzahl



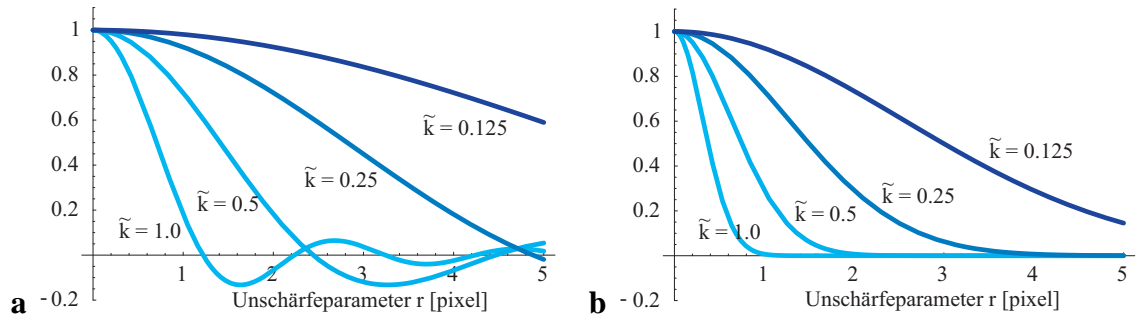
**Abbildung 6.2:** Die Punktantworten **a** Kreisscheiben-Funktion **b** Gaußfunktion mit ihren Transferfunktionen **c** die Airy-Funktionen und **d** wieder einer Gaußfunktion.

im Bild enthalten ist, oder wenn das Schärfemaß exakt diese eine Wellenzahl herausfiltern kann. Höherfrequente Bildstrukturen werden also schneller gedämpft und führen damit zu einem schärfer ausgeprägten Maximum. Für die kreisscheibenförmige Punktantwort erhält man zudem Nebenmaxima, die bei einer Maximumsuche berücksichtigt werden müssen.

In der Praxis sind noch weitere Punkte zu beachten:

- Der Kurvenverlauf ist noch etwas in  $r$ -Richtung skaliert, da man das Schärfemaß nicht über den Unschärfeparameter  $r$ , sondern über der  $Z$ - bzw.  $s$ -Verschiebung auftragen würde. Der Skalierungsfaktor ist dabei die Steigung aus Gleichung (6.4).
- Das Schärfemaß kann keinen so schmalen Wellenzahlbereich herausfiltern, denn aufgrund des reziproken Zusammenhangs zwischen Orts- und Frequenzraum, würde man im Ortsraum einen weit ausgedehnten Filter benutzen müssen und damit die Ortsauflösung verringern. Eine weiter ausgedehnte Transferfunktion des Schärfemaßes hat zudem den Vorteil, daß sich die Abhängigkeit von der Textur des Objekts verringert und daß die Nebenmaxima durch Überlagerung von verschiedenen Wellenzahlen verwischen, da die Nebenmaxima mit der Wellenzahl der Bildstruktur variieren, siehe Abbildung 6.3.
- Das Spektrum des Bildes ist nicht bekannt und soll auch nicht speziell vorausgesetzt werden. Daher kann der Kurvenverlauf auch für ein und dasselbe Schärfemaß bei

unterschiedlichem Bildmaterial nicht exakt berechnet werden.



**Abbildung 6.3:** Die Dämpfung von Bildstrukturen mit den normierten Wellenzahlen  $\tilde{k} = 1, 0.5, 0.25$  und  $0.125$  ist in Abhängigkeit vom Unschärfeparameter  $r$  dargestellt. Dabei wurde in **a** eine kreisscheibenförmige und in **b** eine gaußförmige Punktantwort zugrundegelegt.

## 6.2 Lokale Schärfemaße

Aus dem vorhergehenden Kapitel lassen sich folgende Anforderungen an ein “gutes“ Schärfemaß zusammenfassen:

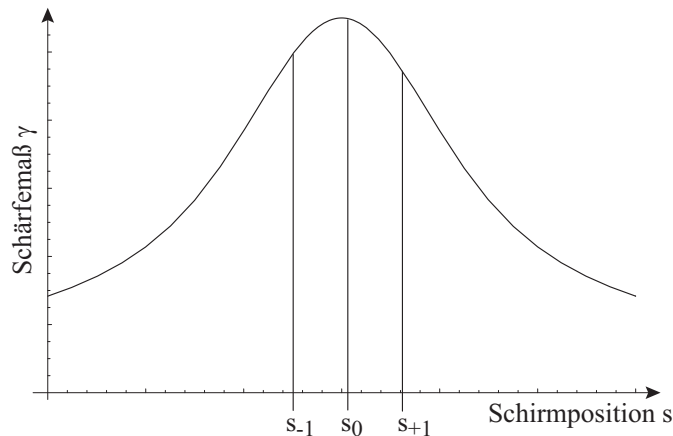
- Es sollte möglichst nur hohe Wellenzahlen berücksichtigen, da dann das Maximum schärfer ausgeprägt ist.
- Es darf aber auch nicht zu schmalbandig sein, damit zum einen der Filterkern im Ortsraum nicht zu groß und daher die Ortsauflösung zu klein wird. Zum anderen wäre ein schmalbandiges Schärfemaß von dem Vorhandensein weniger spezieller Wellenzahlen abhängig.
- Und es sollte möglichst rauschunempfindlich sein.

### 6.2.1 Statistischer Fehler

Bevor nun einige spezielle Schärfenmaße besprochen werden, soll ein Maß für den Fehler in der Maximumbestimmung hergeleitet werden. Qualitativ ist klar, daß dieser sowohl von der (Halbwerts-)Breite des Maximums, als auch von dem Fehler in jedem einzelnen Schärfewert abhängen muß. Wie im vorherigen Abschnitt ausgeführt, ist der Verlauf des

Schärfemaßes  $\gamma$  nicht genau bekannt; man kann den Verlauf in der Nähe des Maximums als nahezu parabelförmig annehmen.

$$\gamma(s) = as^2 + bs + c \quad (6.9)$$



**Abbildung 6.4:** Skizzierter Verlauf des Schärfemaßes  $\gamma$  in Abhängigkeit von der Schirm- bzw. CCD-Position  $s$ .

Hat man nun an drei Schirmpositionen  $s_{-1}$ ,  $s_0$  und  $s_1$  in der Nähe des Maximums den Wert des Schärfemaßes, wie in Abbildung 6.4 skizziert, so kann man die Position des Maximums  $s_z$  bestimmen. Ohne Beschränkung der Allgemeinheit kann man zur einfacheren Berechnung annehmen, daß  $s_{-1} = -\delta$ ,  $s_0 = 0$  und  $s_1 = \delta$  ist. Die Werte des Schärfemaßes an diesen Stellen werden mit  $\gamma_{-1} = \gamma(s_{-1})$ ,  $\gamma_0 = \gamma(s_0)$  und  $\gamma_1 = \gamma(s_1)$  bezeichnet. Weiterhin soll, da  $s$  in der Nähe des Maximums liegt,  $\gamma_0 > \gamma_{-1}$  und  $\gamma_0 > \gamma_1$  gelten. Für die Parameter der Parabel erhält man dann:

$$a = \frac{\gamma_1 + \gamma_{-1} - 2\gamma_0}{2\delta^2}, \quad b = \frac{\gamma_1 - \gamma_{-1}}{2\delta}, \quad c = \gamma(s_0) \quad (6.10)$$

Zur Berechnung der Position des Maximums setzt man nun die erste Ableitung von (6.9) gleich Null und erhält:

$$s_z = -\frac{b}{2a} = \frac{\delta}{2} \frac{\gamma_1 - \gamma_{-1}}{2\gamma_0 - \gamma_1 - \gamma_{-1}} \quad (6.11)$$

Von Interesse ist hier nun aber nicht die Position des Maximums selbst, sondern ihr Fehler. Dazu nimmt man an, daß die Meßwerte des Schärfemaßes aus einem reinen Signalanteil  $\bar{\gamma}_i$  und einem Rauschanteil  $n_i$  bestehen:

$$\gamma_i = \bar{\gamma}_i + n_i, \quad i \in -1, 0, 1 \quad (6.12)$$



Für die Position des Maximums erhält man damit aus (6.11) und (6.12):

$$s_z = \frac{\delta}{2} \left[ \frac{\bar{\gamma}_1 - \bar{\gamma}_{-1}}{2\bar{\gamma}_0 - \bar{\gamma}_1 - \bar{\gamma}_{-1}} \right] \left[ 1 + \frac{n_1 - n_{-1}}{\bar{\gamma}_1 - \bar{\gamma}_{-1}} \right] \left[ 1 + \frac{2n_0 - n_1 - n_{-1}}{2\bar{\gamma}_0 - \bar{\gamma}_1 - \bar{\gamma}_{-1}} \right]^{-1} \quad (6.13)$$

Der letzte Faktor ist nahezu Eins, da

$$|2\bar{\gamma}_0 - \bar{\gamma}_1 - \bar{\gamma}_{-1}| \gg |2n_0 - n_1 - n_{-1}| \quad (6.14)$$

gilt, falls das Rauschen nicht zu stark ist und falls  $\bar{\gamma}_0 > \bar{\gamma}_{-1}$  und  $\bar{\gamma}_0 > \bar{\gamma}_1$  gilt, was in der Nähe des Maximums der Fall ist. Damit erhält man also

$$s_z \approx \bar{s}_z \left( 1 + \frac{n_1 - n_{-1}}{\bar{\gamma}_1 - \bar{\gamma}_{-1}} \right) = \bar{s}_z + \frac{\delta}{2} \left( \frac{n_1 - n_{-1}}{2\bar{\gamma}_0 - \bar{\gamma}_1 - \bar{\gamma}_{-1}} \right) \quad (6.15)$$

wobei  $\bar{s}_z$  folgendermaßen definiert ist:

$$\bar{s}_z = \frac{\bar{\gamma}_1 - \bar{\gamma}_{-1}}{2\bar{\gamma}_0 - \bar{\gamma}_1 - \bar{\gamma}_{-1}} \quad (6.16)$$

Der letzte Term in Gleichung (6.15) stellt also den Fehler in der Position des Maximums dar. Für die Standardabweichung der Maximumsposition erhält man damit:

$$\sigma_{s_z} = \frac{\delta}{2} \frac{\sqrt{\sigma_{-1}^2 + \sigma_1^2}}{2\bar{\gamma}_0 - \bar{\gamma}_1 - \bar{\gamma}_{-1}} \quad (6.17)$$

Dabei ist  $\sigma_{-1}$  und  $\sigma_1$  die Standardabweichung des Rauschens in den jeweiligen Schärfemessungen. Die Standardabweichung ist also proportional zum Quotienten aus dem Abstand  $\delta$ , in dem die Schärfewerte gemessen wurden, und der Differenz zwischen dem höheren und den beiden niedrigeren Schärfewerten.

$$\sigma_{s_z} \propto \frac{\delta}{(\bar{\gamma}_0 - \bar{\gamma}_1) + (\bar{\gamma}_0 - \bar{\gamma}_{-1})} \quad (6.18)$$

Benutzt man noch die Definition des Parameters  $a$ , siehe Gleichung (6.10), erhält man:

$$\sigma_{s_z} \propto \frac{1}{\delta} \frac{1}{a} \quad (6.19)$$

Da  $1/\sqrt{a}$  ein Maß für die Breite des Maximum-Peaks ist, ist die Standardabweichung, wie schon oben vermutet, proportional zum Quadrat aus der Breite des Maximums. Ein Schärfemaß mit einer recht hohen Standardabweichung kann also bei genügend schmalen Maximum-Peak eine geringere Standardabweichung der Position des Maximums bewirken als ein Schärfemaß mit einer kleinen Standardabweichung, aber einem sehr breiten Maximums-Peak.

### 6.2.2 Fehler durch bestimmte Bildstrukturen

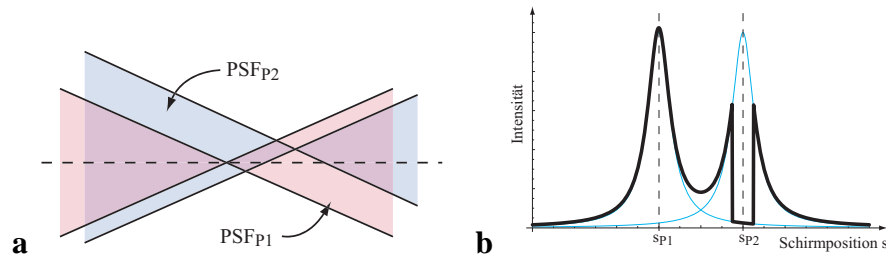
Neben dem im vorherigen Abschnitt beschriebenen statistischen Fehler, der durch das Kamerarauschen bedingt ist, gibt es auch noch Fehlerquellen, die auf Bildstrukturen bzw. auf das Nicht-Zusammenpassen von Modellannahmen und tatsächlichem Bildinhalt zurückzuführen sind. Bisher wurde immer angenommen, daß das Schärfemaß nur ein Maximum enthält und zwar an der Stelle, an der der betrachtete Objektpunkt scharf abgebildet wird. Es kann nun aber auch zwei oder mehr Maxima geben. Dies ist natürlich dann der Fall, wenn man – wie in der Mikroskopie – transparente Objekte untersucht, die eine entsprechende 3D-Struktur aufweisen. Aber auch bei opaken, also  $2\frac{1}{2}$ D-Szenen können an Tiefenkanten mehrere Maxima entstehen. Dafür gibt es im Wesentlichen zwei Ursachen:

Zum einen muß, um das Schärfemaß am Bildpunkt  $x$  zu berechnen, ein räumlich ausgehnter Bildbereich  $U$  um  $x$  herangezogen werden. Liegt innerhalb dieses Bereichs ein Tiefensprung, so wird über Bildpunkte gemittelt, deren zugehörige Objektpunkte unterschiedliche Entfernungen haben. Damit erhält man an diesem Bildpunkt  $x$  immer dann im Verlauf der Fokusserie ein lokales Maximum des Schärfemaßes, wenn einer dieser Objektpunkte gerade scharf abgebildet wird.

Dieses Problem wird also durch die Ausdehnung des Schärfefilters verursacht und ließe sich daher auch abmildern, indem man kleinere Filter benutzt oder den Filter an Kanten adaptiv anpaßt.

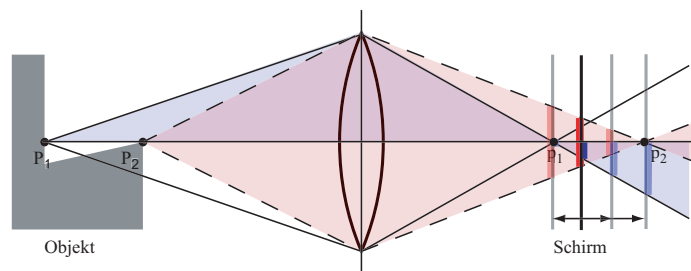
Die zweite Ursache liegt allerdings in der Physik des Abbildungsprozesses selbst und läßt sich nicht durch angepaßte Filter verhindern. Wiederum sind zwei Ursachen zu unterscheiden. Zum einen überlappen sich die 3D-Punktantworten von Objektpunkten, die im Bild benachbarte Bildpunkte haben, vor allem in Bereichen, in denen mindestens ein Objektpunkt unscharf abgebildet wird. Daher wird der Verlauf des Schärfemaßes auch von den in diesem Sinne benachbarten Objektpunkten beeinflußt. Haben diese Objektpunkte, wie das in der Nähe einer Kante der Fall ist, unterschiedliche Entfernungen von der Linse, so können sich kleinere Nebenmaxima ergeben.

Stark vereinfacht kann man sich dies an der Skizze 6.5a verdeutlichen, in der die Punktantworten  $PSF_{P_1}$  und  $PSF_{P_2}$  zweier isolierter Objektpunkte  $P_1$  und  $P_2$  gezeigt sind. Betrachtet man nun den Intensitätsverlauf entlang der gestrichelten Linie, wie in Abbildung 6.5b dargestellt, ergeben sich neben dem Maximum an der Stelle  $s = s_{P_1}$ , an der  $P_1$  scharf abgebildet wird, zwei kleinere Nebenmaxima, die symmetrisch um die Position  $s = s_{P_2}$  angeordnet sind, an der der benachbarte Objektpunkt  $P_2$  scharf abgebildet wird. Die Intensität ist zwar nicht identisch mit dem Schärfemaß und es erfolgt auch nicht die Überlagerung von nur zwei PSFs, sondern von unendlich vielen, trotzdem läßt diese Überlegung Rückschlüsse auf den Verlauf des Schärfemaßes zu, so daß man zusammenfassen



**Abbildung 6.5:** **a** Die sich überlagernden Punktantworten zweier benachbarter Objektpunkte mit unterschiedlicher Entfernung von der Linse sind hier veranschaulicht. In **b** ist der Intensitätsverlauf entlang der gestrichelten Linie skizziert. Neben dem Hauptmaximum ergeben sich zwei kleinere Nebenmaxima, die durch den benachbarten, in einer anderen Entfernung liegenden Objektpunkt entstehen.

kann: Haben nicht alle benachbarten Punkte die gleiche Entfernung, so kann man annehmen, daß sich bei kleinen Entfernungsunterschieden der Maximums-Peak verbreitert. Bei großen Entfernungsunterschieden, wie z.B. an Tiefenkanten, können Nebenmaxima entstehen, auch wenn diese flacher und breiter sein werden, als die in Abbildung 6.5b.

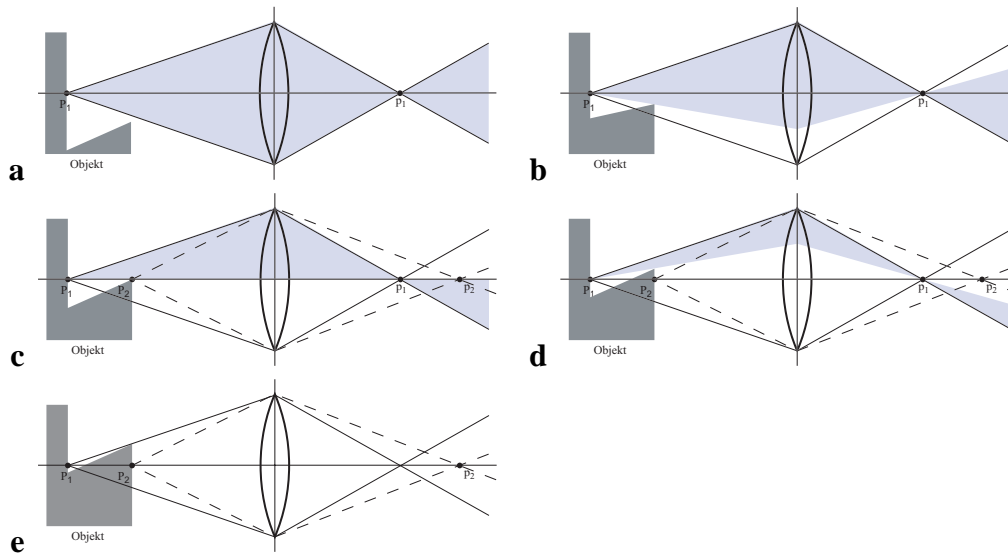


**Abbildung 6.6:** Veranschaulichung der Abbildung eines Objekts mit Tiefenkante: An der Kante werden zwei Objektpunkte mit unterschiedlicher Entfernung auf dieselbe Bildkoordinate abgebildet.

Zum anderen kann auch bei undurchsichtigen Objekten der Fall eintreten, daß, wie in der Mikroskopie transparenter Objekte, zwei Objektpunkte unterschiedlicher Entfernung auf ein und dieselbe Position im Bild abgebildet werden. Dies ist in der Nähe von Tiefenkanten der Fall und im Verlauf der Fokussierung erhält man an diesen Stellen dann zwei Maxima. In Abbildung 6.6 ist ein Objekt mit einer solchen Tiefenkante dargestellt. Die in unterschiedlichen Entfernung befindlichen Objektpunkte  $P_1$  und  $P_2$  werden auf die gleichen Bildkoordinaten abgebildet<sup>2</sup>, so daß im Verlauf der durch die verschiedenen Schirmposi-

<sup>2</sup>Die Punkte liegen hier auf der optischen Achse und entlang dieser hintereinander, um die Zeichnungen übersichtlicher zu halten. Die gleiche Argumentation ließe sich aber auch für andere Punkte durchführen, wenn entweder eine telezentrische Optik benutzt würde oder die entsprechenden Punkte entlang einer Linie vom Durchstoßpunkt der optischen Achse durch die Linse lägen.

tionen angedeuteten Fokussérie zwei Maxima entstehen. Diese zwei Maxima sind nicht nur direkt an der Kante, sondern auch in deren Umgebung zu beobachten; und zwar solange noch ein Teil des von  $P_1$  ausgehenden Lichtkegels die Linse erreicht, oder umgekehrt formuliert, solange der Objektpunkt  $P_1$  noch von irgendeinem Punkt der Linse aus zu sehen ist.



**Abbildung 6.7:** Veranschaulichung der Abbildung einer Tiefenkante: **a** Die Kante hat noch keinen Einfluß auf die Abbildung von  $P_1$ , **b** erste Abschattungen des Lichtkegels, **c** weitere Abschattung und Überlagerung von zweitem Objektpunkt  $P_2$ , **d** auch der Hauptstrahl von  $P_1$  wird abgeschattet, und **e**  $P_1$  ist vollständig verdeckt.

Nähert man sich der Kante, so ist zunächst keine Störung im Verlauf des Schärfenmaßes zu erwarten<sup>3</sup>; es besitzt ein klares Maximum an der Stelle, an der der Objektpunkt  $P_1$  scharf abgebildet wird. Diese Situation ist in Abbildung 6.7a skizziert. Bei einer weiteren Annäherung an die Kante ist zunächst eine teilweise Abschattung des von  $P_1$  ausgehenden Lichtkegels durch weiter vorne liegenden Szenenteile, in diesem Fall durch das Objekt selbst, zu beobachten (s. Teilbild b). Von der z.B. kreisförmigen Punktantwort wird also ein Teil abgedunkelt, aber es ist noch mehr als die Hälfte der Kreisscheibe vorhanden, so daß trotz verändertem Kurvenverlauf noch immer ein deutliches Maximum an der korrekten Position zu erwarten ist. In Teilbild c ist nun schon die Hälfte des Lichtkegels und damit der PSF verdeckt, die Kurvenform hat sich weiter verändert und auch die Höhe des Maximums ist jetzt deutlich reduziert, da nur noch die Hälfte der Intensität den Schirm erreicht. Zudem überlagert sich nun auch der vom Objektpunkt  $P_2$  ausgehende, ungestörte Lichtke-

<sup>3</sup>Durch den weiter oben schon beschriebenen Effekt wird der Verlauf des Schärfenmaßes gegebenenfalls schon verändert. Dieser Effekt bewirkt auch, daß noch bevor die Kante erreicht ist, ein zweites Maximum entsteht.

gel und man erhält ein zweites klares Maximum an der Stelle, an der  $P_2$  scharf abgebildet wird. Auch wenn in Teilbild d der Hauptstrahl von  $P_1$  schon abgeschattet wird, erreicht immer noch Teil des von  $P_1$  ausgehenden Lichtkegels die Linse und erzeugt ein Maximum im Verlauf des Schärfemaßes. Dies ist jetzt im Vergleich zum von  $P_2$  hervorgerufenen Maximum deutlich kleiner und daher als Nebenmaximum zu bezeichnen.

Entfernt man sich noch weiter von der Kante (s. Teilbild e), gelangt ab einem bestimmten Punkt kein Licht von  $P_1$  mehr zur Linse und das Schärfemaß hat damit nur noch ein klares Maximum, das von  $P_2$  hervorgerufen wird. Umgekehrt betrachtet, kann dann von keinem Punkt der Linse mehr um die Kante herum oder an ihr vorbei gesehen werden. Bezeichnet man die Entfernung von  $P_{1,2}$  von der Linse mit  $Z_{1,2}$  und den Linsen- bzw. Blendenradius mit  $r_f$ , kann man den Abstand  $d$  von der Kante, bei dem  $P_1$  keinen Einfluß mehr hat, folgendermaßen schreiben:

$$d = \frac{Z_1 - Z_2}{Z_1} r_f \quad (6.20)$$

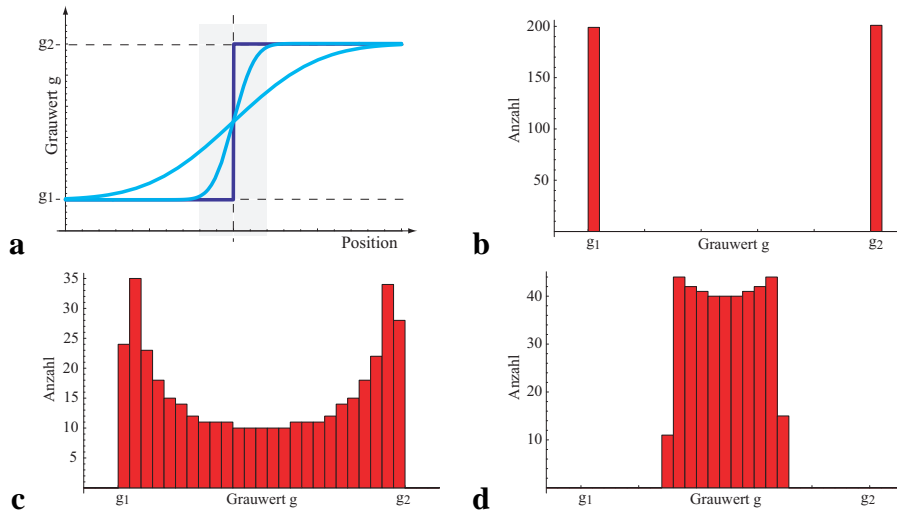
### 6.2.3 Lokale Varianz als Schärfemaß

Die Motivation zur Benutzung der lokalen Varianz als Schärfemaß ergibt sich aus folgender Überlegung: Betrachtet man die Abbildung einer idealen Grauwertkante von Grauwert  $g_1$  nach  $g_2$ , so wird der zunächst abrupte Helligkeitsübergang mit zunehmender Unschärfe glatter, wie in Abbildung 6.8a dargestellt. In Abbildung 6.8b–c ist die lokale Verteilung der Grauwerte bzw. das lokale Grauwert-Histogramm an der Kante dargestellt. Diese geht von einer Verteilung, die nur die beiden Grauwerte  $g_1$  und  $g_2$  enthält, über in eine, die immer mehr Zwischenwerte aufweist. D.h. die Varianz dieser Verteilung nimmt mit zunehmender Unschärfe ab und zwar für  $|g_1 - g_2| = 1.0$  von  $\sigma^2 = 0.25$  über  $\sigma^2 = 0.11$  auf  $\sigma^2 = 0.01$ . Damit eignet sich die lokale Varianz grundsätzlich als Schärfemaß.

Die Berechnung der Varianz der Grauwertverteilung einer kleinen Umgebung im Bild, der *lokalen Varianz*, folgt aus der Formel der Varianz für die  $N$  Grauwerte  $x_0, x_1, \dots, x_N$ :

$$\sigma^2 = \frac{1}{N} \sum_{i=0}^N (x_i - \bar{x})^2 \quad \text{mit} \quad \bar{x} = \frac{1}{N} \sum_{i=0}^N x_i \quad (6.21)$$

Zuerst benötigt man den *lokalen Mittelwert* (in oben stehender Gleichung  $\bar{x}$ ), den man z.B. durch die Faltung des Bildes mit einer Binomialmaske  $B_1$  erhält. Die Binomialmaske bewirkt zusätzlich eine Wichtung der Bildpunkte. Die vom Zentrum der Umgebung weiter entfernten Bildpunkte gehen dadurch mit weniger stark in den lokalen Mittelwert ein als die näher liegenden.



**Abbildung 6.8:** In **a** ist der Grauwertverlauf senkrecht zur Abbildung einer idealen Grauwertkante dargestellt. Mit zunehmender Unschärfe der Abbildung wird die Kante immer glatter. Das Histogramm des grau unterlegten Bereichs ist für die immer glatteren Grauwertverläufe in **b–c** zu sehen. Die zunächst bimodale Verteilung wird mit steigender Unschärfe immer kompakter und damit sinkt ihre Varianz.

Der lokale Mittelwert, das Mittelwertbild, wird dann von den originalen Werten, dem originalen Bild, abgezogen; das Ergebnis wird punktweise quadriert und anschließend mit einer Binomialmaske  $B_2$  wieder lokal gemittelt. Für das Varianzbild  $G_{\text{Varianz}}$  ergibt sich also

$$G_{\text{Varianz}} = B_2 * [(G - B_1 * G) \cdot (G - B_1 * G)] \quad (6.22)$$

und für den Varianzoperator  $\mathcal{V}$  erhält man in Operatorschreibweise:

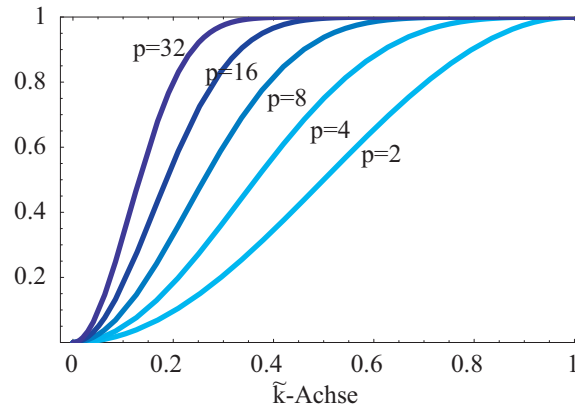
$$\mathcal{V} = B_2 [(1 - B_1) \cdot (1 - B_1)]. \quad (6.23)$$

**Frequenzverhalten:** Die Wahl der Binomialmasken  $B_1$  und  $B_2$  bestimmt die Größe der Umgebung, in der die Varianz berechnet wird und damit auch das Antwortverhalten des gesamten Varianzoperators auf die unterschiedlichen Wellenzahlen der im Bild enthaltenen Strukturen.

Da der Varianz-Operator aber nichtlinear ist, läßt sich keine Transferfunktion für ihn angeben. Trotzdem kann man einige Aussagen über sein Antwortverhalten auf verschiedene Wellenzahlen treffen. Dazu überträgt man Gleichung (6.22) für die lokale Varianz in den Fourierraum, womit aus den Faltungen komplexe Multiplikationen werden und umgekehrt. Man erhält also:

$$\hat{G}_{\text{Varianz}} = \hat{B}_2 \cdot [(\hat{G} - \hat{B}_1 \cdot \hat{G}) * (\hat{G} - \hat{B}_1 \cdot \hat{G})] \quad (6.24)$$

Betrachtet man zunächst nur den linearen Teiloperator  $(1 - \mathcal{B}_1)$ , so kann man aus dessen Transferfunktion ablesen, auf welche Wellenzahlen der Varianzoperator gar nicht reagieren kann, da diese vollständig weggedämpft werden. In Abbildung 6.9 sind diese Transferfunktionen für verschiedene Binomialmasken dargestellt. Man erkennt sofort, daß dieser Filter eine Art Hochpaß-Filter darstellt, da die kleinen Wellenzahlen stark unterdrückt werden und der Mittelwert sogar ganz verschwindet, während hohe Wellenzahlen unvemindert durchgelassen werden. Mit zunehmender Größe der Binomialmaske sinkt die Grenzwellenzahl ab und die Steilheit des Filter nimmt zu.



**Abbildung 6.9:** Transferfunktion des Eingangsfilters  $(1 - \mathcal{B}_1)$  für  $\mathcal{B}_1 = {}^p\mathcal{B}$  mit  $p = 2, 4, 8, 16, 32$ .

Die punktweise Quadratur im Ortsraum entspricht einer Faltung des Spektrums mit sich selbst im Fourierraum. Dies ist der nichtlineare Schritt im Varianzoperator und er kann daher nicht durch eine Transferfunktion beschrieben werden. Trotzdem ist es nützlich, ein einfaches, eindimensionales, periodisches Signal der Form

$$g(x) = g_0 \cos(k_0 x + \phi) \quad (6.25)$$

zu betrachten. Dessen Fouriertransformierte besteht aus zwei Delta-Funktionen im Abstand von jeweils  $k_0$  vom Ursprung:

$$\hat{g}(k) = \frac{g_0}{2} [e^{-i\phi} \delta(k - k_0) + e^{i\phi} \delta(k + k_0)] \quad (6.26)$$

Faltet man nun  $\hat{g}(k)$  mit sich selbst, erhält man drei Delta-Funktionen an den Stellen  $k = 0$  und  $k = \pm 2k_0$ :

$$\hat{g}(k) * \hat{g}(k) = \frac{g_0}{4} [2\delta(k) + e^{-i2\phi}\delta(k - 2k_0) + e^{i2\phi}\delta(k + 2k_0)] \quad (6.27)$$

D.h. der mittlere Grauwert im quadrierten Bild ist proportional zum Quadrat der Amplitude der periodischen Struktur und zusätzlich enthält das quadrierte Bild Strukturen mit der doppelten Wellenzahl  $2k_0$ . Enthält das Bild mehr als eine periodische Struktur mit einer Wellenzahl, so z.B. zwei mit den Wellenzahlen  $k_1$  und  $k_2$ , kann man sich leicht überlegen, daß das quadrierte Bild dann neben dem Mittelwert periodische Strukturen mit allen möglichen Summen und Differenzen der Wellenzahlen enthält, also  $k = k_i \pm k_j$  mit  $i, j \in 1, 2$ .

Allgemein läßt sich also festhalten, daß der Mittelwert des quadrierten Bildes proportional zur Summe der quadrierten Amplituden der periodischen Strukturen ist und daß sich die maximal im Bild enthaltene Wellenzahl verdoppelt. Dieser Mittelwert stellt damit ein gutes Schärfemaß dar, da die kleineren Wellenzahlen durch den  $(1 - \mathcal{B}_1)$  Filter schon unterdrückt wurden. Um diesen Mittelwert im Bild möglichst gut zu bestimmen, wird anschließend mit der Binomialmaske  $B_2$  gemittelt. Die zusätzlich durch das Quadrieren entstehenden Wellenzahlen bringen allerdings einige Probleme mit sich: Da der Filter  $(1 - \mathcal{B}_1)$  die hohen Wellenzahlen fast ungehindert passieren läßt, können Wellenzahlen bis zum Doppelten der Grenzwellenzahl entstehen. Diese können mit der zu Grunde liegenden Abtastung nicht mehr dargestellt werden und führen zu Störungen über den gesamten Wellenzahlbereich, dem sogenannten *Aliasing*. Da dieses Aliasing auch sehr niedrige Wellenzahlen umfassen kann, kann es nicht im nachhinein durch den Tiefpaß-Filter  $\mathcal{B}_2$  eliminiert werden. Es muß also von vornherein darauf geachtet werden, daß Wellenzahlen, die größer als die halbe Grenzwellenzahl sind, nicht mehr im Bild enthalten sind:

$$\hat{g}(\tilde{k}) = 0 \quad \forall |\tilde{k}| > 0.5 \quad (6.28)$$

Dazu kann der Eingangsfiler  $(1 - \mathcal{B}_1)$  entsprechend modifiziert werden.

Allerdings können Störungen mit kleinen Wellenzahlen, die sich schlecht oder gar nicht mit dem Tiefpaß-Filter  $\mathcal{B}_2$  eliminieren lassen, auch direkt durch das Quadrieren entstehen. Wie oben erwähnt, entstehen dabei Wellenzahlen, die den Summen und Differenzen aller im Ausgangsbild enthaltenen Wellenzahlen entsprechen. Eng benachbarte Wellenzahlen im Ausgangsbild bewirken also immer sehr kleine Wellenzahlen im quadrierten Bild, so daß gewisse Schwankungen im Varianzbild trotz konstanter Textur im Ausgangsbild  $g$  nicht vollständig zu vermeiden sind.



## 6.3 Bestimmung des Maximums

Nachdem im vorangegangenen Abschnitt das Schärfemaß  $\gamma(s)$  berechnet wurde, sollen nun verschiedene Methoden beschrieben werden, dessen Maximum zu finden. Neben der maximal erreichbaren Genauigkeit soll dabei auch auf Möglichkeiten einer einfachen und inkrementellen Implementierung Wert gelegt werden. Denn dadurch sind erst Implementierungen in spezieller Hardware z.B. FPGA möglich.

Im Folgenden wird davon ausgegangen, daß bei der Aufnahme der Fokusserie der Schirmabstand  $s$  variiert wurde. Hat man für einen Bildpunkt die Schirmposition  $s_{\max}$  gefunden, an der er am „schärfsten“ abgebildet wird, so kann daraus mittels Gleichung (2.2) die Entfernung des Objektpunkts berechnet werden. Wird im Laufe der Fokusserie statt dem Schirmabstand der Kameraabstand verändert, so entfällt dieser letzte Schritt.

### 6.3.1 Einfache Maximumsuche

Die schnellste und einfachste Methode besteht darin, das globale Maximum über die gesamte Fokussequenz zu suchen und dessen Position zu speichern.

$$s_{\max}^{x,y} = s_i | \gamma_i^{x,y} \geq \gamma_j^{x,y} \quad \forall j = 1, \dots, N \quad \text{und} \quad i \neq j \quad (6.29)$$

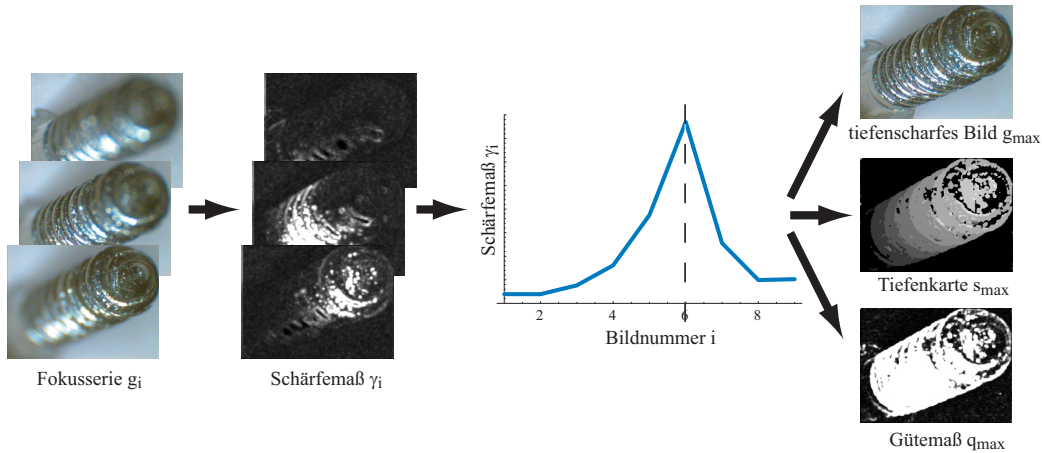
Dabei bezeichnet  $s_i$  die Schirmpositionen mit denen das  $i$ -te der  $N$  Bilder der Fokusserie aufgenommen wurde. Mit  $\gamma_i^{x,y}$  wird der Wert des Schärfemaßes im  $i$ -ten Bild am Bildpunkt  $(x, y)$  bezeichnet und  $s_{\max}^{x,y}$  ist schließlich die Schirmposition an der der Bildpunkt  $(x, y)$  am „schärfsten“ abgebildet wird. Aus  $s_{\max}^{x,y}$  kann mittels Gleichung (2.2) die Entfernung des entsprechenden Objektpunktes berechnet werden.

Aus der Information, in welchem Bild der Serie jeder Bildpunkt scharf abgebildet wird, läßt sich auf einfache Weise ein in weiten Bereichen tiefenscharfes Bild zusammensetzen, wie dies in Kapitel 6.4 beschrieben und auch in Abbildung 6.10 angedeutet ist.

Neben der reinen Information ist auch deren Verlässlichkeit wichtig, daher ist es sinnvoll ein Gütemaß zu berechnen, welches diese Verlässlichkeit wiedergibt. Ein einfacher Ansatz im hier vorliegenden Fall ist die Varianz des Schärfemaßes im Verlauf der Fokusserie.

$$q^{x,y} = \langle (\gamma^{x,y} - \langle \gamma^{x,y} \rangle)^2 \rangle \quad \text{mit} \quad \langle a \rangle = \frac{1}{N} \sum_{i=1}^N a_i \quad (6.30)$$

Liegt ein ausgeprägtes Maximum vor, z.B. bei starker Textur im Bild, so wird dieses größer sein als wenn die Schärfekurve flach verläuft, da keine oder nur schwache Textur im Bild



**Abbildung 6.10:** Die Auswertung einer Fokussierserie vom Originalbild, über das Schärfemaß und dessen Maximum bis hin zu der Tiefenkarte, dem tiefenscharfen Bild und dem Gütemaß.

vorhanden ist. Trotzdem muß beachtet werden, daß dieses Gütemaß auch dann groß sein kann, wenn die Schärfekurve stark verrauscht ist, aber kein deutliches Maximum aufweist.

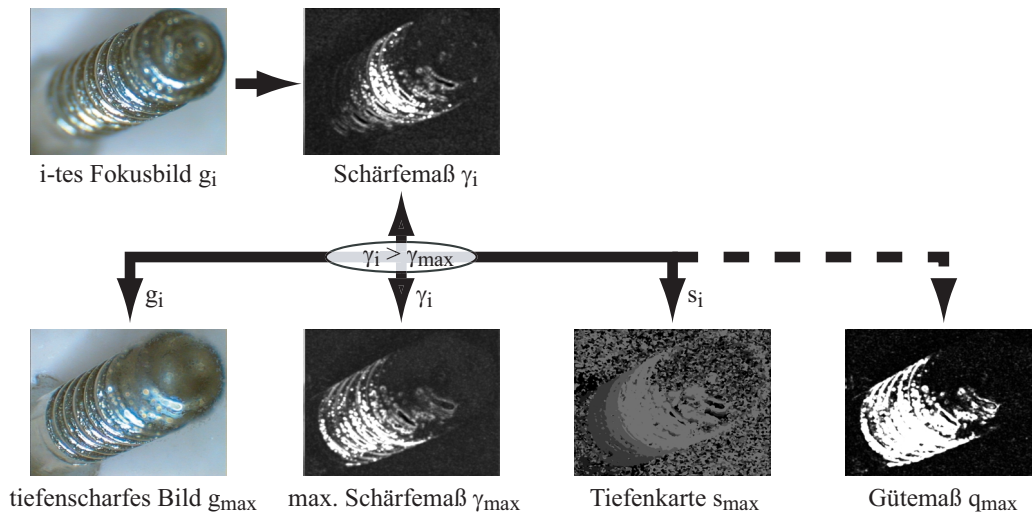
Die Tiefenauflösung dieser Methode ist durch den Abstand der Bilder der Fokussierserie bestimmt. Allerdings kann die in Abschnitt 6.2.1 berechnete Standardabweichung der Maximumposition nicht durch immer kleinere Bildabstände unterschritten werden, da sie durch die Standardabweichung des Schärfemaßes selbst verursacht wird.

### Inkrementelle Implementierung

Der größte Vorteil dieser Methode liegt in ihrer Einfachheit, die eine entsprechend einfache und speicherplatz-sparende inkrementelle Implementierung erlaubt. Dazu reicht es, statt der gesamten Fokussierserie nur zwei Bilder zu speichern, eins mit dem für jeden Bildpunkt im bisherigen Verlauf der Fokussierserie erreichten maximalen Schärfemaß  $\gamma_{\max}^{x,y}$ , und eins mit der zugehörigen Schirmposition  $s_{\max}^{x,y}$ . Für jedes neu aufgenommene Bild  $g_i$  der Fokussierserie wird das Schärfemaß  $\gamma_i$  berechnet und punktweise mit dem bisher maximalen Schärfemaß  $\gamma_{\max}$  verglichen. Wird im neuen Bild ein höherer Schärfewert erreicht, so wird dieser in  $\gamma_{\max}^{x,y}$  übernommen und die entsprechende Schirmposition in  $s_{\max}^{x,y}$  gespeichert. Speichert man zusätzlich noch ein drittes Bild, so kann gleichzeitig ein komplett tiefenscharfes Bild aufgebaut werden, wie dies in Kapitel 6.4 allgemeiner beschrieben wird. Dazu speichert man immer, wenn im neuen Bild ein schärferer Bildpunkt gefunden wurde, parallel zu dessen Schärfewert und Schirmposition dessen Grauwert in  $g_{\text{scharf}}^{x,y}$ . Und auch das oben erwähnte Gütemaß  $q$  kann inkrementell berechnet werden, dazu müssen allerdings zwei Bil-

der gespeichert werden, eins mit der Summe der bisherigen Schärfewerte  $\gamma_i$  und eins mit der Summe der Quadrate. Die Varianz und damit das Gütemaß  $q$  wird wie folgt berechnet:

$$q = \langle \gamma^2 \rangle - \langle \gamma \rangle^2 \quad (6.31)$$



**Abbildung 6.11:** Die inkrementelle Auswertung einer Fokusserie ist vom Originalbild, über das Schärfemaß und dessen Maximum bis hin zu der Tiefenkarte, dem tiefscharfen Bild und dem Gütemaß skizziert.

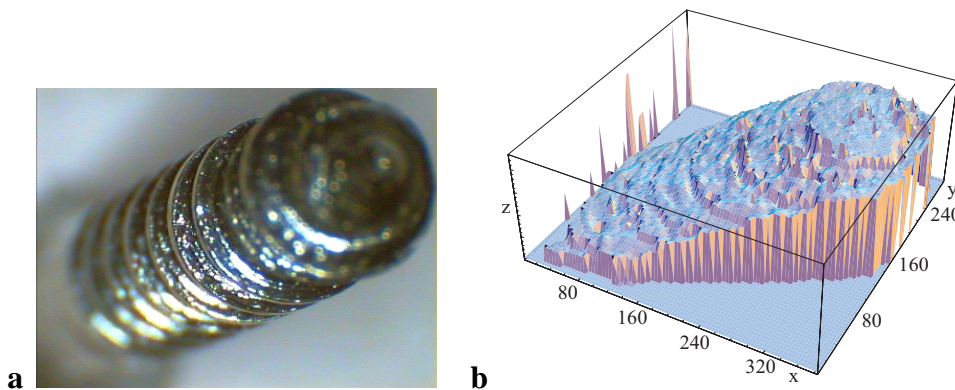
Eine Implementierung dieses einfachen Algorithmus auf einem MikroEnable-FPGA-Board der Firma Silicon-Software wurde im Rahmen des OpenEye-Projekts entwickelt. Durch die massiv parallele Berechnung im FPGA konnte, unabhängig von PC-Prozessor, bei einer Auflösung von  $256 \times 256$  Video-Echtzeit erreicht werden, wobei der begrenzende Faktor für die Bildauflösung nicht die Rechengeschwindigkeit, sondern die begrenzte Speicherkapazität war.

### 6.3.2 Schwerpunktbestimmung

Interpretiert man die Schärfemaßkurve nicht als Schärfemaß an bestimmten Schirmpositionen, sondern als Häufigkeit für bestimmte Schirmpositionen, so kann man als Maximumposition den gewichteten Mittelwert dieser „Verteilung“ annehmen. Physikalisch gesehen ist dies der Schwerpunkt einer entsprechenden Masseverteilung. Dieser Schwerpunkt stimmt mit der gesuchten Maximumposition dann überein, wenn die Kurve des Schärfemaßes symmetrisch um ihr Maximum ist. Bei der Berechnung wird daher ein erster

Schätzwert des Maximums mittels der *Einfachen Maximumsuche* aus dem vorangegangenen Abschnitt ermittelt, um den dann ein symmetrisches Intervall gelegt wird. Nur dieses Intervall wird zur Schwerpunktsbestimmung genutzt, wodurch systematische Fehler minimiert werden.

$$s_{\max}^{x,y} = \frac{\sum_{i=0}^N s_i \gamma_i^{x,y}}{\sum_{i=0}^N s_i} \quad (6.32)$$



**Abbildung 6.12:** **a** Ein Originalbild der Fokusserie. **b** 3D-Darstellung einer Tiefenkarte, die mit der Schwerpunktmethode berechnet wurde. Die Löcher in der Tiefenkarte, die durch ein zu geringes Gütemaß entstanden sind, wurden mittels Regularisierung mit Hilfe der normalisierten Faltung geschlossen.

Prinzipiell ist auch hierbei eine inkrementelle Implementierung denkbar. Dazu muß auch nur ein Bild  $t$  gespeichert werden, welches die Summe der bisherigen  $N'$  Schärfewerte multipliziert mit den entsprechenden Schirmpositionen enthält:

$$t^{x,y} = \sum_{i=0}^{N'} s_i \gamma_i^{x,y} \quad (6.33)$$

Um nun eine vorläufige Schätzung der Maximumposition zu bekommen, muß man nur durch die Summe der bisherigen Schirmpositionen  $\alpha$  teilen:

$$s_{\max}^{x,y} = \frac{t^{x,y}}{\alpha} \quad \text{mit} \quad \alpha = \sum_{i=0}^{N'} s_i \quad (6.34)$$

Da bei dieser inkrementellen Version des Algorithmus kein zum Maximum symmetrisches Intervall zur Berechnung herangezogen wird, sondern einfach alle Schärfewerte, werden

die so geschätzten Maximumpositionen leicht in die Richtung verschoben, in der die Intervallgrenze weiter vom Maximum entfernt ist. Um diese Verschiebung zu verringern, sollte man nur die Schärfewerte  $\gamma$  in die Summe in Gleichung (6.33) eingehen lassen, die eine Mindestschwelle überschreiten, welche sich aus der Standardabweichung des Schärfemaßes ergibt. Liegt das Maximum jetzt ausreichend weit von den Intervallgrenzen entfernt, so ist das Schärfemaß schon vorher unter obigen Schwellenwert gefallen und diese Werte tragen nicht mehr zur Schätzung der Maximumposition bei.

### 6.3.3 Lokale Polynominterpolation

Man legt hierzu durch jeweils drei aufeinanderfolgende Schärfewerte ein Polynom zweiten Grades, überprüft, ob das Maximum dieses Polynoms innerhalb der drei Stützpunkte liegt und sucht über die ganze Schärfekurve nach dem größten Polynommaximum.

$$s_{\text{pmax},j}^{x,y} = \frac{\gamma_{j-1}^{x,y} - \gamma_{j+1}^{x,y}}{2(\gamma_{j-1}^{x,y} - 2\gamma_j^{x,y} + \gamma_{j+1}^{x,y})} \quad \forall j = 1, \dots, N-1 \quad (6.35)$$

$$\gamma_{\text{pmax},j}^{x,y} = \frac{\gamma_j^{x,y} - (\gamma_{j+1}^{x,y} - \gamma_{j-1}^{x,y})^2}{8(\gamma_{j-1}^{x,y} - 2\gamma_j^{x,y} + \gamma_{j+1}^{x,y})} \quad \forall j = 1, \dots, N-1 \quad (6.36)$$

Dabei ist  $s_{\text{pmax},j}^{x,y}$  das Maximum des Polynoms am Bildpunkt  $(x, y)$  und um das Bild  $j$ .  $\gamma_{\text{pmax},j}^{x,y}$  stellen den zugehörigen maximalen Schärfewert dar. Aus diesen  $N-2$  Polynome ergibt sich das Gesamtmaximum als das Maximum mit dem größten Schärfewert  $\gamma_{\text{pmax},j}^{x,y}$  wobei noch gelten muß:

$$s_{j-1} \leq s_{\text{pmax},j}^{x,y} \leq s_{j+1} \quad (6.37)$$

### 6.3.4 Globaler Polynomfit

Bei dieser Methode wird versucht, einer Anzahl von  $N_{\text{fit}}, N_{\text{fit}} > 3$  Schärfewerten ein Polynom zweiten Grades anzunähern. Löst man dieses überbestimmte Problem, kann man neben dem Maximum auch noch den Fehler zwischen Polynom und angefiteten Daten errechnen. Im Allgemeinen ist die Zahl der zum Fit genutzten Wert  $N_{\text{fit}}$  kleiner als die Gesamtzahl der Schärfewerte  $N$ , so daß man mehrere Fits an verschiedenen Stellen durchführen kann und damit auch mehrere Maxima erhält.

$$\left( \sum_{i=0}^{N_{\text{fit}}} [p_j^{x,y}(s_{i+j}) - \gamma_{i+j}^{x,y}]^2 \right) \rightarrow \min \quad \Rightarrow \quad p_j^{x,y}(\cdot) \quad \forall j = 0, \dots, N - N_{\text{fit}} \quad (6.38)$$

Dabei ist  $p_j^{x,y}(z) = a_j^{x,y}z^2 + b_j^{x,y}z + c_j^{x,y}$  das gefittete Polynom am Bildpunkt  $(x, y)$  mit dem Startbild  $j$ . Zu jedem dieser  $N - N_{fit}$  Polynome erhält man das Maximum

$$\gamma_{pmax,j}^{x,y} = p_j^{x,y}(s_{pmax,j}^{x,y}) \quad \text{mit} \quad s_{pmax,j}^{x,y} = -\frac{b_j^{x,y}}{2a_j^{x,y}} \quad (6.39)$$

und den Fehler

$$e_j^{x,y} = \sum_{i=0}^{N_{fit}} |p_j^{x,y}(s_{i+j}) - \gamma_{i+j}^{x,y}| \quad (6.40)$$

Man kann als Gesamtmaximum nun entweder das Maximum mit dem größten Schärfewert  $\gamma_{pmax,j}^{x,y}$  oder mit dem geringsten Fitfehler  $e_j^{x,y}$  auswählen.

## 6.4 Synthese eines tiefenscharfen Bildes

Die Berechnung eines tiefenscharfen Bildes aus der Fokuserie gestaltet sich recht einfach. Hat man neben der Fokuserie  $g_i$  die Position des Schärfemaximums  $s_{max}$  an jedem Bildpunkt, so läßt sich punktweise dasjenige Bild  $g_i$  auswählen, welches der Maximumposition am nächsten liegt. Daraus entnimmt man dann den Grau- bzw. Farbwert für das tiefenscharfe Bild  $g_{max}$ . Diese einfache Vorgehensweise entspricht einer Interpolation nullter Ordnung. Damit verwirft man aber die im vorherigen Abschnitt gegebenenfalls zwischenbild-genau bestimmte Maximumposition; daher sollte man mindestens eine Interpolation erster Ordnung, d.h. eine lineare Interpolation einsetzen. Natürlich sind auch Polynom-Interpolationen höherer Ordnung oder auch Spline-Interpolationen nutzbar, hierzu sei z.B. auf Jähne [1997] verwiesen. In diesem Fall ist die Interpolation aber nicht durch eine Faltung zu bewerkstelligen, da die zu interpolierende Zwischenposition von Bildpunkt zu Bildpunkt variiert, womit die Faltungsmaske nicht verschiebungsinvariant ist.

# Kapitel 7

## Tiefe aus Diffusion

In diesem Kapitel soll eine neues Verfahren zur Entfernungsbestimmung vorgestellt werden. Dabei wird der in Kapitel 5 beschriebene Strukturtensor in seiner erweiterten Variante benutzt, um die Veränderungen in der Unschärfe im Verlauf der Tiefenserie direkt zu schätzen. Da mit dem erweiterten Strukturtensor neben der Diffusionskonstanten auch die Bewegung in den Bildern geschätzt werden kann, ist es möglich, diese Methode auch bei nicht telezentrischen Optiken, d.h. bei Fokusssequenzen mit Skalierungseffekten einzusetzen. In Abschnitt 7.1 wird zunächst gezeigt, daß man Fokussereien als Diffusionsprozeß betrachten kann und welchen Wert die Diffusionskonstante dabei annimmt. Anschließend werden in Abschnitt 7.2 die Objektentfernungen aus der Diffusionskonstante bestimmt und die Ergebnisse an einigen Beispielen überprüft. Eine andere Möglichkeit besteht darin, die Diffusionskonstante als Schärfemaß zu verwenden und entlang der Strahlen nach deren Minimum zu suchen, wie es in Kapitel 6 für andere Schärfemaße beschrieben wird.

### 7.1 Fokusserie als Diffusionsprozeß

In Abschnitt 6.1 wurde gezeigt, daß der Unschärfeparameter  $r$  näherungsweise linear mit der Schirmposition bzw. mit der Objektentfernung variiert. Schreibt man für die Bilder der Fokusserie  $g(x, t)$ , so bezeichnet  $t$  die kontinuierliche Bildnummer, mit der die Kameraparameter variieren. Für den Unschärfeparameter  $r(t)$  erhält man nach Gleichung (6.4):

$$r(t) = |A(Z) \cdot (t - t_0(Z))| \quad (7.1)$$

Dabei ist  $A(Z)$  die Steigung, d.h.  $A(Z)$  beschreibt, wie schnell sich der Unschärfeparameter mit dem im Verlauf der Fokusserie variierten Kameraparameter ändert. Die Position

innerhalb der Fokussierung, an der der entsprechende Objektpunkt scharf abgebildet wird, ist dabei mit  $t_0$  bezeichnet. Ändert sich im Verlauf der Fokussierung der Schirmabstand  $s = \alpha_s t$ , so ergibt sich für  $A(Z)$  und  $t_0(Z)$ :

$$A(Z) = \frac{\alpha_s}{2n_f} \frac{Z - f}{Z} \quad (7.2)$$

$$t_0(Z) = \frac{sZ}{\alpha_s} = \frac{1}{\alpha_s} \frac{Zf}{Z - f} \quad (7.3)$$

Wird dagegen der Objektabstand bzw. die Kameraposition  $Z_K = \alpha_K t$  variiert, erhält man:

$$A(Z) = \frac{\alpha_K}{2n_f} \frac{f^2}{Z_s(Z_s - f)} \quad (7.4)$$

$$t_0(Z) = \frac{1}{\alpha_K} (Z - Z_s) \quad (7.5)$$

Dabei bezeichnet  $Z_s$  die aufgrund des festen Schirmabstands  $s$  scharfgestellte Objektentfernung und  $n_f$  bzw.  $f$  sind, wie auch in den vorangegangenen Kapiteln, die Blendenzahl und die Brennweite der Kamera.

Wie aus Kapitel 2 bekannt ist, kann die Punktantwort unter anderen auch als Gaußfunktion dargestellt werden, deren Standardabweichung  $\sigma$  dabei durch den Unschärfeparameter  $r(t)$  ersetzt wird.

$$h(\mathbf{x}, t)_G = \frac{1}{2\pi r(t)^2} \exp\left(-\frac{\mathbf{x}^2}{2 r(t)^2}\right) \quad (7.6)$$

Setzt man  $r(t)$  nach Gleichung (7.1) ein ergibt sich:

$$h(\mathbf{x}, t)_G = \frac{1}{2\pi A^2(t - t_0)^2} \exp\left(-\frac{\mathbf{x}^2}{2 A^2(t - t_0)^2}\right) \quad (7.7)$$

Die Fokussierung  $g(\mathbf{x}, t)$ , die sich ergibt, wenn man nur einen einzelnen Objektpunkt aufnimmt, entspricht genau der Punktantwort  $h(\mathbf{x}, t)_G$ . Ähnlich wie in Abschnitt 5.3.1 läßt die daraus die mit dem erweiterten Struktortensor geschätzte Diffusionskonstante  $D$  berechnen. Ohne zusätzliche Bewegung, d.h. ohne Skalierungseffekte durch die sich ändernden Kameraparameter, ergibt sich:

$$D = \frac{\partial_t g}{\Delta g} = \frac{\partial_t g}{\partial_x \partial_x g + \partial_y \partial_y g} \quad (7.8)$$



Die partiellen Ableitungen der Fokusserie  $g(\mathbf{x}, t) = h_G(\mathbf{x}, t)$  lauten:

$$\partial_t g(\mathbf{x}, t) = \left[ -\frac{2}{t-t_0} + \frac{x^2}{A^2(t-t_0)^3} \right] g(\mathbf{x}, t) \quad (7.9)$$

$$\partial_x \partial_x g(\mathbf{x}, t) = \left[ -\frac{1}{A^2(t-t_0)^2} + \frac{x^2}{A^4(t-t_0)^4} \right] g(\mathbf{x}, t) \quad (7.10)$$

$$\partial_y \partial_y g(\mathbf{x}, t) = \left[ -\frac{1}{A^2(t-t_0)^2} + \frac{y^2}{A^4(t-t_0)^4} \right] g(\mathbf{x}, t) \quad (7.11)$$

Woraus für die Diffusionskonstante  $D$  folgt:

$$D = \frac{\partial_t g}{\partial_x \partial_x g + \partial_y \partial_y g} = A^2(t-t_0) \quad (7.12)$$

Die geschätzte Diffusionskonstante ist also proportional zu  $t - t_0$ , dem Abstand von dem Bild der Fokusserie, in dem der entsprechende Objektpunkt scharf abgebildet wird. Die Annahme, die dem erweiterte Strukturtensor zugrunde liegt, daß die zu schätzenden Parameter in einer kleinen lokalen Umgebung konstant sind, ist also nicht exakt erfüllt. Geht man allerdings davon aus, daß sich die Diffusionskonstante im Verlauf der Fokusserie nur langsam ändert, kann diese Methode trotzdem angewandt werden. In einer möglichen Erweiterung kann z.B. ein linearer Verlauf von  $D$  modelliert werden.

## 7.2 Schätzung der Tiefe aus der Diffusionskonstanten

Gleichung (7.12) kann nun benutzt werden, um aus der gemessenen Diffusionskonstanten  $D$  die Objektentfernung  $Z$  zu bestimmen. Für eine Fokusserie, in deren Verlauf der Kameraabstand  $Z_K$  variiert wird, setzt man dazu  $t_0$  aus Gleichung (7.5) in Gleichung (7.12) ein und löst nach  $Z$  auf.

$$Z = Z_s + Z_k - \frac{\alpha_K}{A^2} D \quad \text{mit} \quad A = \frac{\alpha_K}{2n_f} \frac{f^2}{Z_s(Z_s - f)} \quad (7.13)$$

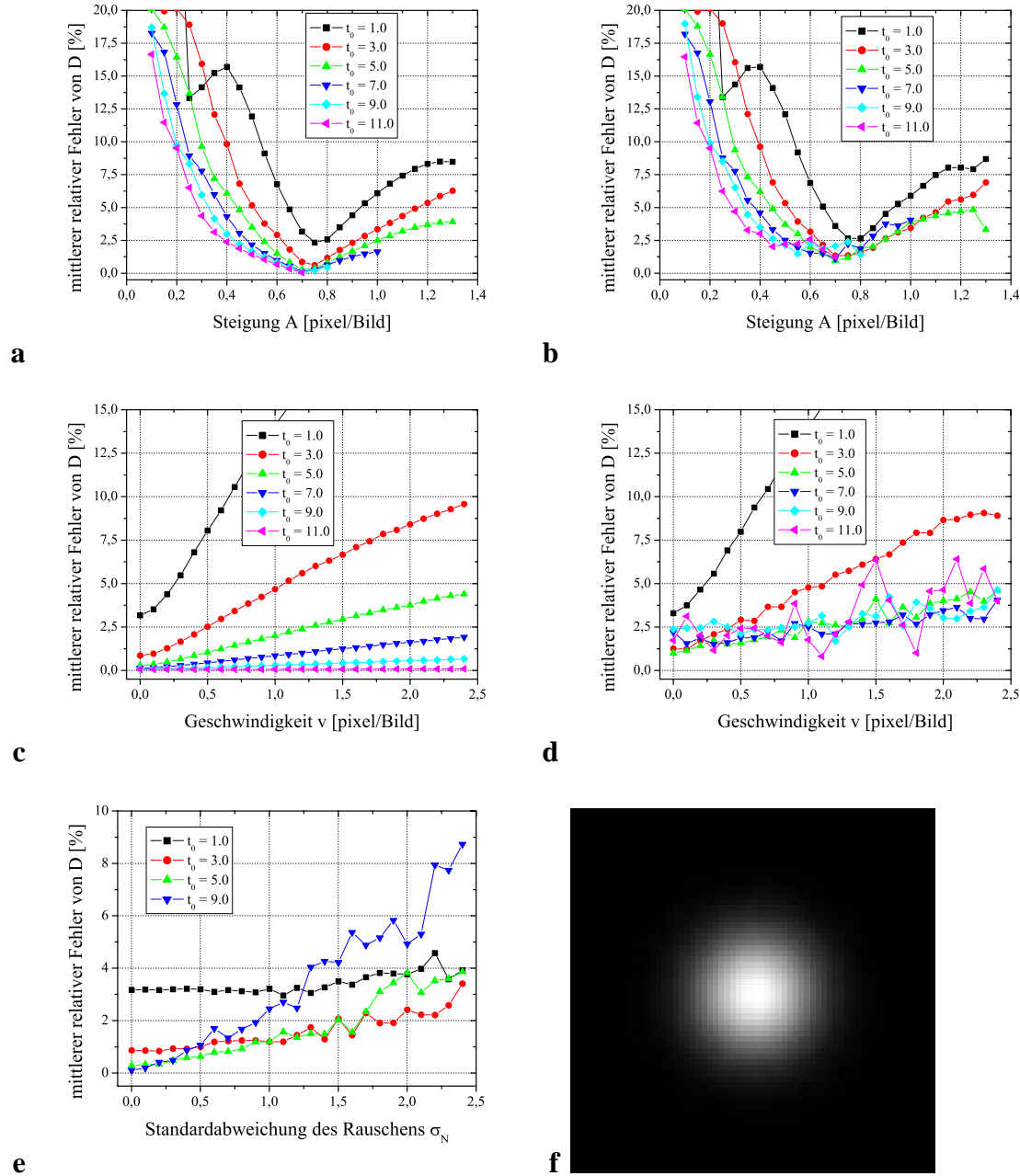
Variiert dagegen die Schirmposition  $s$ , so ist es zunächst sinnvoll  $A(Z)$  aus Gleichung (7.1) für  $Z \gg f$  zu nähern:

$$A(Z) = \frac{\alpha_s}{2n_f} \frac{Z - f}{Z} \approx \frac{\alpha_s}{2n_f} \quad (7.14)$$

Damit ist  $A$  unabhängig von  $Z$ . Setzt man nun  $t_0$  aus Gleichung (7.3) in Gleichung (7.12) ein und löst nach  $s_Z$  auf, erhält man:

$$s_Z = s - \frac{\alpha_s}{A^2} D \quad \text{mit} \quad A \approx \frac{\alpha_s}{2n_f} \quad (7.15)$$

Woraus sich die Objektentfernung leicht mittels  $Z = s_Z f / (s_Z - f)$  berechnen läßt.



**Abbildung 7.1:** Mittlerer relativer Fehler der mit dem erweiterten Strukturtenor geschätzten Diffusionskonstante  $D$  für verschiedene  $t_0$  **a** aufgetragen gegen die Steigung  $A$  für  $\sigma_N = 0.0$  und **b** für  $\sigma_N = 1.0$ . **c** aufgetragen gegen die Geschwindigkeit  $v$  für  $\sigma_N = 0.0$  und **d** für  $\sigma_N = 1.0$ . **e** aufgetragen gegenüber der Standardabweichung des Rauschens  $\sigma_N$  für  $A = 0.7$  Pixel/Bild. **f** zeigt ein Bild der zur Berechnung verwendeten Fokussierier.

Im Folgenden soll der mittlere relative Fehler in der Schätzung von  $D$  betrachtet werden. Dafür wurden mit Gleichung 7.7 Fokussereien von einem einzelnen Objektpunkt für unterschiedliche Werte von  $A$  und  $t_0$  synthetisiert. Zudem wurden die Fokussereien mit mittelwertfreiem normalverteiltem Rauschen mit der Standardabweichung  $\sigma_N$  versehen. Anschließend wurden die Sequenzen mit dem erweiterten Strukturtenor ausgewertet und die gemessene Diffusionskonstante mit der nach Gleichung (7.12) berechneten verglichen. In Abbildung 7.1a und b ist der so berechnete mittlere relative Fehler gegen die Steigung  $A$  für unterschiedliche  $t_0$  aufgetragen, wobei für Teilbild a kein Rauschen addiert wurde, während für Teilbild b die Standardabweichung des addierten Rauschens  $\sigma_N = 1.0$  beträgt. Es ist deutlich zu erkennen, daß der Fehler bei  $A \approx 0.7$  ein Minimum hat und dies relativ unabhängig von den Werten für  $t_0$  und  $\sigma_N$ . Wenn möglich, muß bei einer Messung  $A$  in der Nähe dieses Minimums liegen. Nähert man  $A$  aus Gleichung (7.4) für  $Z \gg f$  erhält man:

$$A = \frac{\alpha_K}{2n_f} \frac{f^2}{Z_s^2} \quad (7.16)$$

Mit den realistischen Werten, siehe Kapitel 9:

$$\begin{aligned} \text{Bildabstand } \alpha_K &= 15 \text{ mm,} \\ \text{Blendenzahl } n_f &= 1.4, \\ \text{Brennweite } f &= 25 \text{ mm und} \\ \text{Beobachtungabstand } Z_s &= 800 \text{ mm} \end{aligned}$$

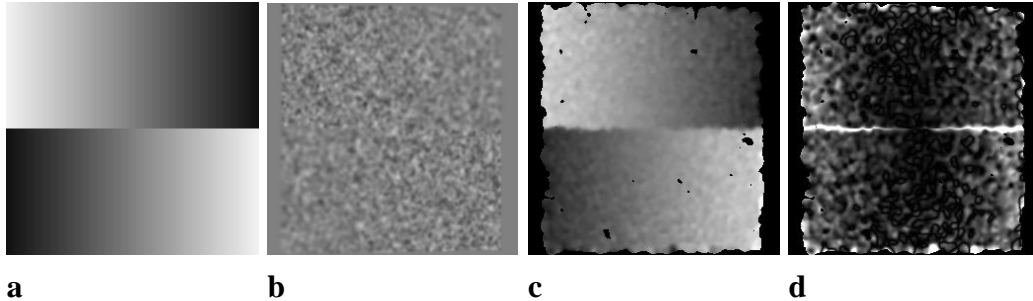
erhält man z.B. ein  $A$  von  $5,232 \cdot 10^{-3}$  mm/Bild. Berücksichtigt man noch die Größe der Bildpunkte von  $7.5 \mu\text{m}$  erhält man  $A = 0,698$  Pixel/Bild, was sehr nah am Minimum liegt.

Weiterhin ist zu erkennen, daß das Minimum für kleinere  $t_0$  schärfer ist, d.h. man sollte nicht direkt an dem Bild in der Fokusserie messen, an dem der entsprechende Objektpunkt scharf abgebildet wird ( $t_0 = 0$ ), sondern eher einige Bilder weiter entfernt davon ( $t_0 = 5$ ).

In den Teilbildern 7.1c und d ist der Fehler gegen die Geschwindigkeit  $v$  aufgetragen, wobei die Standardabweichung des Rauschens wieder  $\sigma_N = 0.0$  bzw.  $\sigma_N = 1.0$  beträgt. Die Steigung  $A$  ist mit  $A = 0,7$  Pixel/Bild in die Nähe des Minimums gelegt. Solche zusätzlichen Bewegungen treten auf, wenn man eine nicht telezentrische Optik verwendet, wodurch die Bilder im Verlauf der Fokusserie zusätzlich skaliert werden. Der dargestellte Fehler steigt mit der Geschwindigkeit an, wobei die Steigung dieses Anstiegs maßgeblich von  $t_0$  abhängt. Wie schon in Abschnitt 5.3.1 läßt sich das damit erklären, daß die Gaußkurve für kleine  $t_0$  recht scharf/steil ist; sie enthält damit hohe Wellenzahlen, die auch schon bei kleinen Geschwindigkeiten zu Störungen führen.

Abschließend ist der Fehler in Teilbild 7.1e gegen die Standardabweichung des Rauschens

aufgetragen. Zu bemerken ist hier, wie auch in Teilbild d, daß hohe  $t_0$  stärker auf das Rauschen reagieren. Da die Gaußkurven für höhere  $t_0$  flacher sind, ist für sie das Signal-zu-Rausch-Verhältnis schlechter, was zu diesem Effekt führt.



**Abbildung 7.2:** Beispiel einer synthetischen Fokussiererei: **a** die zu Grunde liegende Tiefenkarte (schwarz=705mm, weiß=735mm), **b** ein Originalbild der Sequenz, **c** die berechnete Tiefenkarte (schwarz=705mm, weiß=735mm) und **d** der absolute Fehler der berechneten Tiefenkarte (schwarz=0mm, weiß= $\pm 5$ mm).

Die Ergebnisse der Auswertung einer synthetischen Fokussiererei sind in Abbildung 7.2 zu sehen. Die Berechnung der Fokussiererei ist in Abschnitt 8.1 beschrieben. Hier erfolgte sie ausgehend von der in Teilbild a dargestellten Tiefenkarte mit einem Beobachtungsabstand von  $Z_s = 719.44$ mm, einer Brennweite von  $f = 25.0$ mm und einer Blendenzahl von  $n_f = 1.0$ . Als Oberflächentextur diente dabei ein Rauschmuster, wie an dem Originalbild in Teilbild b zu sehen ist. In Teilbild c und d ist die mittels Gleichung (7.15) aus der Diffusionskonstante berechnete Tiefenkarte und deren Abweichung von der zu Grunde liegenden Tiefenkarte zu sehen. Es fällt auf, daß der Fehler zum rechten und linken Rand leicht ansteigt und an der horizontalen Tiefenkante deutlich erhöht ist. Die Erhöhung des Fehlers an der Tiefenkante ist darauf zurückzuführen, daß die lokale Konstanz der Diffusionskonstante an der Kante massiv verletzt ist. Eine Verbesserung an solchen Kanten könnte durch die Verwendung von robuster Statistik, wie sie von Garbe [2001] zur Detektion von Reflexen benutzt wurde, erreicht werden.

## **Teil III**

# **Messungen und Auswertung**



# Kapitel 8

## Synthetische Sequenzen

### 8.1 Berechnung der Sequenzen

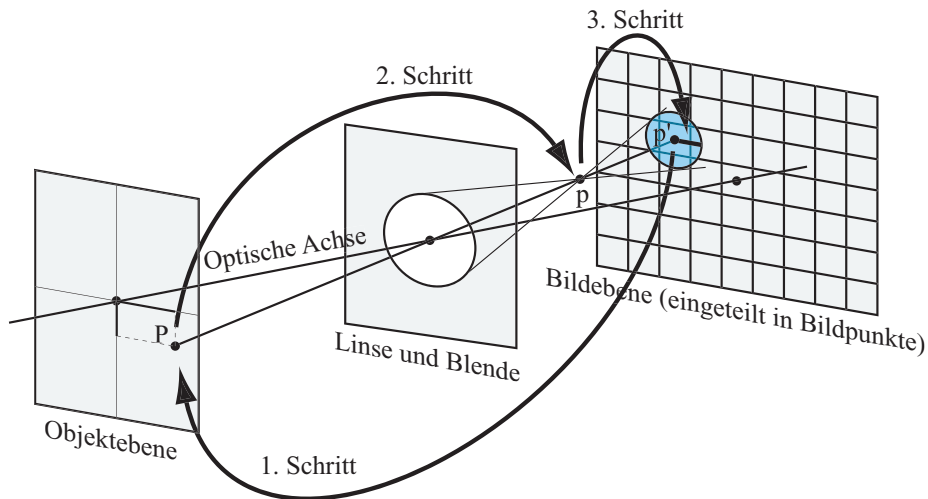
Die Berechnung der Fokusequenzen erfolgt in einzelnen Bildern, wobei die entsprechenden Kameraparameter von Bild zu Bild variiert werden. Dadurch können verschiedenste Parameter auf einfache Weise verändert werden, ohne daß dazu die Berechnungsroutine geändert werden müßte. In den folgenden beiden Abschnitten werden die Arbeitsschritte beschrieben, in die die Berechnung des Einzelbilder unterteilt werden kann. Im ersten Schritt wird der Mittelpunkt und der Unschärfeparameter bzw. der Unschärferadius der Punktantwort für jeden Objektpunkt berechnet. Anschließend wird die entsprechend skalierte Punktantwort in das zu synthetisierende Bild eingearbeitet.

#### 8.1.1 Geometrie

Die naheliegende Vorgehensweise aus den Objektpunkten die Bildpunkte zu berechnen, hat den entscheidenden Nachteil, daß die Punktantworten der einzelnen Objektpunkte, die sich in der Bildebene überlagern, nicht gleichmäßig dicht in dieser verteilt sind. Da das Integral über die gesamte PSF auf eins normiert ist, führen Variationen in der Dichte der PSFs auch bei konstant hellen Objekten zu unterschiedlichen Intensitäten in der Bildebene. In Vortests hat sich gezeigt, das diese Intensitätsschwankungen nur unzureichend durch eine Dichtekorrektur kompensiert werden können.

Daher wird im Folgenden der umgekehrte Weg beschritten. Wie in Abbildung 8.1 skizziert, wird in einem ersten Schritt - ausgehend von jedem Pixel der Bildebene - der entsprechende Objektpunkt berechnet. Dazu wird der Strahl durch den jeweiligen Pixelmittelpunkt  $p'$  und

das Zentrum der Linse bzw. Blende mit den abzubildenden Objekten, z.B. einer Ebene, geschnitten. Man erhält neben den Weltkoordinaten des Objektpunkts  $\mathbf{P} = [XYZ]^T$  auch die Koordinaten  $[\lambda_X \lambda_Y]^T$  in der Ebene bezüglich der sie aufspannenden Vektoren, die später zur Bestimmung der Helligkeit des Objektpunkts herangezogen werden.



*Abbildung 8.1: Skizze zur Berechnung der synthetischen Fokussierien.*

Im zweiten Schritt wird für den Objektpunkt  $P$  mit der Gleichung für dünne Linsen (2.2) sein scharf abgebildeter Bildpunkt  $p$  berechnet. Aus dessen Abstand von der Linse, dem Abstand der Bildebene und dem Blendenradius kann im dritten und letzten Schritt der Unschärfeparameter bzw. Unschärfekreis berechnet werden.

### Telezentrische Optiken

Sollen Fokussierien ohne Skalierungseffekte, wie sie mit telezentrischen Optiken zu erreichen sind, synthetisiert werden, wird nur der erste Schritt zur Berechnung des zu jedem Pixel gehörenden Objektpunkts modifiziert; die weitere Vorgehensweise bleibt gleich. Anstatt den Strahl vom Pixelmittelpunkt durch den Mittelpunkt der Linse zu legen, verläuft er nun parallel zur optischen Achse durch den Pixelmittelpunkt. Man erhält statt der obigen Zentralprojektion eine Parallelprojektion.

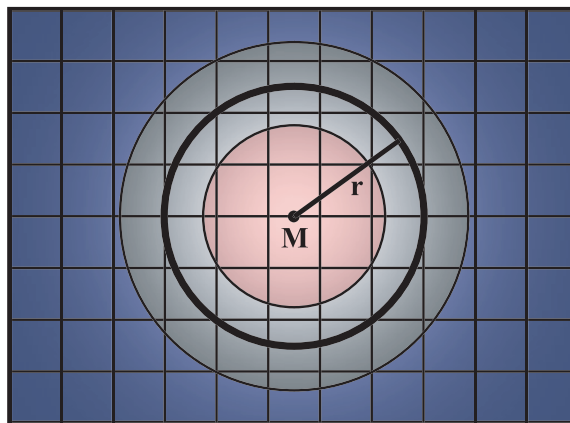
### 8.1.2 2D-Punktantwort

Um das endgültige Bild zusammensetzen zu können, fehlen noch zwei Dinge: Zum einen muß der Helligkeitswert jedes Objektpunkts bestimmt werden und zum anderen ist noch



die Punktantwort für den jeweiligen Unschärfeparameter zu berechnen. Die entsprechend skalierte Punktantwort wird dann mit dem Helligkeitswert multipliziert und ins Bild addiert, was im Prinzip einer Faltung mit einer ortsabhängigen Maske entspricht.

Der Helligkeitswert jedes Objektpunkts läßt sich recht einfach durch seine Koordinaten  $[\lambda_X \lambda_Y]^T$  innerhalb der Ebene<sup>1</sup> bestimmen. Sind die Basisvektoren, die die Ebene aufspannen, senkrecht zueinander und gleich lang, läßt sich der Helligkeitswert direkt aus  $[\lambda_X \lambda_Y]^T$  und einer 2D-Funktion, die die Helligkeitsverteilung beschreibt, verzerrungsfrei berechnen. Steht statt der funktionalen Beschreibung der Helligkeitsverteilung eine Rastergrafik zur Verfügung, kann die Helligkeit z.B. mittels bilinearer Interpolation berechnet werden.

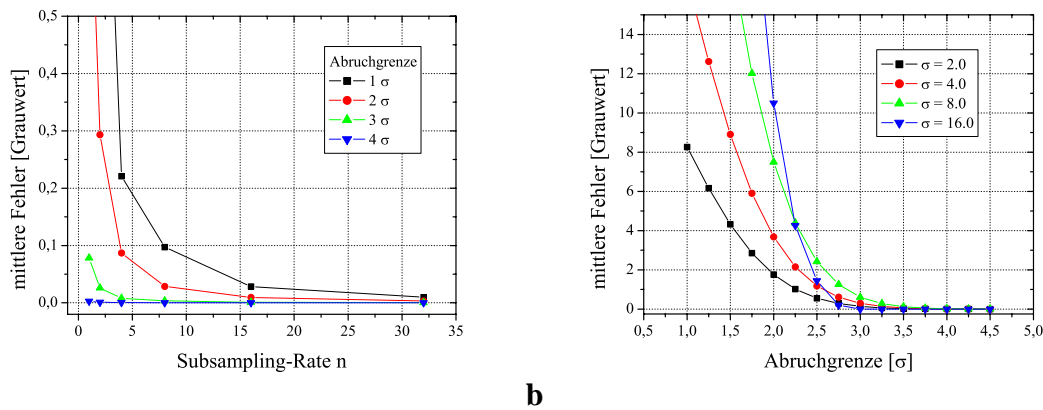


**Abbildung 8.2:** Veranschaulichung der Berechnung der Punktantwort. Während für die im roten Bereich liegenden Pixel nur eine Funktionsauswertung nötig ist, müssen die Pixel im anschließenden grauen Bereich in Subpixel unterteilt werden. Pixel aus dem blauen Bereich bleiben unverändert gleich Null.

Bei den noch zu berechnenden PSF handelt es sich um die aus Kapitel 2 bekannte Gauß- bzw. Kreisscheiben-Funktion. Sie müssen für unterschiedliche Standardabweichungen bzw. Radien möglichst effizient und genau berechnet werden. Es ergibt sich dabei das Problem, daß die eine Kreisscheibe mit beliebigem Radius nicht korrekt im diskreten Raster dargestellt werden kann. Ähnlich verhält es sich bei der Gaußfunktion; sie ist zwar unendlich ausgedehnt, wird aber sinnvollerweise nach einigen Standardabweichungen abgebrochen. Diese Abbruchgrenze der Gaußfunktion und der Rand der Kreisscheibe können durch *Subsampling* angenähert werden. Dabei werden die Pixel des Bildes in  $n^2$  Subpixel unterteilt, für jeden Subpixel wird der Funktionswert berechnet und die Ergebnisse werden dann anteilig zum Gesamtpixel zusammengefaßt. Um die Rechenzeit dabei nicht zu stark ansteigen zu lassen, wurden die Pixel, die vollständig inner- bzw. außerhalb der Grenze bzw. des Ra-

<sup>1</sup>In der Computergrafik werden solche Koordinaten häufig als *Texturkoordinaten* bezeichnet.

dus liegen, bestimmt. Für sie ist ein Subsampling nicht notwendig, da alle ihre Subpixel auf der gleichen Seite der Grenze liegen. In Abbildung 8.2 ist das Vorgehen verdeutlicht: Hat die Abbruchgrenze den Radius  $r$ , so befinden sich alle Pixel mit einem Abstand ihres Pixelmittelpunkts zum Mittelpunkt  $M$  von  $r' < r - \sqrt{2}/2$  vollständig innerhalb und alle mit einem Abstand von  $r' > r + \sqrt{2}/2$  vollständig außerhalb. Diese Bereiche sind in Abbildung 8.2 rot und blau eingefärbt; nur für die Pixel, deren Mittelpunkt im nicht eingefärbten Bereich liegt, wird das Subsampling durchgeführt.



**Abbildung 8.3:** Mittlere Grauwertabweichung der berechneten gaußförmigen PSF: **a** aufgetragen gegen die lineare Unterteilung  $n$ , die entsprechenden Pixel werden damit in  $n^2$  Subpixel unterteilt. Die verschiedenen Kurven stehen für unterschiedliche Abbruchgrenzen von eins bis vier  $\sigma$ . **b** aufgetragen gegen die Abbruchgrenze in Vielfachen des Unschärfeparameters  $\sigma$ . Die verschiedenen Kurven zeigen PSFs mit unterschiedlichem Unschärfeparameter  $\sigma$ .

In Abbildung 8.3 ist in Teilbild (a) der mittlere Grauwertfehler gegen die Subsampling-Unterteilung für verschiedene Abbruchgrenzen aufgetragen. Der Fehler ist dabei die Abweichung von einer Referenz-PSF, die mit einer sehr hohen Subsampling-Unterteilung von  $n = 128$  berechnet wurde. Daraus geht hervor, daß eine Subsampling-Unterteilung zwischen  $n = 4$  und  $n = 16$  je nach Abbruchgrenze sinnvoll ist. Bei höheren Abbruchgrenzen ist die Gaußfunktion an dieser Grenze schon stark abgefallen und der Gewinn durch besseres Subsampling fällt nicht mehr so hoch aus. In Teilbild (b) ist der Fehler gegen die Abbruchgrenze für verschiedene PSF-Größen aufgetragen. Als Referenz-PSF diente hier eine PSF, deren Berechnung erst an den Bildgrenzen abgebrochen wurde. Eine Abbruchgrenze zwischen  $3\sigma$  und  $4\sigma$  ist ein guter Kompromiß aus Maskengröße und Fehler.

## Normierung

Trotz des Subsamplings sind die Fehler durch die Diskretisierung noch so groß, daß die Normierung mittels Division durch den integralen Wert der PSF, also

$$\sqrt{2\pi}\sigma \quad \text{für} \quad \exp -\frac{\mathbf{x}^2}{2\sigma^2} \quad \text{bzw.} \quad (8.1)$$

$$\pi r^2 \quad \text{für} \quad \Pi\left(\frac{|\mathbf{x}|}{r}\right) \quad (8.2)$$

nicht ausreichte. Daher wird vor der Berechnung der eigentlichen Maske des PSF mit dem gleichen Verfahren die Summe der Grauwerte berechnet. Diese wird bei der endgültigen Berechnung als Normierungsfaktor benutzt.

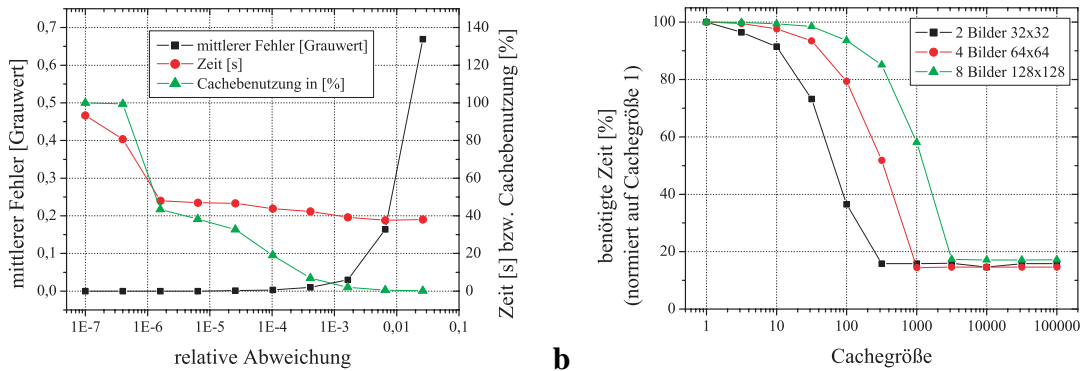
## PSF-Cache

Bedenkt man, daß die PSF in jedem Bild der Serie für jeden Bildpunkt berechnet werden muß und daß eine PSF insbesondere bei größeren Gaußmasken bis zu  $64 \times 64$  Bildpunkte umfassen kann, wird klar, daß weitere Optimierungen notwendig sind.

$$16 \text{ Bilder} \times 512^2 \text{ Bildpunkte} \times 64^2 \text{ Maskenpunkte} = 16 \text{ Mrd. Funktionsauswertungen} \quad (8.3)$$

Daher wurde ein Puffer-Speicher benutzt, in dem die bisher schon berechneten Punktantworten zusammen mit den zu ihrer Erstellung notwendigen Parametern gespeichert werden. Wird eine Punktantwort benötigt, wird ihr Parametersatz zuerst im Puffer-Speicher gesucht, ist dort eine passende PSF vorhanden, wird diese benutzt, anderenfalls wird die PSF neu berechnet und im Puffer gespeichert. Um den Puffer-Speicher nicht zu groß werden zu lassen, was neben dem hohen Speicherbedarf auch die Suchzeit verlängern würde, müssen die Parametersätze bei der Suche nicht exakt identisch sein. Die erlaubte Abweichung wird so eingestellt, daß sich ein sinnvoller Kompromiß aus Rechenzeitbedarf und dem durch die Abweichung entstehenden Fehler ergibt.

In Abbildung 8.4a ist die Grauwertabweichung der PSF-Masken, der Zeitbedarf für deren Berechnung und der Grad der Cachebenutzung gegen die maximal zulässige relative Abweichung der Parametersätze aufgetragen. Man sieht wie der Grauwertfehler zurückgeht und der Zeitbedarf sowie die Cachebenutzung zunächst moderat ansteigen, wenn die Anforderungen an die Parametersätze gesteigert werden. Ab einem gewissen Punkt ist der Cache voll und der Zeitbedarf steigt sprunghaft an. Dies ist auch in **b** zu erkennen. Hier ist der normierte Zeitbedarf zur Berechnung einer Fokuserie unterschiedlicher Größe gegen die Cachegröße aufgetragen. Ab einer bestimmten Cachegröße sinkt der Zeitbedarf



**Abbildung 8.4:** Der Einfluß der maximal zulässigen Abweichung zwischen den Parametersätzen der angeforderter PSF und der im Cache gefundenen auf die Grauwertabweichung der PSF-Masken, den Grad der Cachebenutzung und der Zeitbedarf ist in (a) dargestellt. b zeigt den Zeitbedarf zur Erzeugung unterschiedlich großer Fokusssequenzen in Abhängigkeit von der Cachegröße.

drastisch, was darauf zurückzuführen ist, daß nun alle berechneten PSFs für den späteren Wiedergebrauch zwischengespeichert werden können. Vor dieser Schwelle wird ein Teil der berechneten PSFs gleich nach ihrer Benutzung verworfen, da der Cache zu klein ist, um sie zu speichern. Deutlich ist zu erkennen, daß diese Schwelle mit der Größe der zu berechnenden Fokusssequenzen steigt. Damit der Cache effektiver arbeiten kann, muß er mindestens 10000 bis 100000 PSFs speichern können.

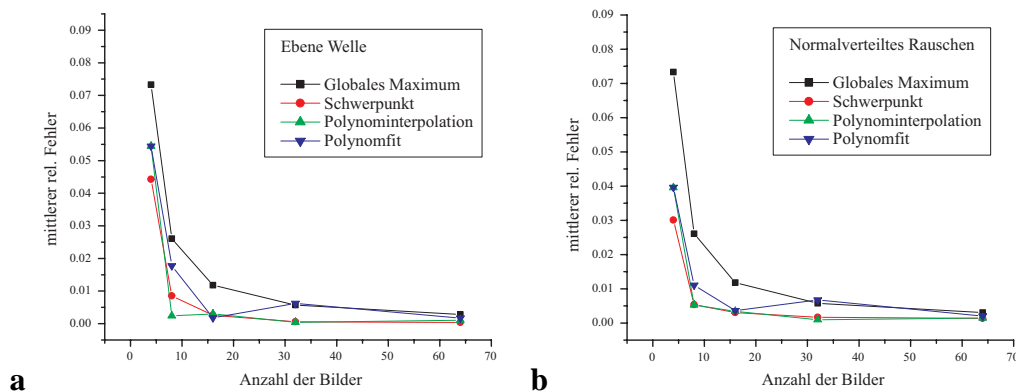
## 8.2 Test an synthetischen Sequenzen ohne Skalierung

Zur Analyse der in Kapitel 6 vorgestellten unterschiedlichen Algorithmen, werden Fokusssequenzen mit variierenden Kameraparametern erstellt. Diese werden anschließend von den verschiedenen Algorithmen analysiert und die daraus berechnete Tiefekarte wird mit der der Synthese der Fokusssequenzen zu Grunde liegenden Tiefekarte verglichen. Im Folgenden finden sich nun die Schaubilder, die den relativen Fehler jedes Algorithmus in Abhängigkeit von Parametern wie z.B. Rauschen, Bildabstand, usw. darstellen.

Als Helligkeitsverteilung wurde, falls nicht explizit anders angegeben, im jeweils linken Teilbild (a) eine ebene Welle mit einer Wellenlänge von 14 Bildpunkten und einer Amplitude von 100 Grauwerten benutzt, im rechten Teilbild (b) normalverteiltes Rauschen mit einer Standardabweichung von 100 Grauwerten. Das Rauschbild wurde anschließend geglättet um die hohen Frequenzanteile abzdämpfen. Als vorgegebene Tiefekarte wurde eine schiefe Ebene benutzt, deren minimale Tiefe 25% und deren maximale Tiefe 75% der Tiefe des Gesamtvolumens von  $100 \times 100 \times 20$  Bildpunkten beträgt.

### 8.2.1 Bildanzahl / -abstand

In Abbildung 8.5 ist zunächst der mittlere relative Fehler gegen die Anzahl der Bilder der Fokussiererei aufgetragen. Es ist deutlich zu erkennen, daß die einfache Maximumsuche aus Abschnitt 6.3 den höchsten Fehler aufweist, während die Schwerpunkt-Methode bzw. die lokale Polynominterpolation ungefähr gleich niedrige Fehler produzieren. Der grundsätzliche Trend, daß der Fehler mit steigender Bildanzahl fällt, ist bei allen Verfahren zu beobachten.



**Abbildung 8.5:** Der mittlere relative Fehler der Tiefenkarte in Abhängigkeit von der Dichte bzw. der Anzahl der Bilder der Fokussiererei.

### 8.2.2 Rauschabhängigkeit

Eine ähnliche Situation ist in Abhängigkeit vom Bildrauschen zu sehen, siehe Abbildung 8.6. Die einfache Maximumsuche weist den höchsten Fehler auf, während die Schwerpunkt-Methode und die lokale Polynominterpolation ähnliche Fehler erzeugen. Bei allen Verfahren steigt der Fehler mit dem Rauschen leicht an, aber die Polynomfit-Methode wird sehr bald instabil und der Fehler steigt ab einem bestimmten Punkt sprunghaft an.

### 8.2.3 Numerische Apertur

Wie in Abbildung 8.7 zu sehen, zeigt die Abhängigkeit von der Numerische Apertur(NA) das aus der Theorie erwartete Verhalten. Mit steigender NA wird die Blendenöffnung größer und damit fällt die Schärfentiefe. Also fällt auch der Fehler aller Verfahren mit steigender NA ab.

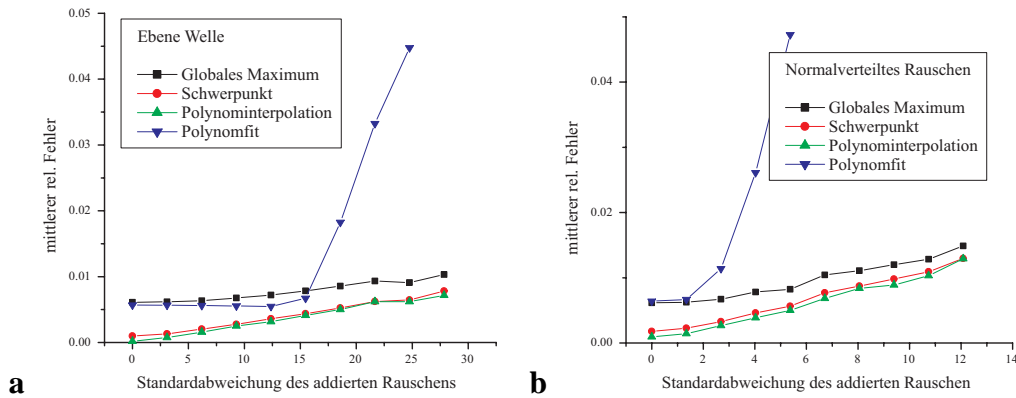


Abbildung 8.6: Der mittlere relative Fehler der Tiefenkarte in Abhängigkeit von der Standardabweichung des addierten normalverteilten Rauschens.

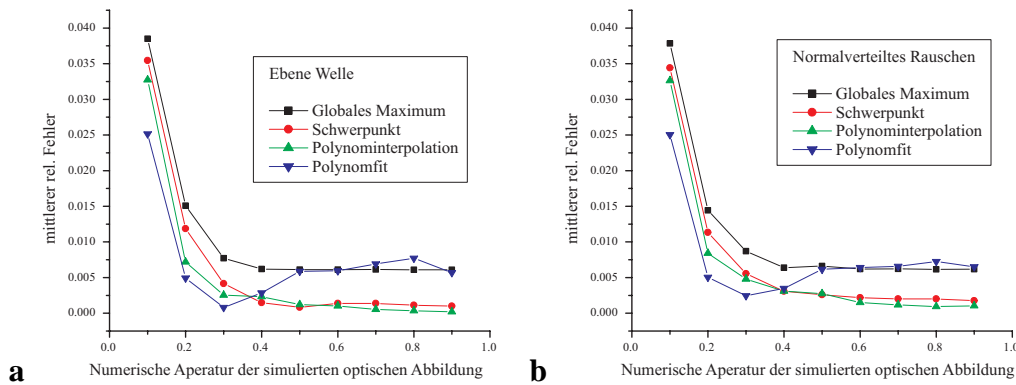
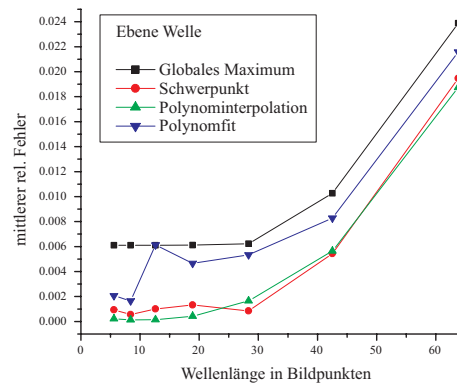


Abbildung 8.7: Der mittlere relative Fehler der Tiefenkarte in Abhängigkeit von der Blendenzahl der simulierten optischen Abbildung.

### 8.2.4 Wellenlänge der Bildstrukturen

Auch in Abhängigkeit von der Wellenlänge der Bildstrukturen zeigt sich, daß die einfache Maximumsuche den größten Fehler aufweist, wie in Abbildung 8.8 dargestellt. Der globale Polynomfit ist nur bei sehr kurzen Wellenlängen deutlich besser. Die Schwerpunkt-Methode und die lokale Polynominterpolation zeichnen sich durch den niedrigsten Fehler aus.

Zusammenfassend ergibt sich durch den erheblichen Mehraufwand des globalen Polynomfits kein geringerer Fehler. Das einfachste Verfahren, die einfache Maximumsuche, scheidet deutlich schlechter ab, als z.B. die Schwerpunkt-Methode oder die lokale Polynominterpolation. Beide Verfahren lassen sich auch inkrementell implementieren, so daß man sie der einfachen Maximumsuche vorziehen sollte.



**Abbildung 8.8:** Der mittlere relative Fehler der Tiefenkarte in Abhängigkeit von der Wellenlänge der Bildstruktur.

## 8.3 Test an synthetischen Sequenzen mit Skalierung

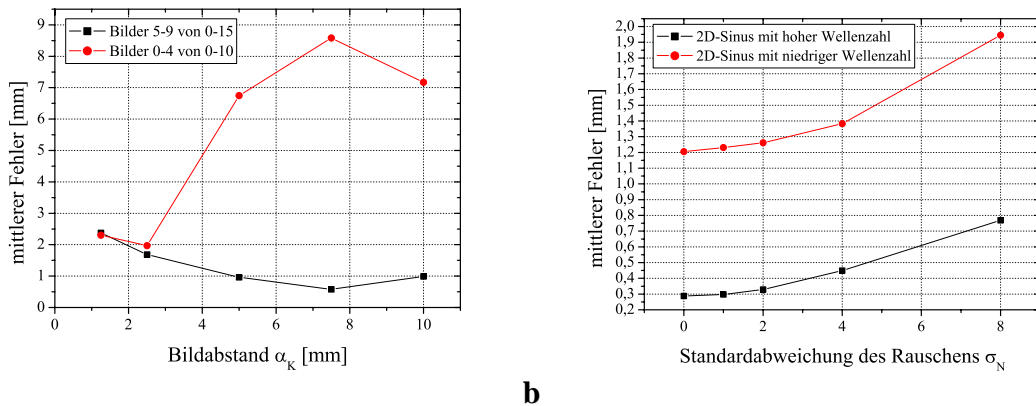
Die in Kapitel 7 beschriebene Methode zur Gewinnung von Tiefenkarten aus Fokussereien soll hier an einigen synthetischen Testsequenzen überprüft werden. Dazu soll die Abhängigkeit von verschiedenen Parametern, wie Bildrauschen und Bildabstand, sowie der Einfluß von Szenengeometrie, Oberflächentextur und der Wahl der PSF untersucht werden.

### 8.3.1 Rauschen und Bildabstand

Zuerst wurden Fokussereien aus 16 Bildern mit unterschiedlichen Bildabstand  $\alpha_k$  synthetisiert, wobei die restlichen Parameter konstant bei

Parameter	Wert
Beobachtungsabstand $Z_s$	719.44 mm
Brennweite $f$	25.00 mm
Blendenzahl $n_f$	1.0

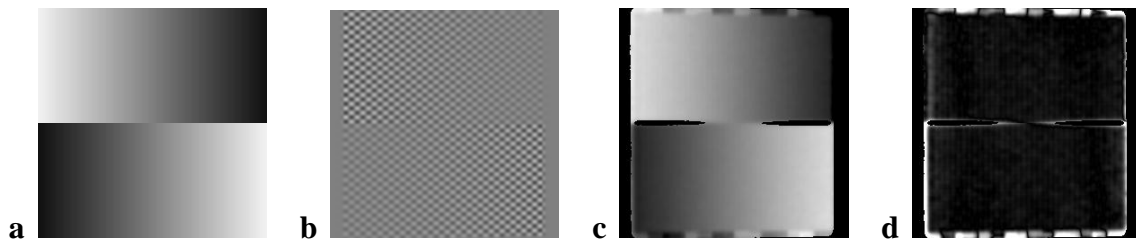
gelassen wurden. Aus den Fokussereien wurde mit der in Kapitel 7 beschriebenen Methode zuerst die Diffusionskonstante und daraus dann die Szenentiefe bestimmt. Dazu wurden jeweils die Bilder 5 bis 9 der jeweiligen Fokusserei herangezogen, außer es wird explizit auf die Verwendung anderer Bilder hingewiesen. Aus der berechneten Szenentiefe und der der Synthese zu Grunde liegenden Tiefenkarte wird schließlich der mittlere Fehler ermittelt.



**Abbildung 8.9:** Mittlerer Fehler der berechneten Tiefenkarte **a** in Abhängigkeit der Schrittweite  $\alpha_K$  und **b** in Abhängigkeit vom Bildrauschen  $\sigma_N$ .

In Abbildung 8.9a ist dieser Fehler gegen den Bildabstand  $\alpha_K$  aufgetragen. Für die schwarze Kurve, bei der die mittleren Bilder 5-9 benutzt wurden, ist ein Minimum bei  $\alpha_K \approx 7,5\text{mm}$  zu erkennen. Dies entspricht dem Minimum von  $A$  aus Abbildung 7.1a und mit den Werten der anderen Kameraparameter ergibt sich für  $A$  ein Wert von ungefähr 0.6. Nimmt man zu Auswertung allerdings die äußeren Bilder 0-4 so ergibt sich ein deutlich höherer Fehler.

Die Abhängigkeit vom Bildrauschen ist in Teilbild 8.9b zu sehen. Wie zu erwarten war, steigt der mittlere Fehler mit dem Bildrauschen an. Die unterschiedlichen Oberflächen-texturen, wie hohe bzw. niedrige Wellenzahl des hierbei verwendeten 2D-Sinusmusters, haaben allerdings einen deutlich größeren Einfluß. Für eines dieser Muster sind in Abbildung 8.10 die zu Grunde liegende Tiefenkarte, ein Originalbild, die daraus berechnete Tiefenkarte sowie der absolute Fehler dargestellt.



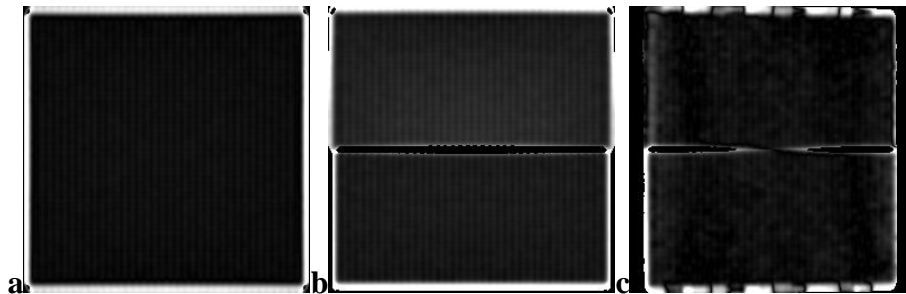
**Abbildung 8.10:** Beispiel einer synthetischen Fokussiersequenz: **a** die zugrundeliegende Tiefenkarte (schwarz=705mm, weiß=735mm), **b** ein Originalbild der Sequenz, **c** die berechnete Tiefenkarte (schwarz=705mm, weiß=735mm) und **d** der absolute Fehler der berechneten Tiefenkarte (schwarz=0mm, weiß= $\pm 5\text{mm}$ ).



### 8.3.2 Szenengeometrie

Der Einfluß der Szenengeometrie soll an drei Beispielen verdeutlicht werden. Dabei handelt es sich um

- eine zur optischen Achse senkrechte Ebene,
- zwei ebenso orientierte Ebenen, die um 20mm in Richtung der optischen Achse versetzt sind und damit eine Tiefenkante aufweisen und
- zwei Ebenen, die um 30 Grad zur Kamera hin- bzw. von ihr weggeneigt sind.

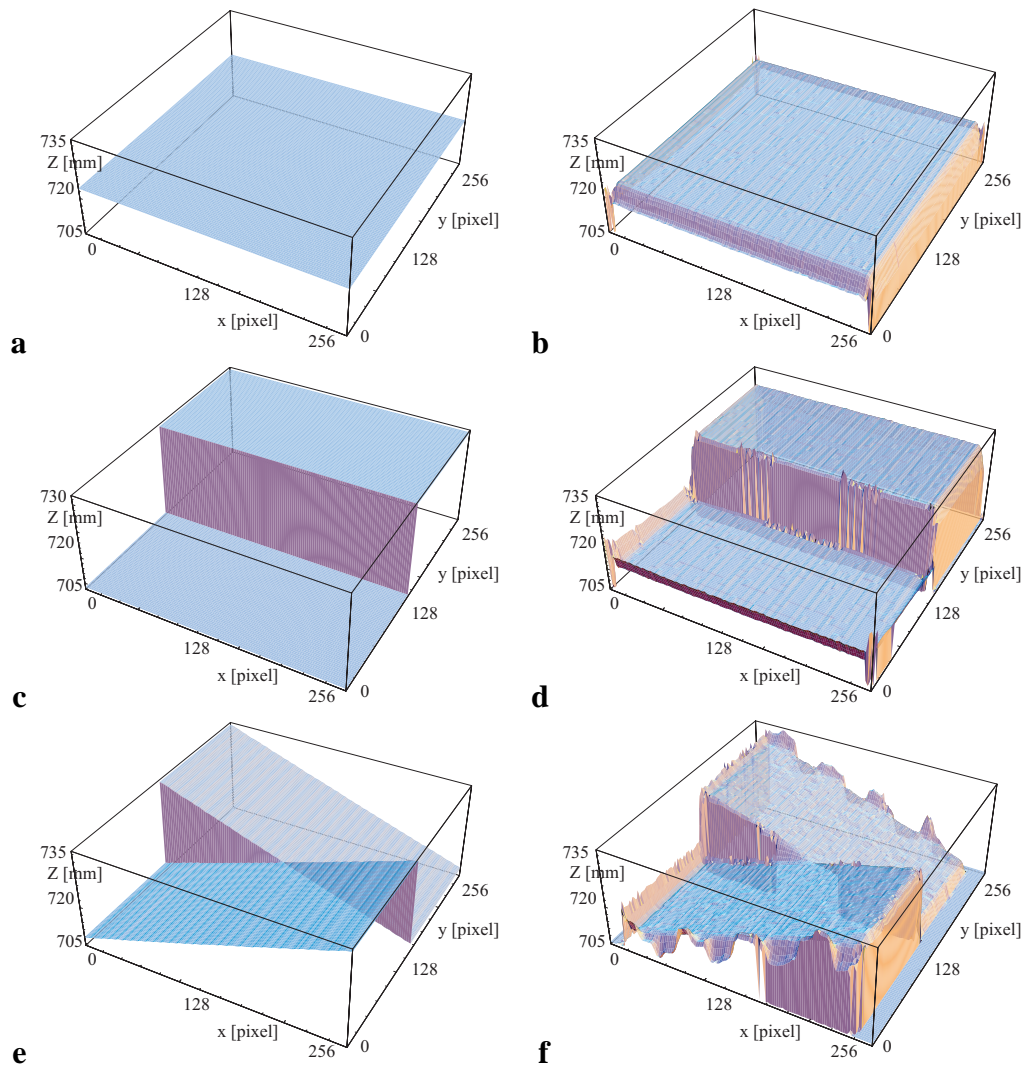


*Abbildung 8.11: Der mittlere Fehler der Tiefenkarten ist für **a** eine Ebene senkrecht zur optischen Achse, **b** zwei Ebenen senkrecht zur optischen Achse mit einem Versatz von 20mm und **c** zwei um jeweils 30 Grad geneigte Ebenen.*

In Abbildung 8.12 sind diese Szenengeometrien abgebildet, wobei die Teilbilder a,c,e die der Synthese zu Grunde liegenden Tiefenkarten zeigen, während b,d,f die berechnete Tiefeninformation darstellt. Der mittlere Tiefenfehler ist in Abbildung 8.11 zu sehen, wobei gut zu erkennen ist, daß der Fehler vor allem an den Tiefenkanten deutlich höher als in den restlichen Bereichen ist.

### 8.3.3 Oberflächentextur

Um den Einfluß der Oberflächentextur einschätzen zu können wurden Fokussereien mit unterschiedlichen Texturen erzeugt. In der folgenden Tabelle sind die mittleren Fehler der Tiefenrekonstruktion notiert:

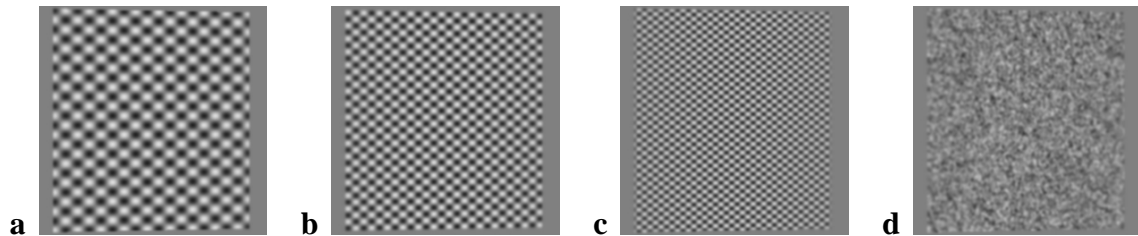


**Abbildung 8.12:** In **a** , **c** und **e** sind die der Synthese der Fokusserie zu Grunde liegenden Tiefenkarten für unterschiedliche Szenengeometrien dargestellt. Die Teilbilder **b** , **d** und **f** enthalten die jeweiligen berechneten Tiefenkarten.

Oberflächentextur	mittlerer Tiefenfehler [mm]
Rauschen	0.581
Rauschen(hochfreq.)	0.661
2D Sinus (niedrig)	1.205
2D Sinus (mittel)	0.580
2D Sinus (hoch)	0.288

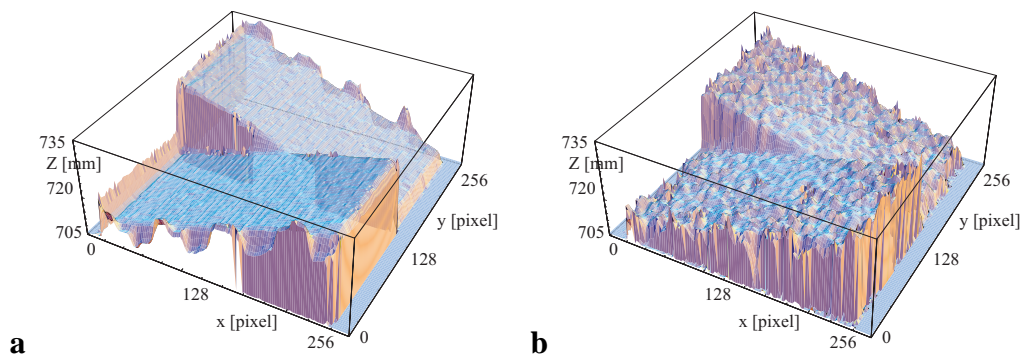
Die Bezeichnung hinter dem 2D-Sinus beschreibt die Wellenzahl. Die Fehler steigen für die 2D-Sinus-Muster sinkender Wellenzahl an, wobei die Rauschmuster ungefähr den Fehl-

ler des Sinusmusters mit der mittleren Wellenzahl aufweisen. In Abbildung 8.13 sind die unterschiedlichen Sinus-Muster und ein Rauschmuster gezeigt.



**Abbildung 8.13:** Unterschiedliche Oberflächentextur, die zur Erzeugung von synthetischen Fokusserien dienen.

In Abbildung 8.14 sind die resultierenden Tiefenkarten für das Sinus-Muster mit der hohen Wellenzahl und das Rauschmuster dargestellt. Während für das Sinus-Muster eine relative glatte Tiefenkarte berechnet wurde, ergibt sich für das Rausch-Muster eine deutlich unruhigere Tiefenkarte.



**Abbildung 8.14:** Tiefenkarten berechnet aus Fokusserien, die sich nur in der Oberflächentextur der Objekte unterscheiden: **a** 2D Sinus-Muster, **b** Rausch-Muster

### 8.3.4 PSF

Zuletzt soll noch die Abhängigkeit von der zur Synthese benutzten PSF betrachtet werden. Dazu wurde lediglich die benutzte PSF zwischen den Fokusserien ausgetauscht, alle anderen Parameter blieben gleich. In der folgenden Tabelle sind die Ergebnisse für die Gauß- und Kreisscheiben-PSF und für unterschiedliche Szenengeometrien zusammengetragen.

Geometrie	Kreisscheiben-PSF		Gauß-PSF	
	7	5	7	5
Stufe/gerade	1.098	7.860	0.853	8.537
Stufe/schräg	0.646	4.745	0.582	2.860
Ebene	0.461	4.001	0.568	2.373

Zusätzlich wurde die Tiefe noch aus unterschiedlichen Teilen der Fokusserie errechnet. Die Bildnummer in der Tabelle gibt dabei das jeweilig mittlere Bild an. Betrachtet man zunächst die Auswertung an Bild 7 der Fokusserie. Der mittlere Tiefenfehler ist hier bei der Kreisscheiben-PSF grundsätzlich leicht höher als bei der Gauß-PSF. Wird die Auswertung nun aber leicht verschoben, z.B. zu Bild 5, führt das bei der Kreisscheiben-PSF unabhängig von der Szenengeometrie zu einer deutlicheren Verschlechterung als bei der Gauß-PSF.

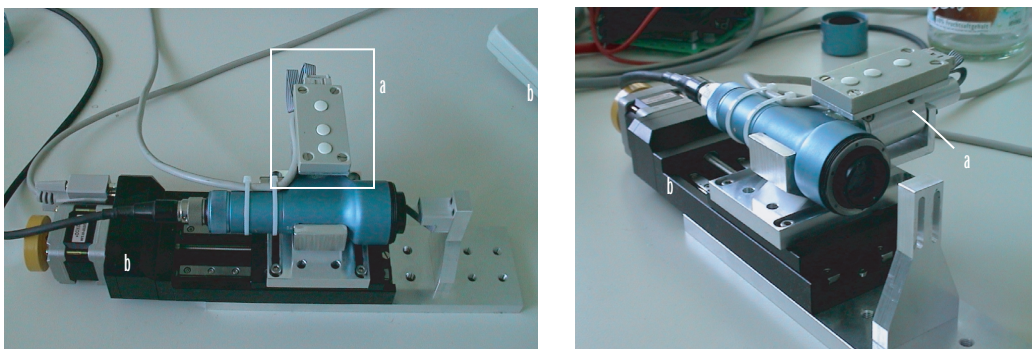
# Kapitel 9

## Reale Sequenzen

### 9.1 Telezentrische Aufnahmen (Handmikroskop)

#### 9.1.1 Der Aufbau

Das Handmikroskop wurde von der Firma A-Tec für die Aufnahmen mit einer motorgesteuerten Fokusverstellung ausgestattet. Es handelt sich dabei um einen Linear-Motor mit Impulsgeber, der über eine Datatranslation DT300 PCI-Karte vom PC gesteuert und abgefragt wird. Um im Gegensatz zu der unbekanntem Fokusverstellung durch den Motorfokus



*Abbildung 9.1: Handmikroskop mit motorgesteuertem Fokus (a) auf einem Verschiebetisch (b).*

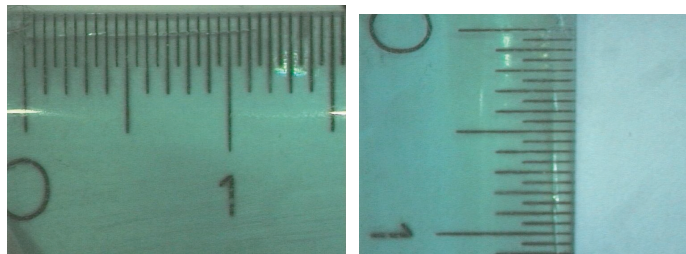
noch eine exakt bekannte Verstellung der Fokusebene zu ermöglichen, wurde das Handmikroskop auf einen Verschiebetisch montiert, der ebenfalls vom PC gesteuert werden konnte.

Das Bildsignal wurde über ein SVHS-Kabel und einen PC-Eye2 Framegrabber der Firma

Eltec an den PC übermittelt. Obwohl auf dieser hochfrequente Störungen auftraten konnte leider keine höherwertige Verbindung zwischen Handmikroskop und Framegrabber genutzt werden. Insgesamt müßte deshalb ein relativ hoher Rauschpegel mit einer Standardabweichung von 3,3 Grauwerten toleriert werden.

### 9.1.2 Auflösung des Handmikroskops

Um die Auflösung des Handmikroskops in X- und Y-Richtung zu bestimmen, wurden der Einfachheit halber zwei Bilder von einem Lineal aufgenommen.



*Abbildung 9.2: Handmikroskopaufnahmen von einem Lineal in horizontaler und vertikaler Richtung.*

In diesen wurde dann der Abstand zweier Meßstriche in Bildpunkten ermittelt.

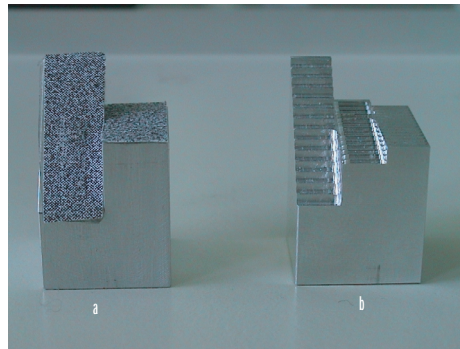
$$X: 689 \pm 1 \text{ Bildpunkt} \hat{=} 1,5 \text{ cm} \quad \rightarrow \quad 1 \text{ Bildpunkt} \hat{=} 21,77 \pm 0,03 \mu\text{m}$$

$$Y: 455 \pm 1 \text{ Bildpunkt} \hat{=} 1,0 \text{ cm} \quad \rightarrow \quad 1 \text{ Bildpunkt} \hat{=} 21,98 \pm 0,05 \mu\text{m}$$

### 9.1.3 Die verwendeten Testkörper

Es wurden zwei Testkörper vom IWR bzw. IMS hergestellt, um die Eigenschaften des Handmikroskops vermessen zu können. Bei dem ersten Testkörper (Abb. 9.3 b) vom IWR handelt es sich um drei Treppen mit Schritthöhen von 1.0, 0.25 und 0.1 mm nebeneinander. Das Material erwies sich aber als zu stark glänzend, so daß sich mit ihm keine geeigneten Aufnahmen machen ließen. Auch konnte er nicht, wie zuerst geplant, sandgestrahlt werden, da sonst seine Genauigkeit verloren gegangen wäre.

Daher wurde vom IMS in Stuttgart ein zweiter Testkörper angefertigt, der aus zwei im Winkel von 45 Grad zueinander geneigten Ebenen besteht. Auch bei ihm erwies sich das Material als stark glänzend, aber durch die stärker ausgeprägten Bearbeitungsspuren waren

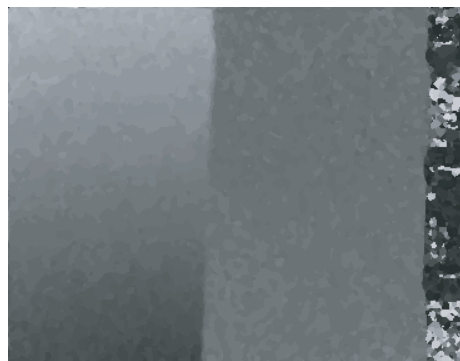


*Abbildung 9.3: Testkörper (a) vom IMS (beklebt) und (b) vom IWR.*

die so aufgenommenen Fokussereien noch auswertbar. Um bessere Oberflächeneigenschaften und Kontraste zu erhalten, wurden die Ebenen des Testkörpers für weitere Aufnahmen mit einem mit Rauschmustern bedruckten Papier beklebt.

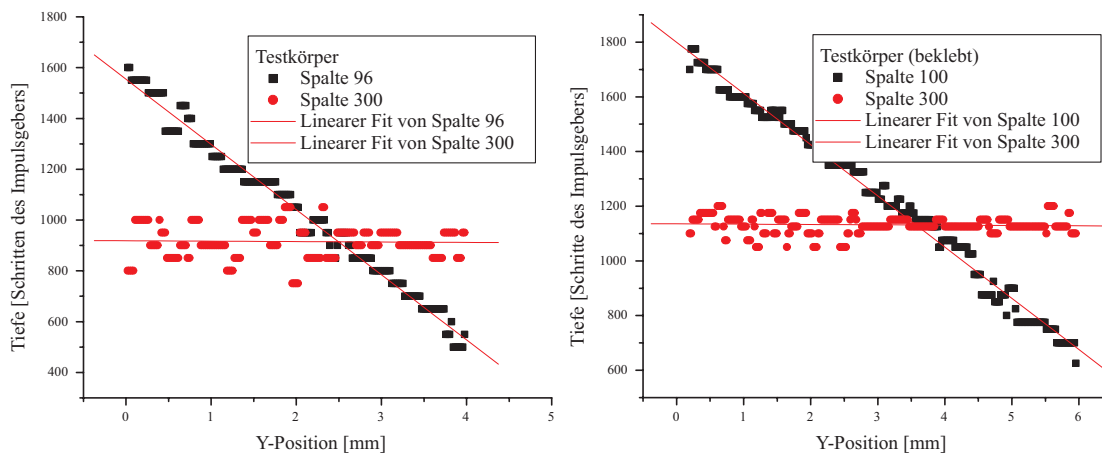
#### **9.1.4 Aufnahmen des IMS-Testkörpers**

Hier ist nun die Auswertung einiger Fokussereien zu sehen. Benutzt wurde im Folgenden die Fokusverstellung des Handmikroskops, womit sich anschließend die Anzahl der Signale des Impulsgebers in eine Tiefe in Millimetern umrechnen lässt. Abbildung 9.4 zeigt exemplarisch das Aussehen einer Tiefenkarte. In ihr ist die Tiefe in Grauwerten bzw. Helligkeit codiert.



*Abbildung 9.4: Tiefenkarte des IMS-Testkörpers. Links ist die um 45 Grad geneigte Ebene zu sehen.*

Zur weiteren Auswertung wurden nun jeweils eine Spalte aus dem Bereich der geneigten Ebene und der nahezu senkrecht zur optischen Achse stehenden Ebene herausgegriffen. In Abb. 9.5 sind sie und die zugehörigen Ausgleichsgeraden dargestellt.



**Abbildung 9.5:** Ausgewählte Spalten der rekonstruierten Tiefe des IMS-Testkörpers. (Links unbelebt und recht belebt.)

Bei der Serie des unbelebten Testkörpers mußten einige Bilder weggelassen werden, da sich bei ihnen anscheinend die Motorverstellung verklemmt hatte. Trotz der nur ca. 30 % geringeren Anzahl an Bildern erhöht sich der Fehler beim Fit von  $\pm 1,2$  auf  $\pm 4,2$  Schritte des Impulsgebers um fast den Faktor vier. Daran ist zu erkennen, wie wichtig ausreichender Kontrast und geeignete Oberflächeneigenschaften sind.

Unter der Annahme, daß der Winkel der Ebenen des IMS-Testkörpers zueinander 45 Grad beträgt, läßt sich berechnen, um wie weit die Fokusebene bei einem Signal des Impulsgebers verschoben wird. Für die Steigung der geneigten Ebene gegenüber der nicht geneigten ergibt sich ein Wert von  $-186,2 \pm 1,1$  Impulsen pro Millimeter. Bei einem Winkel von 45 Grad ist das Verhältnis von Tiefe zu Y-Position genau 1 : 1 und damit erhält man:

$$Z: 186,2 \pm 1,1 \text{ Impulse} \hat{=} 1,0 \text{ mm} \quad \rightarrow \quad 1 \text{ Impuls} \hat{=} 5,37 \pm 0,03 \mu\text{m} \quad (9.1)$$

Diese scheinbare Genauigkeit von ca.  $6 \mu\text{m}$  wird nur mit einem Fit von 200 bis 300 Meßpunkten erreicht. Die Genauigkeit einer Einzelmessung, also der Tiefeninformation an einem Bildpunkt, liegt schon aus statistischen Gründen ca. 10-20 Mal so hoch, d.h. bei ca. 0,1 mm.

## 9.2 Nicht telezentrische Aufnahmen

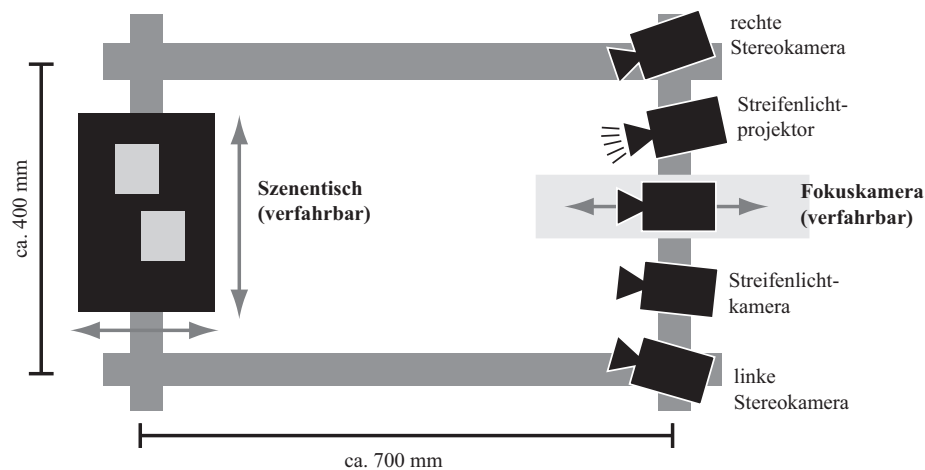
Die Aufnahmen von Fokussereien mit nicht telezentrischen Objektiven wurden zusammen mit anderen Aufnahmen in einem Aufbau durchgeführt. Die jeweiligen Szenen wurden zusätzlich zu den Fokussereien von vier Kameras zur passiven Stereoauswertung und von



einem aktive Stereosystem mit einem Streifenlicht-Projektor aufgenommen. Dabei wurden die Aufnahmen im infra-roten Licht durchgeführt, um die Störlichteinflüsse zu minimieren. Das Kamerarauschen wurde verringert, in dem statt Einzelaufnahmen zu machen ganze Sequenzen von 10-100 Bildern aufgenommen wurden, von denen je ein Original, der Mittelwert und die Varianz gespeichert wurde.

### 9.2.1 Der Aufbau

In Abbildung 9.6 ist der gesamte Meßaufbau skizziert. Als Basis dient ein aus X95-Streben aufgebautes Rechteck. Auf der linken Seite ist auf zwei gekreuzt montierten Owis-Verschiebetischen der Szenentisch angebracht. Gegenüber befinden sich die vier Kameras vom Typ Sony X55 und der Streifenlichtprojektor mit einer improvisierten Automatikblende. Diese fünf Aufnahmeelemente sind paarweise angeordnet: Ganz außen sind die Kameras für das passive Stereosystem, dann folgen der Streifenlichtprojektor und die zugehörige Kamera und in der Mitte befindet sich die auf einem Owis-Verschiebetisch angebrachte Fokuskamera. Jeweils die beiden Stereokameras und die Fokus- und die Streifenlichtkamera sind über einen Eltec PCEye-Framegrabber mit dem PC verbunden, so daß die Stereokameras synchron betrieben werden konnten. Von den drei Verschiebetischen und der Blende für den Streifenlichtprojektor sind je nach Bedarf bis zu drei Einheiten mit einer Steuerkarte im PC verbunden und lassen sich automatisch steuern.



**Abbildung 9.6:** Skizze des Aufbaus zur Aufnahme von Fokuserien. Links die zu beobachtende Szenen und auf der rechten Seite die vier Kameras und der Streifenlichtprojektor.

Für eine Szenen bzw. Szenenposition wurden jeweils vier Aufnahmen der einzelnen Kameras, eine Streifenlichtaufnahme und eine Fokuserie gemacht. Um die Position und die

restlichen Kameraparameter möglichst genau zu erhalten, wurde für alle Kameras eine Kalibrierung durchgeführt. In der folgenden Tabelle sind die daraus errechneten Daten für die Fokuskamera aufgelistet:

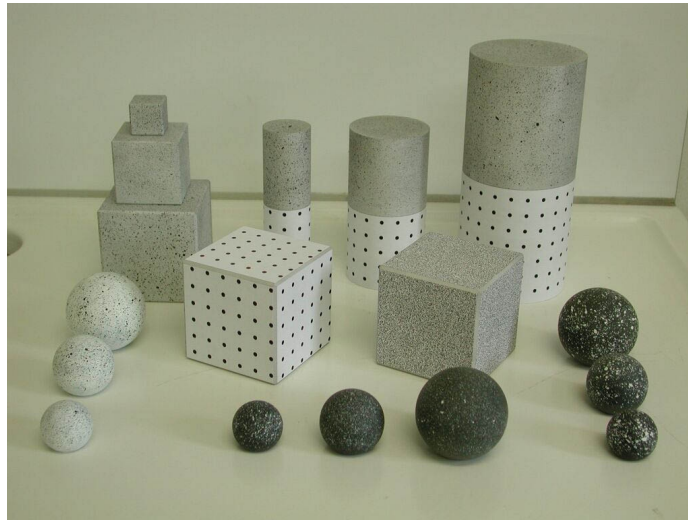
Parameter	Wert
Position $(X, Y, Z)$	3.76 mm, 5.95 mm, -727.27 mm
eff. Brennweite $f$	25.19 mm
Hauptpunkt $(x, y)$	303.06 pixel, 251.67 pixel
Richtung $(w, p, r)$	-0.00 Grad, -0.01 Grad, -3.14 Grad



**Abbildung 9.7:** Foto des Aufbaus zur Aufnahme von Fokusserien. Links die zu beobachtende Szenen und auf der rechten Seite die vier Kameras und der Streifenlichtprojektor.

## 9.2.2 Testkörper

Als Testkörper wurden unter anderen Würfel, Zylinder und Kugeln in unterschiedlichen Größen, bzw. Durchmessern verwendet. Die Würfel und Zylinder bestehen aus sandgestrahltem Aluminium und wurden für die Aufnahmen mit Mustern beklebt. Die Kugeln aus sandgestrahltem Plastik wurden mit Farbe weiß besprenkelt, bzw. komplett weiß eingefärbt und schwarz besprenkelt. In Abbildung 9.8 ist ein Übersichtsfoto der Testkörper zu sehen.

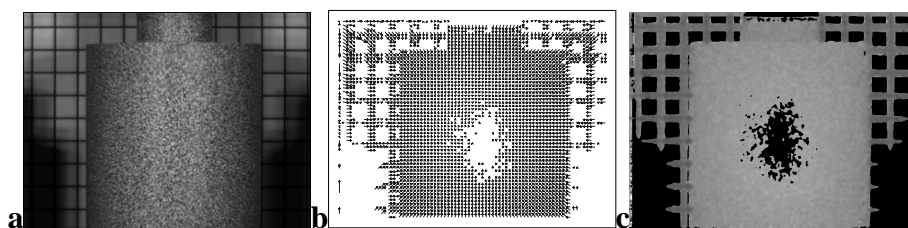


**Abbildung 9.8:** Foto der benutzen Testkörper. Die Würfel und Zylinder bestehen aus sandgestrahltem Aluminium und wurden mit verschiedenen Mustern beklebt. Die Kugeln sind aus Plastik und wurden ebenfalls sandgestrahlt, um eine nicht reflektierende Oberfläche zu erhalten.

### 9.2.3 Testszenen

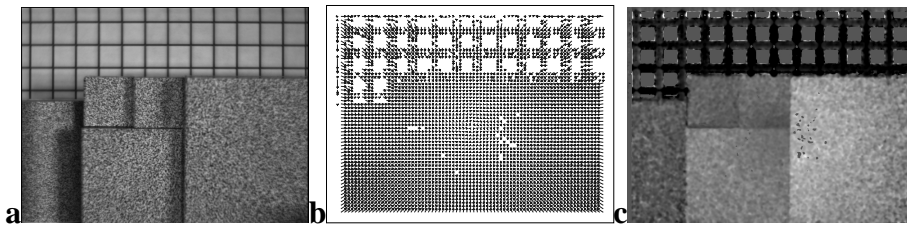
An zwei Testszenen soll demonstriert werden, daß die in Kapitel 7 vorgestellte Methode auch auf realen Fokuserien qualitativ richtige Tiefendaten liefert.

In Abbildung 9.9 und 9.10 sind zwei aus den oben beschriebenen Testkörpern zusammengesetzte Szenen zu sehen. Teilbild a zeigt jeweils das Originalbild, während in Teilbild b und c die berechnete Bewegung bzw. die Diffusionskonstante zu sehen ist.



**Abbildung 9.9:** Beispielsequenz einer realen Szene (Zylinder): **a** Originalbild **b** 2D-Bewegung und **c** Diffusionskonstante

Man kann deutlich erkennen, daß von dem Verfahren die von der Kameraverschiebung im Laufe der Fokuserie herrührende Bewegung im Bild erkannt wird, und auch die Diffusionskonstante zeigt einen zur Szenentiefe proportionalen Verlauf.



**Abbildung 9.10:** Beispielsequenz einer realen Szene (Würfel): **a** Originalbild **b** 2D-Bewegung und **c** Diffusionskonstante

# Kapitel 10

## Zusammenfassung und Ausblick

Ziel der vorliegenden Arbeit war die genaue Untersuchung von Depth-From-Focus-Verfahren. Im Rahmen des *OpenEye-Projekts* sollten deren Einsatzmöglichkeiten im Handmikroskop der A-Tec analysiert werden.

Dazu wurde in Kapitel 2 der Prozess der Bildentstehung eingehend untersucht, wobei besonders auf die Entstehung der Unschärfe im Bild Wert gelegt wurde. Die 2D- und 3D-Punktantworten und ihre Transferfunktionen wurden hergeleitet. In Kapitel 3 wurden die DFF-Verfahren mit anderen 3D-Meßverfahren verglichen und entsprechend ihrem zu Grunde liegenden physikalischen Meßprinzip eingeordnet. Dabei konnte gezeigt werden, daß die DFF-Verfahren und die Stereo-Verfahren auf dem gleichen Meßprinzip, der Triangulation, basieren und daher bei gleichen Systemdimensionen ähnliche Probleme aufweisen. Es wurden aber auch Unterschiede, wie z.B. die Ausdehnung der Punktantwort des gesamten Systems und die daraus resultierenden Auswirkungen von Verdeckungen herausgearbeitet. Kapitel 4 gibt einen Überblick über die in der Literatur beschriebenen DFF-Verfahren, während in Kapitel 5.2 eine Methode zur Parameterschätzung, insbesondere der Diffusionskonstante, in Bildsequenzen grundlegend erläutert wurde. Eine ausführliche Beschreibung und Charakterisierung des Verlaufs des Unschärfeparameters erfolgte in Kapitel 6, wobei besonders der Einfluß spezieller Szenengeometrien, wie z.B. Tiefenkanten, herausgearbeitet wurde. Weiterhin konnte die lokale Varianz als ein Schärfemaß eingeführt und deren Verhalten bezüglich verschiedener Wellenzahlen analysiert werden. Es wurden schließlich verschiedene Methoden zur Bestimmung des Maximums vorgestellt und untersucht, inwieweit diese in einfacher und speichersparender Weise implementiert werden können, so daß sie sich z.B. für den Einsatz in FPGA-Hardware eignen.

Eine neuartige Methode zur Tiefenschätzung aus Fokussereien wurde in Kapitel 7 vorgestellt und ihre Eignung anhand synthetischer Daten erprobt. Da diese Methode auf ei-

ner Parameterschätzung in der Fokusserie basiert, konnte neben der Diffusionskonstanten, aus der anschließend die Tiefeninformation berechnet wurde, auch die Bewegung in der Sequenz geschätzt werden. Skalierungseffekte, wie sie bei der Verwendung einer nicht-telezentrischen Optik auftreten, werden mitgeschätzt und stören die Tiefenbestimmung nicht.

Die effiziente und genaue Erzeugung von synthetischen Fokusserien wurde in Kapitel 8 beschrieben. Mit den so berechneten synthetischen Fokusserien, konnten die vorgestellten Verfahren getestet werden. Abschließend wurden in Kapitel 9 zwei Aufbauten zu Aufnahme realer Fokusserien beschrieben und die Verfahren an einigen realen Sequenzen erfolgreich erprobt.

In Zukunft sind vor allem zwei Erweiterungen der neuartigen Methode zur Tiefenbestimmung aus der Diffusionskonstanten erfolgversprechend:

- Bisher wurde die Diffusionskonstante in einer kleinen Umgebung örtlich und zeitlich konstant modelliert. Örtlich konstant ist sie nicht, denn sie variiert mit der Szenentiefe und das wird ja auch von der Methode zu Tiefenbestimmung genutzt. Zeitlich bzw. im Verlauf der Fokusserie ist sie auch nicht konstant, wie in Kapitel 7 gezeigt. Eine höhere Genauigkeit könnte man daher erwarten, wenn man eine lineare Abhängigkeit der Diffusionskonstanten in örtlicher und zeitlicher Richtung direkt in die der Parameterschätzung zu Grunde liegende Differentialgleichung modelliert.
- An Kanten ist die örtliche Konstanz stark verletzt, was zu hohen Fehlern in der Diffusionskonstanten und damit in der Tiefenschätzung führt. Mit Hilfe der robusten Statistik, wie sie von Garbe [2001] zur Detektion von Reflexen auf der Wasseroberfläche eingesetzt wurde, könnten die jeweils nicht ins Modell passenden Bildpunkt aus der lokalen Umgebung weniger stark gewichtet werden. Damit kann das Ergebnis der Tiefenschätzung an Kanten verbessert werden.

# Anhang





# Anhang A

## Fehlerfortpflanzung

Um den Einfluß von Bildrauschen zu untersuchen geht man davon aus, daß sich das verrauschte Bild  $G_N$  aus einem Signal- und einem Rauschanteil zusammensetzt:

$$G_N = G + N \quad (\text{A.1})$$

Dabei soll das Rauschen einer mittelwertfreien Normalverteilung mit der ortsunabhängigen Standardabweichung  $\sigma_N$  entsprechen, so daß für  $N$  gilt:

$$\langle N \rangle = 0 \quad (\text{A.2})$$

$$\langle (N - \langle N \rangle)^2 \rangle = \sigma_N^2 \quad (\text{A.3})$$

Damit ergibt sich für den Erwartungswert und die Varianz des verrauschten Bildes  $G_N$ :

$$\langle G_N \rangle = \langle G \rangle + \langle N \rangle = G \quad (\text{A.4})$$

$$\sigma_{G_N}^2 = \langle (G_N - \langle G_N \rangle)^2 \rangle = \langle (G + N - G)^2 \rangle = \langle N^2 \rangle = \sigma_N^2 \quad (\text{A.5})$$

### A.1 Erwartungswert und Varianz linearer verschiebungs-invarianter Filter

Zur einfacheren Schreibweise sollen die Indizes, über die die Faltung ausgeführt wird, durch einen Index  $i$  ersetzt werden, womit man für die Anwendung des Filters  $A$  auf das Bild  $G$  schreiben kann:

$$G' = \sum_i a_i G_i \quad \text{mit} \quad \sum_i a_i = 1 \quad (\text{A.6})$$

Für das verrauschte Bild  $G_N$  ergibt sich daraus:

$$G'_N = \sum_i a_i G_{N,i} = \sum_i a_i G_i + \sum_i a_i N_i \quad (\text{A.7})$$

Und für den Erwartungswert und die Varianz erhält man:

$$\begin{aligned} \langle G'_N \rangle &= \sum_i a_i G_i + \sum_i a_i \langle N_i \rangle \\ &= \sum_i a_i G_i \end{aligned} \quad (\text{A.8})$$

$$\sigma_{G'_N}^2 = \langle (G'_N - \langle G'_N \rangle)^2 \rangle \quad (\text{A.9})$$

$$\begin{aligned} &= \left\langle \left( \left( \sum_i a_i G_i + \sum_i a_i N_i \right) - \left( \sum_i a_i G_i \right) \right)^2 \right\rangle \\ &= \left\langle \left( \sum_i a_i N_i \right)^2 \right\rangle \\ &= \left\langle \sum_i \sum_j a_i a_j N_i N_j \right\rangle \\ &= \sum_i \sum_j a_i a_j \langle N_i N_j \rangle \\ &= \sum_i a_i^2 \sigma_i^2 \end{aligned} \quad (\text{A.10})$$

$$= \sigma^2 \sum_i a_i^2 \quad (\text{A.11})$$

Dabei wurde im vorletzten Schritt vorausgesetzt, daß das Rauschen räumlich unkorreliert ist, so daß  $\langle N_i N_j \rangle = \sigma_{N,i}^2 \delta_{ij}$  ist. Im letzten Schritt wurde angenommen, daß die Varianz des Rauschens  $\sigma_{N,i}^2 = \sigma_N$  im ganzen Bild konstant ist.

## A.2 Erwartungswert und Varianz des quadrierten Bildes

Sei  $G'_N$  das Quadrat des verrauschten Bildes  $G_N = G + N$ , dann gilt:

$$G'_N = (G_N)^2 = (G + N)^2 = G^2 + N^2 + 2GN \quad (\text{A.12})$$

Für den Erwartungswert erhält man damit:

$$\begin{aligned} \langle G'_N \rangle &= \langle G^2 \rangle + \langle N^2 \rangle + \langle 2GN \rangle \\ &= G^2 + \sigma^2 + 2G \langle N \rangle \\ &= G^2 + \sigma^2 \end{aligned} \quad (\text{A.13})$$

Für die Varianz ergibt sich:

$$\begin{aligned}
 \sigma_{G'_N}^2 &= \langle (G'_N - \langle G'_N \rangle)^2 \rangle \\
 &= \langle ((G + N)^2 - (G^2 + \sigma^2))^2 \rangle \\
 &= \langle (2GN + N^2 - \sigma^2)^2 \rangle \\
 &= \langle 4G^2N^2 + N^4 + \sigma^4 + 4GN^3 - 4GN\sigma^2 - 2N^2\sigma^2 \rangle \\
 &= 4G^2 \underbrace{\langle N^2 \rangle}_{=\sigma^2} + \underbrace{\langle N^4 \rangle}_{=3\sigma^4} + \sigma^4 + 4G \underbrace{\langle N^3 \rangle}_{=0} - 4G \underbrace{\langle N \rangle}_{=0} \sigma^2 - 2 \underbrace{\langle N^2 \rangle}_{=\sigma^2} \sigma^2 \\
 &= 4G^2\sigma^2 + 3\sigma^4 + \sigma^4 - 2\sigma^4 \\
 &= 4G^2\sigma^2 + 2\sigma^4
 \end{aligned} \tag{A.14}$$



# Anhang B

## Beschreibung der Software

### B.1 Zusätzliche Heurisko-Operatoren

Es wurden verschiedene Windows-DLLs geschrieben, um die benötigten Operationen unter Heurisko4 verfügbar zu machen. Neben den Operationen zur Berechnung von Tiefenkarten und der Synthese von Fokusserien werden noch DLLs zur Steuerung des Verschiebetischs und der Motorverstellung des Fokus des Handmikroskops benötigt.

#### B.1.1 focus.dll

Alle Operatoren, die im folgenden kurz beschrieben werden, arbeiten ausschließlich mit FLOAT-Bildern, d.h. die Bildpunkte sind Fließkommawerte einfacher Genauigkeit. Zur weiteren Beschleunigung der Algorithmen ließen sich diese auch für 8bit-wertige Bildpunkte implementieren.

Die Operator-Namen sind dickgedruckt und die Anzahl der eckigen Klammern hinter den Ein- bzw. Ausgabeobjekten gibt deren Dimension an.

$\text{tiefenkarte}[\ ][\ ] = \mathbf{ArgMax}(\text{fokusserie}[\ ][\ ][\ ])$

*ArgMax* berechnet punktweise das globale Maximum entlang der Z-Koordinate in der Eingangssequenz *fokusserie* und speichert das Ergebnis im Bild *tiefenkarte*.

$\text{tiefenkarte}[\ ][\ ] = \mathbf{ArgMaxStatistik}(\text{fokusserie}[\ ][\ ][\ ])$

*ArgMaxStatistik* berechnet punktweise und subbildgenau das globale Maximum entlang der Z-Koordinate in der Eingangssequenz *fokusserie*, d.h. daß die Genauigkeit nicht mehr auf den Abstand der Bilder der Sequenz beschränkt ist (genauerer siehe im Abschnitt *Algo-*

*rithmik*). Dazu muss im Bild *tiefenkarte* schon das Ergebnis von *ArgMax* vorhanden sein, anschließend wird das Bild *tiefenkarte* mit den verbesserten Ergebnissen überschrieben.

`tiefenkarte[][] = ArgMaxPoly(fokusserie[][][])`

*ArgMaxPoly* berechnet punktweise und subbildgenau das globale Maximum entlang der Z-Koordinate in der Eingangssequenz *fokusserie*, d.h. daß die Genauigkeit nicht mehr auf den Abstand der Bilder der Sequenz beschränkt ist (genaueres siehe im Abschnitt *Algorithmik*). Dazu wird ein Polynom durch jeweils drei aufeinanderfolgende Werte gelegt und dessen Maximum im Bild *tiefenkarte* gespeichert.

`tiefenkarte[][],fehler[][] = ArgMaxPoly2(fokusserie[][][])`

*ArgMaxPoly2* berechnet punktweise und subbildgenau das globale Maximum entlang der Z-Koordinate in der Eingangssequenz *fokusserie*, d.h. daß die Genauigkeit nicht mehr auf den Abstand der Bilder der Sequenz beschränkt ist (genaueres siehe im Abschnitt *Algorithmik*). Dazu wird ein Polynom durch jeweils ein Drittel der Werte gelegt und dessen Maximum im Bild *tiefenkarte* gespeichert. Das Bild *fehler* enthält den Fehler dieses linearen Fits.

`tiefenscharfesBild[][] = Merge(tiefenkarte[][],fokusserie[][][])`

*Merge* berechnet aus der Sequenz *fokusserie* und dem Bild *tiefenkarte*, welches die vorher berechnete Tiefeninformation enthalten muß, ein tiefenscharfes Bild und speichert dies im Bild *tiefenscharfesBild*.

`fokusserie[][][] = SynFocusSeq Pillbox(tiefenkarte[][],bild[][],ebenenpositionen[],aperatur)`

*SynFocusSeq Pillbox* berechnet eine Bildsequenz *fokusserie*, wozu eine vorgegebene Tiefenkarte *tiefenkarte*, eine Oberflächenhelligkeit *bild*, die Fokusposition der einzelnen zu berechnenden Bilder *ebenenpositionen* und die numerische Aperatur der zu simulierenden Optik *aperatur* benötigt wird.

## B.2 Workspaces

Alle Workspaces sind für Heurisko4 ausgelegt und enthalten weitere Dokumentation im Code selbst, so daß hier auf eine ausführliche Beschreibung verzichtet wurde.

### B.2.1 handmikroskop aufnahme.ws

Dieser Heurisko4 Workspace enthält Operatoren zur Aufnahme von Bildsequenzen mit dem Handmikroskop. Dazu wird eine Datatranslation DT300 Karte zum Anschluß der Fokussteuerung bzw. ein Verschiebetisch inklusive der entsprechenden Ansteuerungskarte ge-

braucht. Weiterhin wird ein PC-Eye2 Framegrabber der Firma Eltec verwendet.

### **B.2.2 newFokus.ws**

Mit diesem Workspace wurden alle Auswertungen gemacht. Er beherrscht alle hier erwähnten Algorithmen, allerdings werden alle Daten insbesondere auch die Bildsequenzen komplett im Hauptspeicher gehalten, wodurch dieser Workspace nur für kleinere Sequenzen oder für Rechnern mit viel Speicher geeignet ist.

### **B.2.3 newFokusInc.ws**

Diese Workspace ist speziell für Rechner mit geringerer Hardwareausstattung gedacht. Er berechnet ein tiefenscharfes Bild, das Gütemaß und die Tiefenkarte inkrementell, d.h. es wird jeweils ein Bild geladen und sofort verarbeitet.





# Literaturverzeichnis

- J. L. Barron, D. J. Fleet, and S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, 1994.
- M. Born and E. Wolf. *Prinziples of Optics*. Pergamon Press, 6 edition, 1980.
- S. Chaudhuri and A. Rajagopalan. *Depth From Focus: A Real Aperature Imaging Approach*. Springer-Verlag, 1999.
- T. Darell and K. Wohn. Depth from focusing using a pyramid architecture. *Pattern Recognition Letters*, 11(12):787 – 796, December 1990.
- T. Dresler, G. Häusler, and V. H. Three-dimensional sensing of rough surfaces by coherence radar. *Applied Optics*, 31(7):919 – 925, March 1992.
- J. Ens and P. Lawrence. An Investigation of Methods for Determining Depth from Focus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(2):97 – 108, February 1993.
- A. Erhardt, G. Zinser, D. Komitowski, and J. Bille. Reconstructing 3-D light-microscopic images by digital image processing. *Applied Optics*, 24(2):194–200, January 1985.
- O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint*. The MIT Press, Cambridge, MA, 1993.
- D. J. Fleet. *Measurement of Image Velocity*. Kluwer Academic Publishers, Dordrecht, The Netherlands, 1992.
- C. S. Garbe. *Heat Flow*. PhD thesis, University of Heidelberg, Heidelberg, Germany, 2001. in preparation.
- P. Geißler. *Depth-from-Focus zur Messung der Größenverteilung durch Wellenbrechung erzeugter Blasenpopulationen*. PhD thesis, IWR, Universität Heidelberg, 1998.

- P. Geißler and T. Dierig. Depth-From-Focus. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications*, volume 2, pages 592–623. Academic Press, 1999.
- P. Grossmann. Depth from focus. *Pattern Recognition Letters*, 5(1):63 – 69, January 1987.
- G. Häusler and E. Körner. Simple focusing criterion. *Applied Optics*, 23(15):2468–2469, August 1984.
- G. Häusler and D. Ritter. Parallel three-dimensional sensing by color-coded triangulation. *Applied Optics*, 32(35):7164 – 7169, December 1993.
- H. Haußecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):661–673, June 2001.
- H. Haußecker, C. S. Garbe, H. Spies, and B. Jähne. A total least squares for low-level analysis of dynamic scenes and processes. In *DAGM*, pages 240–249, Bonn, Germany, 1999. Springer.
- H. Haußecker and H. Spies. Motion. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications*, volume 2, chapter 13, pages 309–396. Academic Press, San Diego, 1999.
- E. Hecht. *Optik*. Addison-Wesley, 2 edition, 1989.
- B. Jähne. *Digitale Bildverarbeitung*. Springer, 4 edition, 1997.
- R. A. Jarvis. A perspective in range-finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5:122–139, March 1983.
- E. Krotkov. Focusing. *International Journal of Computer Vision*, 1:223 – 237, 1987.
- B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *DARPA Image Understanding Workshop*, pages 121–130, 1981.
- S. K. Nayar, M. Watanabe, and M. Noguchi. Real-time focus range sensor. *PAMI*, 18(12): 1186–1199, December 1996.
- A. P. Pentland. A new sense of depth of field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(4):523 – 531, July 1987.
- A. Rajagopalan and S. Chaudhuri. A Variational Approach to Recovering Depth From Defocused Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(10):1158 – 1164, October 1993.

- A. N. Rajagopalan and S. Chaudhuri. Simultaneous depth recovery and image restoration from defocused images. In *CVPR'99*, Fort Collins, Colorado, June 1999.
- Y. Y. Schechner and N. Kiryati. Depth from defocus vs. stereo: How different really are they? *International Journal of Computer Vision*, 39(2):141–162, 2000.
- T. Scheuermann, G. Pfundt, P. Eyerer, and B. Jähne. Oberflächenkonturvermessung mikroskopischer Objekte durch Projektion statistischer Rauschmuster. In *DAGM*, pages 319 – 326, 1995.
- T. Scheuermann, G. Wiora, and M. Graf. Topographical Maps of Microstructures generated by Depth from Focus Techniques. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications*, volume 3, pages 379 – 410. Academic Press, 1999.
- T. J. Scholz. *Ein Depth from Focus-Verfahren zur On-Line Bestimmung der Zellkonzentration bei Fermentationsprozessen*. PhD thesis, IWR, Universität Heidelberg, 1995.
- R. Schwarte, H. Heinol, B. Buxbaum, T. Ringbeck, Z. Xu, and K. Hartmann. Principles of three-dimensional imaging techniques. In B. Jähne, H. Haußecker, and P. Geißler, editors, *Handbook of Computer Vision and Applications*, volume 1, pages 463–484. Academic Press, 1999.
- R. Schwarte, H. Heinol, B. Buxbaum, Z. Xu, T. Ringbeck, Z. Zhang, W. Tai, K. Hartmann, W. Kleuver, and X. Luan. Neuartige 3D-Visionssysteme auf der Basis Layout-optimierter PMD-Strukturen. *Technisches Messen*, pages 264 – 271, July - August 1998.
- H. Spies. *Analysing Dynamic Processes in Range Data Sequences*. PhD thesis, University of Heidelberg, Heidelberg, Germany, July 2001.
- H. Spies and B. Jähne. A general framework for image sequence analysis. In *Fachtagung Informationstechnik*, pages 125–132, Magdeburg, Germany, March 2001. Otto-von-Guericke-Universität Magdeburg.
- M. Subbarao. Optical transfer function of a diffraction-limited system for polychromatic illumination. *Applied Optics*, 29(4):554–558, February 1988.
- M. Subbarao and J.-K. Tyan. A perspective in range-finding techniques for computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):864–870, August 1998.
- S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*. Society for Industrial and Applied Mathematics, Philadelphia, 1991.



# Danksagung

Beim Rückblick auf die Planung, die Durchführung und den Abschluß dieser Arbeit wird deutlich welche mannigfaltigen Hilfen ich in dieser Zeit erhalten habe.

An erster Stelle muß ich an Herrn Prof. Dr. Jähne denken und möchte ihm insbesondere dafür danken, daß er mich als Doktorand in seine Arbeitsgruppen integrierte, mich in Bereichen arbeiten ließ, die meiner Interessenlage besonders entsprachen, und daß er die Betreuung dieser Arbeit übernahm.

Herrn Prof. Dr. Bille möchte ich dafür danken, daß er sich als Zweitkorrektor zur Verfügung gestellt hat.

Die fruchtbaren Diskussionen und die zahlreichen Anregungen aus dem Kreis meiner Kollegen in der Arbeitsgruppe haben mir bei der Erstellung dieser Dissertation geholfen. Vor allem danke ich Hagen Spieß und Christoph Garbe für ihre Hilfestellung bei der Durchführung und Auswertung der Experimente.

Nicht zuletzt möchte ich mich ganz herzlich bei meiner Familie und Eva-Maria Lehmann bedanken, die mir besonders in den letzten Tagen bei der Endfassung dieser Arbeit mit Rat und Tat entscheidend geholfen haben.