

Verena Lauren

Semilineare Approximation in der Bildrekonstruktion

Eingereicht zur Erlangung des Grades eines Doktors der Naturwissenschaften an der
Fakultät für Mathematik und Informatik der Universität Passau im November 2004



Zusammenfassung

In der Bildverarbeitung spielen Segmentierungsalgorithmen, bei denen die Art der zu segmentierenden Menge schon im Voraus festgelegt werden kann und nur noch ihre Größe und Lage angepaßt werden muß, eine eher untergeordnete Rolle. Gründe hierfür sind vor allem komplizierte Zielfunktionen und daraus resultierende lange Rechenzeiten, die zudem meist kein optimales Ergebnis liefern. Dabei kann eine mengenbasierte Segmentierung durchaus sinnvoll eingesetzt werden, wenn gewisse Rahmenbedingungen eingehalten werden. In dieser Arbeit wird eine Theorie zur allgemeinen mengenbasierten Segmentierung vorgestellt und untersucht, unter welchen Bedingungen optimale Segmentierungsergebnisse erreicht werden können. Die anschließenden Anwendungen bestätigen die Nützlichkeit dieser Theorie.

Vorwort

Die vorliegende Arbeit entstand im Laufe meiner Tätigkeit als wissenschaftliche Mitarbeiterin bei FORWISS unter der Leitung von Prof. Dr. K. Donner an der Universität Passau.

Mein Dank gilt Herrn Prof. Dr. K. Donner, der diese Arbeit angeregt hat. Durch fortlaufende wissenschaftliche Diskussionen und Ratschläge hat er mir geholfen, diese Arbeit zu schreiben.

Ich danke Herrn Prof. Dr. S. Graf für die Übernahme des Zweitgutachtens.

Bedanken möchte ich mich auch bei meinen Kollegen für die zahlreichen fachlichen Diskussionen und nicht zuletzt auch für ihre Geduld, da ich mich (bedingt durch drei Babypausen und Teilzeitarbeit mit wechselnden Stundenzahlen) als recht unzuverlässige Kollegin im Hinblick auf die Entlastung anderer erwies.

Mein besonderer Dank geht natürlich an meine Familie. Hier alles Positive aufzuzählen, wäre wohl zuviel des Guten; daher sei nur kurz angemerkt, daß ihre Unterstützung die vielen durchwachten Nächte und vollen Windeln bei weitem wett gemacht hat. Trotz der doch recht langen Zeit, die diese Dissertation bis zu ihrer Fertigstellung benötigt hat (immerhin zehn Jahre), würde ich mich immer wieder für Familie *und* Arbeit entscheiden. Natürlich hätte ich diese Dissertation lieber in einer kürzeren Zeitspanne abgeschlossen, aber die aktuelle Betreuungssituation für Kinder (und nicht zuletzt das gesellschaftliche Umfeld) sieht die Mutter immer noch als hauptamtlich für die Kinder zuständig. Eine adäquate Wochenarbeitszeit zu reservieren, um meine Promotion schneller voranzutreiben, war daher für mich nicht zu schaffen. Aber nun ist sie fertig und ich darf mich (zumindest in dieser Hinsicht) zurücklehnen und gespannt sein, ob meine Tochter in Zukunft bessere Möglichkeiten vorfindet, um Familie und Arbeit unter einen Hut zu bringen.

Passau, im März 2006

Verena Lauren

Inhaltsverzeichnis

Symbolverzeichnis	9
1 Einleitung	11
2 Bildsegmentierung durch Funktionsquantisierungen	13
2.1 Zwei-Niveau-Passung	13
2.2 Funktionsquantisierungen	22
2.3 Methoden der Variationsrechnung	23
2.4 Adaption parametrischer Grauwertmodelle	25
3 Semilineare Approximation auf euklidischen Vektorräumen	31
3.1 Semilineare Funktionen	31
3.2 Semilineare Approximation	37
3.3 Existenz lokaler Optima	40
4 Sensoren und Sensoraufnahmen	55
4.1 Das Sensorarray	55
4.2 Das Linsensystem	62
4.3 Einbeziehung von Sensorarrays in semilineare Funktionen	63
5 Ein mögliches Mengenmodell	67
5.1 Die lokale Hausdorff-Topologie	67
6 Subpixelgenaue Objektpassung auf Grauwertbildern	73
6.1 Ein semilinearer Ansatz zur subpixelgenauen Objektpassung	75
6.2 Die modifizierte Grauwertmodelladaption	80
6.3 Test beider Verfahren	85
7 Objektrekonstruktion in der Computer-Tomographie	93
7.1 Problemstellung	93
7.2 Absorptionsverhalten von Röntgenstrahlung	95
7.3 Modellierung der Objektform	98
7.4 Ergebnisse und Ausblick	102

8 Zusammenfassung und Ausblick	107
8.1 Mehrkanalige Bilder	108
8.2 Erweiterte Mengensysteme und überlappende Mengen	109
A Algebraische Grundlagen	111
B Ergebnisse der Paßverfahren aus Kapitel 6	113
B.1 Modifizierte Grauwertmodelladaption	113
B.2 Kreisscheibenpassung und Vergleich beider Verfahren	118
Literaturverzeichnis	121

Symbolverzeichnis

\mathbb{N}	Menge der natürlichen Zahlen, $\mathbb{N} = \{1, 2, 3, \dots\}$
\mathbb{R}	Menge der reellen Zahlen
$\mathfrak{P}(X)$	Potenzmenge von X
$\mathfrak{A}(X)$	Menge der abgeschlossenen Teilmengen von X
$\mathfrak{K}(X)$	Menge der kompakten Teilmengen von X
\bar{X}	Abschluß von X
$\mathcal{B}(X)$	Menge der Borelschen Teilmengen von X
$\lambda^d(A)$	d -dimensionales Lebesgue-Maß auf A
$\mathcal{L}^2(A, B)$	Vektorraum der quadratintegrierbaren Funktionen von A nach B
$L^2(A, B)$	Quotientenvektorraum der quadratintegrierbaren Funktionen von A nach B
$L^1_\mu(A, B)$	Quotientenraum der μ -integrierbaren Funktionen von A nach B
$\mathfrak{S}(f)$	Bildraum der Funktion $f : X \rightarrow Y$, $\mathfrak{S}(f) \subseteq Y$
$\mathbb{1}_A$	Indikatorfunktion auf A
∇f	Gradient der Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$
$\text{dist}(A, B)$	Distanzfunktion, Abstand von A und B
$\dim(X)$	Dimension des Vektorraumes X
$\text{proj}_U y$	Orthogonalprojektion des Vektors y auf den Unterraum U
$\langle x y \rangle$	Skalarprodukt von x und y
$\ x\ $	Norm von x
$\text{linH}\{x_1, \dots, x_n\}$	lineare Hülle der Vektoren x_1, \dots, x_n
$\text{Hom}(E, F)$	Menge der linearen Abbildungen von E nach F
$\text{sup}(f)$	Support (Träger) der Funktion f

1 Einleitung

Ausgangspunkt dieser Arbeit war ein Segmentierungsproblem aus dem Bereich der zerstörungsfreien Prüfung. Gegeben war ein Aluminiumstück, das aus zwei miteinander verschweißten Teilen bestand. Die Aufgabe bestand nun darin zu bestimmen, wie belastbar dieses Stück ist. Dazu muß man wissen, daß der wesentliche Schwachpunkt in der Stabilität in dem Bereich liegt, an dem die beiden einzelnen Teile miteinander verschweißt wurden. Eine Prüfung der Stabilität des Aluminiumstückes besteht daher im wesentlichen in der Untersuchung der Güte des Schweißnahtbereiches. Von besonderem Interesse bei dieser Untersuchung ist dabei die sogenannte Einbrandtiefe, die beschreibt wie weit die Schweißnaht in das Aluminium hineinreicht. Nur eine ausreichende Einbrandtiefe gewährleistet eine dauerhafte Belastbarkeit.

Im industriellen Umfeld ist eine Gütekontrolle der Einbrandtiefe bisher nur zerstörend möglich. Dazu wird das verschweißte Aluminiumstück auseinandergesägt und die zersägten Flächen anschließend einer speziellen chemischen Behandlung unterzogen. Nach dieser Behandlung ist der Schweißnahtbereich als scharf abgegrenzter Bereich sichtbar und die Einbrandtiefe kann nun begutachtet und vermessen werden. Diese zerstörende Prüfung arbeitet sehr zuverlässig. Ihr größter Nachteil liegt aber auf der Hand, denn natürlich ist ein zersägtes Aluminiumstück für die Weiterverarbeitung unbrauchbar geworden. Eine zerstörungsfreie Prüfungsmöglichkeit wäre daher ein wünschenswertes Ziel. Diese angedachte Prüfung sollte als Ergebnis ein Bild des Aluminiumquerschnittes liefern in dem der Schweißnahtbereich ebenso deutlich abgegrenzt vom Rest des Aluminiumstückes erscheint wie nach einer zerstörenden Prüfung, so daß am Bild eine Vermessung der Einbrandtiefe möglich ist.

Aus Gründen der Wirtschaftlichkeit kommt als einzige Prüfungsmethode, die das Innere eines Werkteils sichtbar machen kann, die Röntgen-Tomographie in Frage. Neben medizinischen Tomographen existieren auch für den Einsatz in der Materialprüfung zahlreiche Geräte in diversen Größen und mit verschiedenen Strahlenergien. Zur Rekonstruktion des Bildmaterials wird dabei — wie auch im medizinischen Bereich — eine gefilterte Rückprojektion verwendet. Für die hier erwähnten Aluminiumstücke lieferte diese Rekonstruktionsmethode jedoch nicht die gewünschten Ergebnisse, denn in den rekonstruierten Bildern war der Schweißnahtbereich vom restlichen Aluminiumstück nicht unterscheidbar. Es schien daher zunächst so, als ob eine zerstörungsfreie Prüfung von verschweißten Aluminiumteilen nicht möglich wäre.

An dieser Stelle kam die Idee ins Spiel, den Schweißnahtbereich nicht in den bereits rekonstruierten Bildern zu suchen, sondern in den noch nicht rekonstruierten Rohdaten

des Tomographen nach Hinweisen auf die Einbrandtiefe zu schauen. Im Laufe der Problemanalyse wurde dann klar, daß dieses Problem einen — nicht offensichtlichen aber dennoch deutlichen — Zusammenhang zur Zwei-Niveau-Passung aufweist, die vom Forschungsgruppenleiter Prof. Donner entwickelt wurde und bei FORWISS bereits mehrfach eingesetzt wurde. Zu dieser Zeit wurde ein weiteres Segmentierungsproblem aktuell — eine subpixelgenaue Kreisscheibenpassung. Obwohl dieses Problem mit dem oben beschriebenen Problem der Schweißnahterkennung keinerlei Berührungspunkte zu haben schien, konnte auch hier ein Zusammenhang zur Zwei-Niveau-Passung nachgewiesen werden. Die Entwicklung einer Theorie, die diese Zusammenhänge herausarbeitet, wurde dann zum Thema dieser Dissertation.

Parallel zu dieser Arbeit wurde — ebenfalls bei FORWISS — von Tobias Hanning in seiner Dissertation eine Theorie entwickelt, die ebenfalls auf der Zwei-Niveau-Passung aufbaut ([Han01]). Aber obwohl seine *vektorielle Mehrniveaupassung* und die hier vorgestellte semilineare Approximation dieselbe Basis besitzen, gehen sie doch von Anfang an getrennte Wege. In [Han01] liegt der Schwerpunkt auf der Segmentierung von Bildern in sinnvolle Bereiche, ohne daß a-priori genauere Informationen über das gegebene Bild vorhanden sind. Dagegen beschäftigt sich diese Arbeit damit, wie ein vorher bekanntes Objekt — zum Beispiel ein Rechteck oder auch ein Kreis — in einem Bild wiedergefunden werden kann. Diese Vorgehensweise kann auch als *Objektsegmentierung* bezeichnet werden, im Gegensatz zur allgemeineren *Bildsegmentierung* aus [Han01].

Die vorliegende Arbeit gliedert sich wie folgt: Kapitel 2 befaßt sich mit der Zwei-Niveau-Passung und stellt die Stärken und Schwächen dieses algorithmischen Prinzips heraus. Außerdem werden mögliche Verallgemeinerungen angesprochen sowie bestehende vergleichbare Algorithmen aus der Literatur vorgestellt, die im Hinblick auf ihre Verwendbarkeit in der Objektsegmentierung untersucht werden. In Kapitel 3 wird dann ein allgemeines Prinzip zur Bildsegmentierung vorgestellt, das auf im Voraus bereitgestellten Mengensystemen basiert. Die Einbindung dieses Segmentationsprinzips in eine vorhandene (Kamera-)Sensorik wird in Kapitel 4 besprochen. Ein Vorschlag für ein im Voraus bereitgestelltes Mengensystem wird in Kapitel 5 unterbreitet. Auf einem derartigen Mengensystem basiert eines der in Kapitel 6 vorgestellten Verfahren zur subpixelgenauen Kreisscheibenpassung. Kapitel 7 schließlich behandelt die oben erwähnte Einbrandtiefenbestimmung und zeigt, welche Kompromisse aufgrund der Komplexität der Aufgabenstellung eingegangen werden mußten, um überhaupt brauchbare Ergebnisse erzielen zu können. Abschließend werden in Kapitel 8 die Ergebnisse dieser Arbeit zusammengefaßt sowie ein Ausblick auf mögliche Weiterentwicklungen gegeben.

2 Bildsegmentierung durch Funktionsquantisierungen

Wie bereits in der Einleitung beschrieben wurde, liegt der Schwerpunkt dieser Arbeit in der Objektsegmentierung. Daher war ein wesentlicher Ausgangspunkt dieser Arbeit die in [Don91] beschriebene *Zwei-Niveau-Passung*. Dieses Segmentierungsverfahren besitzt im Gegensatz zu den meisten anderen Segmentierungsalgorithmen den Vorteil, daß die Art der zu segmentierenden Menge durch ein vorher gewähltes Mengenmodell bereits im Voraus festgelegt werden kann. Die Zwei-Niveau-Passung ordnet sich in den Kontext der *Funktionsquantisierungen* ein, wie sie zum Beispiel in [Käm00] für eindimensionale Funktionen beschrieben werden. Allerdings wird dort das zweidimensionale Mengenmodell aufgrund seiner Komplexität nur sehr eingeschränkt verwendet. Auch in die Methoden der Variationsrechnung läßt sich ein derartiges Mengenmodell kaum einbinden. Gerade dieses Mengenmodell kann aber in der Praxis eine deutliche Hilfestellung darstellen, da sich damit Vorwissen über die Art des zu segmentierenden Bereiches formulieren läßt. Ein weiteres Verfahren, das auf der Verwendung eines Mengenmodells basiert, ist die in [Dre98] vorgestellte Grauwertmodelladaptation, die eine Subpixelpassung auf der Basis von Mengenmodellen beschreibt. Im folgenden wird ein Überblick über diese Verfahren gegeben.

2.1 Zwei-Niveau-Passung

Vereinfachend gesprochen besteht die Zwei-Niveau-Passung darin, für eine gegebene Funktion h eine „möglichst gut“ passende Treppenfunktion mit nur zwei Stufen — den *Niveaus* — zu finden. Dabei sind die Höhen der Niveaus frei wählbar, während die Niveaugrenzen durch ein Mengenmodell vorgegeben sind. Die Zwei-Niveau-Passung kann allgemein auf \mathbb{R}^d , $d \in \mathbb{N}$ durchgeführt werden, auch wenn der zweidimensionale Fall in der Bildsegmentierung der bedeutendste ist.

Sei also $d \in \mathbb{N}$ und M eine kompakte Teilmenge des \mathbb{R}^d . Weiterhin sei $\mathcal{A} \subseteq \mathcal{B}(M)$ ein System von Teilmengen von M . Die Aufgabe besteht nun darin, für eine gegebene Funktion $h \in L^2(M, \mathbb{R})$ eine Menge $A \in \mathcal{A}$ sowie zwei Zahlen $\alpha, \beta \in \mathbb{R}$ zu finden, so daß der Wert

$$\int_M (h - \alpha \mathbf{1}_A - \beta \mathbf{1}_{M \setminus A})^2 d\lambda^d \quad (2.1)$$

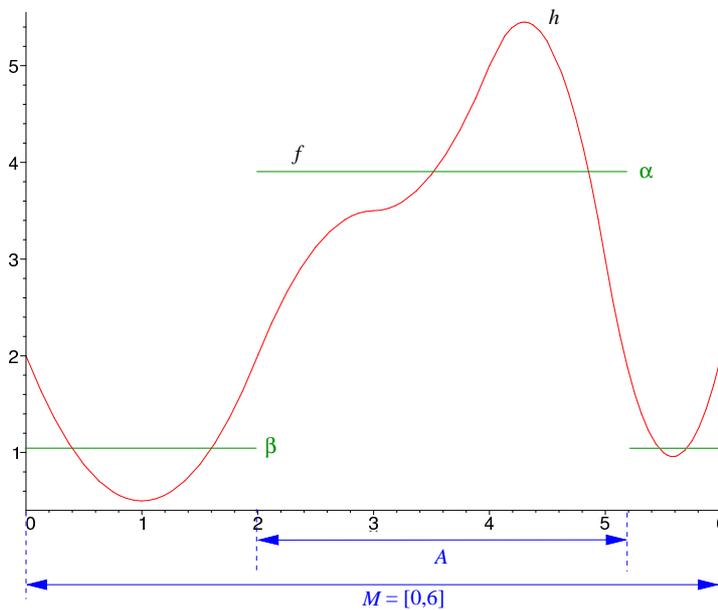


Abbildung 2.1: Zwei-Niveau-Passung einer Funktion $h : M \rightarrow \mathbb{R}$ (rot) durch eine Funktion $f = \alpha \mathbb{1}_A + \beta \mathbb{1}_{M \setminus A}$ (grün). Das Intervall A sowie die Niveaus α und β der Treppenfunktion f sind willkürlich gewählt.

minimal wird. Für $B \in \mathcal{B}(M)$ wird dabei mit $\mathbb{1}_B$ die Indikatorfunktion von B auf M

$$\mathbb{1}_B : M \rightarrow \mathbb{R}, \quad \mathbb{1}_B(x) := \begin{cases} 1, & x \in B \\ 0, & x \notin B \end{cases}$$

bezeichnet. \mathcal{A} steht für das gewählte Mengenmodell (Mengensystem), das je nach Kontext allgemeiner gewählt oder auch stärker eingeschränkt sein kann. Ein sehr allgemeines Mengenmodell wäre zum Beispiel das System aller zusammenhängenden Teilmengen von M . Weiter eingeschränkte Mengenmodelle im zweidimensionalen Fall sind dagegen Rechteckflächen, Kreisscheiben oder auch die in [Don91] vorgeschlagenen kegelförmigen Mengen. In der Regel ist die Wahl eines zu allgemeinen Mengenmodells nicht zu empfehlen, ein extremes Beispiel in dieser Richtung stellt $\mathcal{A} = \mathfrak{P}(M)$ für diskretes M dar. Dieser Fall entspricht dem gänzlichen Fehlen eines Mengenmodells, denn es gelten keinerlei Einschränkungen mehr an \mathcal{A} — jede Wahl ist möglich. Die eigentliche Stärke des Mengenmodells, die Form des zu findenden Bereiches bereits im Voraus bestimmen zu können, geht dabei völlig verloren.

Eine beispielhafte Anwendung von (2.1) an einer eindimensionalen Funktion zeigt Abbildung 2.1. Das Mengenmodell \mathcal{A} besteht hier aus den zusammenhängenden abgeschlossenen Teilmengen von $M = [0, 6]$. Das Intervall $A \in \mathcal{A}$ in Abbildung 2.1 ist willkürlich gewählt, die Höhen α und β der Treppenfunktion f ebenfalls. Es stellt sich nun die Frage, ob für dieses Intervall A die Werte α und β so angepaßt werden können, daß sich die Passung verbessert (d.h. daß der Wert von (2.1) kleiner wird) und ob sogar optimale

Werte α_0 , β_0 und A_0 bestimmt werden können. Gesucht wird also eine optimale Passung, d.h. zumindest ein lokales Optimum der Funktion

$$\tilde{\varphi} : \mathbb{R}^2 \times \mathcal{A} \rightarrow \mathbb{R}, \quad \tilde{\varphi}(\alpha, \beta, A) := \int_M (h - \alpha \mathbb{1}_A - \beta \mathbb{1}_{M \setminus A})^2.$$

Um überhaupt über lokale Optima sprechen zu können, wird als Erstes ein Abstandsbegriff auf \mathcal{A} benötigt. Dazu sei zunächst

$$\delta : \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}, \quad \delta(A, B) := \lambda^d((A \setminus B) \cup (B \setminus A)),$$

wobei λ^d das d -dimensionale Lebesgue-Maß bezeichnet. Allerdings bildet δ auf \mathcal{A} nur eine Pseudometrik, denn für zwei Mengen $A, B \in \mathcal{A}$, die sich nur durch λ^d -Nullmengen unterscheiden, gilt $\delta(A, B) = 0$. Um dies zu umgehen, wird die Äquivalenzrelation

$$\sim \subseteq \mathcal{A} \times \mathcal{A}, \quad A \sim B :\iff \delta(A, B) = 0$$

definiert und statt \mathcal{A} die Quotientenmenge $\mathcal{A}^* := \mathcal{A}/\sim$ betrachtet. Bezeichnet man nun mit δ^* die von δ auf \mathcal{A}^* induzierte Metrik, wird $(\mathcal{A}^*, \delta^*)$ zu einem metrischen Raum.

Bemerkung 2.1.1

Die Wahl der Metrik und die daraus resultierende Wahl der Quotientenmenge \mathcal{A}^ als grundlegendes Mengenmodell wird im wesentlichen dadurch beeinflusst, daß das Mengensystem nur indirekt — über Indikatorfunktion und Integral — in die Optimierung eingeht. In diesem Sinne kann (2.1) als unsensibel gegenüber Änderungen von A , die vom Integral unterdrückt werden, bezeichnet werden. Im Falle der Zwei-Niveau-Passung handelt es sich dabei nur um λ^2 -Nullmengen, die durch den Wechsel von \mathcal{A} auf \mathcal{A}^* eliminiert werden (vgl. dazu auch Abbildung 4.4 auf Seite 61).*

Mit diesen Voraussetzungen kann man nun nach lokalen Minima der jetzt nicht mehr auf \mathcal{A} , sondern auf \mathcal{A}^* definierten Funktion

$$\varphi : \mathbb{R}^2 \times \mathcal{A}^* \rightarrow \mathbb{R}, \quad \varphi(\alpha, \beta, A) := \int_M (h - \alpha \mathbb{1}_A - \beta \mathbb{1}_{M \setminus A})^2$$

suchen. Die Optimierung läuft zunächst über Tripel $(\alpha, \beta, A) \in \mathbb{R}^2 \times \mathcal{A}^*$, wird jedoch durch den folgenden Satz, der das Kernstück der Optimierung in der Zwei-Niveau-Passung bildet, auf eine Optimierung nur über \mathcal{A}^* reduziert:

Satz 2.1.2

Bezeichne im folgenden \emptyset diejenige Äquivalenzklasse aus $\mathcal{B}(M)/\sim$, für die $\lambda^d(\emptyset) = 0$ gilt und sei $\gamma_h : \mathcal{B}(M)/\sim \setminus \{\emptyset\} \rightarrow \mathbb{R}$ definiert durch

$$\gamma_h(B) := \frac{1}{\lambda^d(B)} \int_B h d\lambda^d.$$

Ist unter diesen Voraussetzungen für $\mathcal{A}^* \subseteq \mathcal{B}(M)/\sim$ die Funktion $\psi : \mathcal{A}^* \setminus \{\emptyset, M\} \rightarrow \mathbb{R}$ durch

$$\psi(A) := \gamma_h(A) \int_A h + \gamma_h(M \setminus A) \int_{M \setminus A} h$$

gegeben, ist ein Tripel (α_0, β_0, A_0) genau dann ein lokales Minimum der Funktion φ , wenn $\alpha_0 = \gamma_h(A_0)$, $\beta_0 = \gamma_h(M \setminus A_0)$ und A_0 ein lokales Maximum von ψ ist.

Von der Optimierung dreier Parameter ist man durch Satz 2.1.2 bei einer Optimierung nur über $\mathcal{A}^* \setminus \{\emptyset, M\}$ gelangt. Statt nach einem lokalen Minimum von φ suchen zu müssen, genügt es nun ein lokales Maximum von ψ zu bestimmen. Die optimalen Niveaus α und β zu jedem $A \in \mathcal{A}$ ergeben sich dann automatisch als

$$\alpha = \gamma_h(A) \quad \text{und} \quad \beta = \gamma_h(M \setminus A).$$

Für den Beweis von Satz 2.1.2 sei auf [Don91] verwiesen, hier genügt ein Hinweis auf den folgenden zentralen Zusammenhang von φ und ψ :

$$\varphi(\gamma_h(A), \gamma_h(M \setminus A), A) = \int_M h^2 - \psi(A).$$

Für den Optimierungsprozeß selbst existiert ein weiteres Kriterium, das ausgehend von einer bereits segmentierten Menge A_0 die Suche nach einer neuen Menge A_1 insoweit einschränkt, daß nur noch die Differenzmenge $A_1 \setminus A_0$ (bzw. $A_0 \setminus A_1$) zwischen bereits segmentierter und neuer Menge bei der Berechnung von ψ berücksichtigt werden muß ([Don98]). Seien dazu drei Mengen $A, B, C \in \mathcal{A}^* \setminus \{\emptyset, M\}$ so gewählt, daß sie eine Partition von M bilden. Angenommen, der Optimierungsprozeß hat bereits die Menge A gefunden und es soll nun geprüft werden, ob eine Erweiterung von A auf $A \cup B$ eine Verbesserung in der Passung bringt¹ d.h. ob

$$\psi(A \cup B) > \psi(A)$$

gilt. Da sich für alle Mengen $D \in \mathcal{A}^* \setminus \{\emptyset, M\}$ der Ausdruck $\gamma_h(D) \int_D h$ als

$$\gamma_h(D) \int_D h = \lambda^d(D) \gamma_h^2(D)$$

schreiben läßt, ergibt sich für $\psi(A \cup B) - \psi(A)$ zunächst²

$$\begin{aligned} \psi(A \cup B) - \psi(A) &= \gamma_{A \cup B} \int_{A \cup B} h + \gamma_C \int_C h - \gamma_A \int_A h - \gamma_{B \cup C} \int_{B \cup C} h \\ &= \lambda_{A \cup B} \gamma_{A \cup B}^2 + \lambda_C \gamma_C^2 - \lambda_A \gamma_A^2 - \lambda_{B \cup C} \gamma_{B \cup C}^2. \end{aligned}$$

¹In diesem Fall muß $\forall A, B \in \mathcal{A}^* : A \cup B \in \mathcal{A}^*$ zusätzlich vorausgesetzt werden

²im folgenden Abschnitt werden zur besseren Übersicht für $D \in \mathcal{A}^* \setminus \{\emptyset, M\}$ die Abkürzungen γ_D statt $\gamma_h(D)$ und λ_D statt $\lambda^d(D)$ verwendet

Weiter gilt wegen $A \cap B = \emptyset$ für $\gamma_{A \cup B}$

$$\gamma_{A \cup B} = \frac{\int_{A \cup B} h}{\lambda_{A \cup B}} = \frac{\int_A h + \int_B h}{\lambda_A + \lambda_B} = \frac{\lambda_A \gamma_A + \lambda_B \gamma_B}{\lambda_A + \lambda_B}$$

und analog für $\gamma_{B \cup C}$

$$\gamma_{B \cup C} = \frac{\lambda_B \gamma_B + \lambda_C \gamma_C}{\lambda_B + \lambda_C},$$

so daß sich $\psi(A \cup B) - \psi(A)$ nun in der Form

$$\psi(A \cup B) - \psi(A) = \frac{(\lambda_A \gamma_A + \lambda_B \gamma_B)^2}{\lambda_A + \lambda_B} + \lambda_C \gamma_C^2 - \lambda_A \gamma_A^2 - \frac{(\lambda_B \gamma_B + \lambda_C \gamma_C)^2}{\lambda_B + \lambda_C}$$

darstellen läßt. Für die weiteren Überlegungen wird die Symmetrie zwischen den beiden Termen $\lambda_{A \cup B} \gamma_{A \cup B}^2 - \lambda_A \gamma_A^2$ und $\lambda_{B \cup C} \gamma_{B \cup C}^2 - \lambda_C \gamma_C^2$ ausgenutzt. Aus

$$\psi(A \cup B) - \psi(A) > 0$$

kann dann

$$\frac{(\lambda_A \gamma_A + \lambda_B \gamma_B)^2}{\lambda_A + \lambda_B} - \lambda_A \gamma_A^2 > \frac{(\lambda_B \gamma_B + \lambda_C \gamma_C)^2}{\lambda_B + \lambda_C} - \lambda_C \gamma_C^2,$$

hergeleitet werden, was sich zu

$$\frac{\lambda_B(\lambda_B \gamma_B^2 + 2\lambda_A \gamma_A \gamma_B - \lambda_A \gamma_A^2)}{\lambda_A + \lambda_B} > \frac{\lambda_B(\lambda_B \gamma_B^2 + 2\lambda_C \gamma_C \gamma_B - \lambda_C \gamma_C^2)}{\lambda_B + \lambda_C}$$

und wegen $\lambda_B \neq 0$ weiter zu

$$\frac{\lambda_B \gamma_B^2 + 2\lambda_A \gamma_A \gamma_B - \lambda_A \gamma_A^2}{\lambda_A + \lambda_B} > \frac{\lambda_B \gamma_B^2 + 2\lambda_C \gamma_C \gamma_B - \lambda_C \gamma_C^2}{\lambda_B + \lambda_C} \quad (2.2)$$

umformen läßt. Die beiden Zähler von (2.2) lassen sich weiter vereinfachen: Zunächst gilt für den Zähler $\lambda_B \gamma_B^2 + 2\lambda_A \gamma_A \gamma_B - \gamma_A^2 \lambda_A$ der linken Seite

$$\begin{aligned} \lambda_B \gamma_B^2 + 2\lambda_A \gamma_A \gamma_B - \gamma_A^2 \lambda_A &= \lambda_B \gamma_B^2 + 2\lambda_A \gamma_A \gamma_B - \gamma_A^2 \lambda_A + \underbrace{\lambda_A \gamma_B^2 - \lambda_A \gamma_B^2}_{\text{Nullterm addiert}} \\ &= \gamma_B^2(\lambda_A + \lambda_B) - \lambda_A(\gamma_A - \gamma_B)^2 \end{aligned}$$

Analoges ergibt sich für den Zähler der rechten Seite von (2.2). Setzt man diese Ergebnisse in (2.2) ein, ist $\psi(A \cup B) - \psi(A) > 0$ genau dann wenn

$$\frac{\lambda_A}{\lambda_A + \lambda_B}(\gamma_A - \gamma_B)^2 < \frac{\lambda_C}{\lambda_B + \lambda_C}(\gamma_B - \gamma_C)^2. \quad (2.3)$$

Um 2.3 weiter vereinfachen zu können, muß an dieser Stelle eine weitere Annahme getroffen werden. Diese Annahme beruht auf der Tatsache, daß sowohl das gesuchte Objekt als auch der Hintergrund in der Zwei-Niveau-Passung durch konstante Niveaus approximiert

werden. Man kann daher davon ausgehen, daß sich der bereits gefundene Objektbereich A in seinem Grauwertmittel γ_A deutlich vom Grauwertmittel γ_C des Hintergrundes unterscheidet. Im folgenden wird $\gamma_A > \gamma_C$ angenommen, was einem hellen Objekt auf dunklem Hintergrund entspricht. Für den fraglichen Teilbereich B kann dann angenommen werden, daß sein Mittel γ_B zwischen γ_A und γ_C liegt, also $\gamma_A > \gamma_B > \gamma_C$ gilt. Nähme man nämlich $\gamma_B > \gamma_A$ an, würde daraus sofort $\gamma_{A \cup B} > \gamma_A$ sowie $\gamma_C < \gamma_{B \cup C}$ folgen, was im weiteren Verlauf der Überlegungen automatisch zu $\psi(A \cup B) > \psi(A)$ führen würde. Eine Hinzunahme von B zu A bedürfte in diesem Fall keiner weiteren Untersuchung. Analoge Überlegungen gelten, wenn $\gamma_B < \gamma_C$ angenommen wird.

Ist also $\gamma_A > \gamma_B > \gamma_C$ läßt sich (2.3) zu

$$\sqrt{\lambda_A} \sqrt{\lambda_B + \lambda_C} (\gamma_A - \gamma_B) < \sqrt{\lambda_C} \sqrt{\lambda_A + \lambda_B} (\gamma_B - \gamma_C)$$

und weiter zu

$$\gamma_B > \frac{\gamma_A \sqrt{\lambda_A(\lambda_B + \lambda_C)} + \gamma_C \sqrt{\lambda_C(\lambda_A + \lambda_B)}}{\sqrt{\lambda_A(\lambda_B + \lambda_C)} + \sqrt{\lambda_C(\lambda_A + \lambda_B)}} \quad (2.4)$$

vereinfachen. Um nun zu überprüfen, ob unter den obigen Bedingungen das Hinzufügen einer Menge B zu der bereits gefundenen Paßmenge A eine Verbesserung in der Zielfunktion ψ ergibt, muß man nicht mehr die gesamte Funktion ψ auswerten, sondern kann sich auf Kriterium (2.4) zurückziehen. Analoge Kriterien ergeben sich, wenn bereits $A \cup B$ gefunden wurden und geprüft werden soll, ob eine Entfernung von B aus A eine Verbesserung ergibt.

Die gesamten Überlegungen lassen sich natürlich auch durchführen, wenn ein dunkles Objekt auf einem hellen Hintergrund gefunden werden soll. In diesem Fall wird $\gamma_A < \gamma_C$ vorausgesetzt, der Rest ergibt sich analog.

In der vorliegenden Form mag Kriterium (2.4) nicht viel handlicher sein als eine direkte Auswertung von ψ . Eine weitere große Vereinfachung ergibt sich aber, wenn λ_B im Vergleich zu λ_A und λ_C vernachlässigt werden kann. In diesem Fall vereinfacht sich (2.4) zu

$$\gamma_B > \frac{1}{2}(\gamma_A + \gamma_C). \quad (2.5)$$

Kriterium (2.5) ist sehr einfach anzuwenden und bewirkt bei einer Implementierung des Verfahrens einen deutlichen Geschwindigkeitsvorteil gegenüber einer direkten Auswertung der Zielfunktion ψ . Es sollte aber beachtet werden, daß die Annahme einer infinitesimalen Größe von B in praktischen Bildsegmentierungsproblemen nicht erfüllt werden kann: Wenn bereits die Grundmenge M durch eine Pixelmenge repräsentiert wird, kann die kleinste für B erreichbare Größe oberhalb der leeren Menge nur 1 Pixel sein. Dennoch kann Kriterium (2.5) als gute Näherung an den realen Sachverhalt angewendet werden.

In [Don91] werden zwei weitere Forderungen an \mathcal{A}^* gestellt, die unten vorgestellt werden. Der dabei im folgenden benutzte Operator \subseteq ist bezüglich der Äquivalenzklassen zu verstehen: Für $A, B \in \mathcal{A}^*$ gilt $A \subseteq B$ falls $A \neq B$ ist und Mengen A_1 aus der Äquivalenzklasse von A und B_1 aus der Äquivalenzklasse von B existieren so daß $A_1 \subseteq B_1$ gilt. Analoges gilt für die Operatoren \cup und \cap sowie für die Intervallbildung:

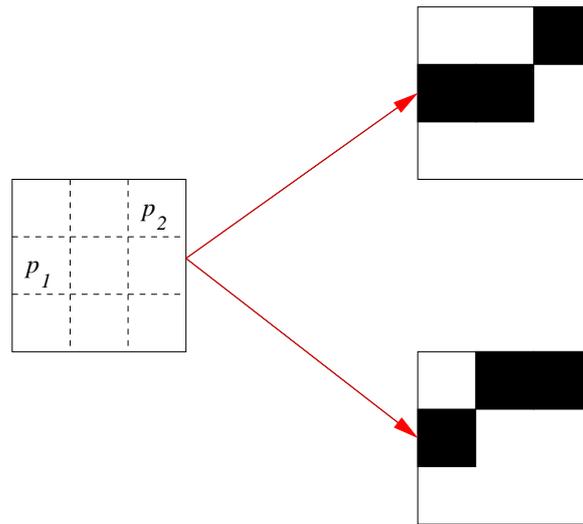


Abbildung 2.2: Zwei Diskretisierungsmöglichkeiten einer Linie vom Pixel p_1 zum Pixel p_2 durch den Bresenham-Algorithmus für $M = \{1, 2, 3\}^2$.

1. $\forall A, B \in \mathcal{A}^* : A \cup B \in \mathcal{A} \wedge A \cap B \in \mathcal{A}$ und
2. $\forall A, B \in \mathcal{A}^*, A \subseteq B: \lambda(B) = \sup_{C \in]A, B[} \lambda(C) \wedge$
 $\lambda(A) = \inf_{C \in]A, B[} \lambda(C)$ (Dichteigenschaften).

Diese zwei Forderungen dienen der leichteren Klassifikation möglicher lokaler Optima. Gleichzeitig wird aber vor allem durch die erste Forderung auch das Mengensystem \mathcal{A}^* sehr stark eingeschränkt. In [Don91] werden kegelsternförmige Mengen vorgeschlagen, die den Forderungen genügen und vor allem zur Passung kreisförmiger Objekte verwendet werden können. Die Verwendung rotierbarer Strukturen ist aber durch die erste Forderung nur sehr eingeschränkt möglich, so daß beide Forderungen im folgenden nicht weiter verwendet werden.

Ohne die obigen Forderungen besitzt man zunächst einen sehr großen Vorrat an möglichen Mengenmodellierungen. Da diese Mengenmodelle in der praktischen Bildverarbeitung aber stets auf einer diskreten Grundmenge M basieren, stößt man sehr schnell auf Probleme. Das folgende Beispiel verdeutlicht eins dieser Probleme, nämlich das der Mehrdeutigkeit:

Beispiel 2.1.3

Ein mögliches Mengensystem über einem diskreten Bildraum $M = \{1, \dots, n\}^2$, $n \in \mathbb{N}$ ist die Menge aller geraden Linien zwischen zwei Pixeln $p_1, p_2 \in M$. Da dieses Mengensystem über M definiert ist und M diskret ist, handelt es sich bei den Linien im Mengensystem ebenfalls um diskrete Linien. Zur Erzeugung diskreter Strecken zwischen zwei Pixeln eignet sich der Bresenham-Algorithmus (siehe dazu zum Beispiel [FDF⁺94]). Die allgemeinste Form von \mathcal{A} — im diskreten Fall identisch zu \mathcal{A}^* — ist damit die fol-

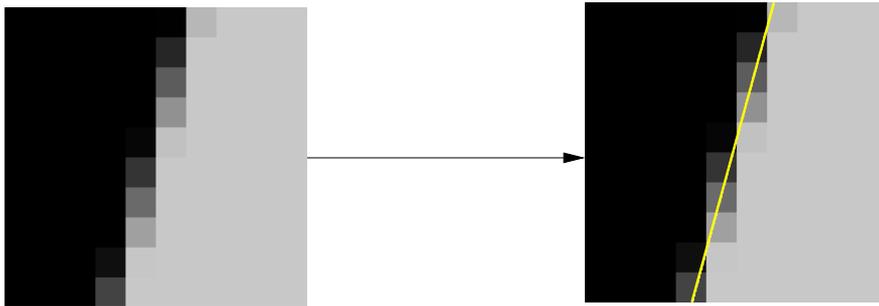


Abbildung 2.3: Ein Bildausschnitt mit einer diskreten Bildkante. Die kontinuierliche Kante, die der Diskretisierung im linken Bild zugrunde lag, ist im rechten Bild als Gerade einkisziert.

gende:

$$\mathcal{A} = \{A \subseteq M \mid \text{es existieren Punkte } p_1, p_2 \in M, \text{ so daß } A \text{ alle Pixel einer durch den Bresenham-Algorithmus erzeugten Linie zwischen } p_1 \text{ und } p_2 \text{ enthält}\}.$$

Ist \mathcal{A} nun eindeutig in dem Sinne, daß zwischen zwei Punkten p_1 und p_2 innerhalb von M eine eindeutig bestimmte Linie existiert? Leider ist das nicht der Fall wie Abbildung 2.2 verdeutlicht. Beide Diskretisierungen in Abbildung 2.2 sind optimal bezüglich des beim Bresenham-Algorithmus verwendeten Quadratmittelfehlers. Und je nachdem, welcher der beiden Punkte bei der Berechnung als Start- und welcher als Endpunkt verwendet wird, werden durch den Algorithmus auch tatsächlich beide Versionen erzeugt. Um Eindeutigkeit im Mengensystem zu erreichen, muß \mathcal{A} also weiter eingeschränkt werden. Dies läuft in der Praxis auf eine zusätzliche Überprüfung jeder erzeugten Pixellinie hinaus, was den algorithmischen Aufwand — und damit die Laufzeit des Verfahrens — deutlich erhöht.

Das in Beispiel 2.1.3 dargestellte Nichteindeutigkeitsproblem vervielfacht sich, wenn als Mengenmodell komplexere Strukturen wie etwa Polygonzüge angedacht werden. Die Zwei-Niveau-Passung wird durch diese Problematik in ihrer Algorithmik unnötig aufgebläht, denn gerade ein Polygonzug ist sehr einfach über die Menge seiner Eckpunkte beschreibbar. Einen weitaus einfacheren Zugang würde daher ein Mengenmodell bieten, das nicht auf dem Grundraum M aufsetzt und daher nicht auf Pixelbasis beschrieben werden müßte. Diese Art der kontinuierlichen Modellierung würde einen zusätzlichen Vorteil bieten, der in der folgenden Bemerkung aufgezeigt wird:

Bemerkung 2.1.4

Betrachtet man die Ränder nicht-achsenparalleler Kanten in digitalen Bildern, fällt auf, daß die Grauwerte am Rand die „Schräglage“ der Kante im Subpixelbereich angeben. Abbildung 2.3 zeigt ein Beispiel für eine derartige digitale Kante. Die unterschiedlichen Grauwertabstufungen von schwarz zu grau sind im linken Bild deutlich sichtbar. Mit

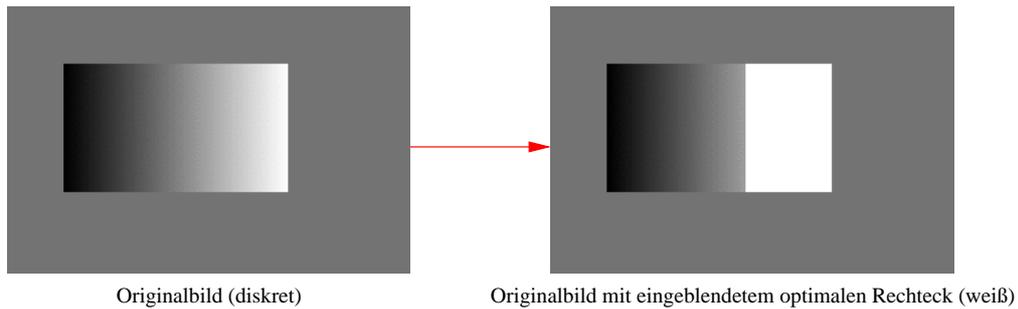


Abbildung 2.4: Rechteck mit inhomogenem Grauwert (links) mit dem Ergebnis der Zwei-Niveau-Passung (rechts). Als \mathcal{A}^* wurde die Menge aller achsenparallelen Rechtecke gewählt. Das gefundene optimale Rechteck wurde rechts weiß in das Originalbild eingeblendet.

dem Vorwissen, daß es sich bei der beobachteten Kante ursprünglich um eine gerade Linie handelte, sollte sich diese Information nutzen lassen können, um die ursprüngliche Position der Kante im Bild vor ihrer Digitalisierung zu bestimmen (wie im rechten Bild von Abbildung 2.3 einskizziert). Da die Zwei-Niveau-Passung auf digitalen (und damit diskreten) Bildern aber nur diskrete Mengenmodelle erlaubt, läßt sich dieser Wunsch so zunächst nicht realisieren.

Bemerkung 2.1.4 führt zur Problematik der Objektpassung im Subpixelbereich. Offensichtlich geben die Grauwerte im Kantenbereich von Abbildung 2.3 einen Hinweis auf die optimale Lage einer kontinuierlichen Kante. Beobachtungen dieser Art lassen sich in das Mengenmodell der Zwei-Niveau-Passung nicht einbeziehen, wünschenswert wäre daher eine Erweiterung der Zwei-Niveau-Passung, die kontinuierliche Mengenmodelle mit berücksichtigt.

Ein weiteres Problem der Zwei-Niveau-Passung ist, daß in jedem zu segmentierenden Bild $h \in L^2(M, \mathbb{R})$ nach Voraussetzung sowohl Objekt als auch Hintergrund weitgehend homogene Grauwerte besitzen müssen, damit sie durch den Paßprozeß gefunden werden können. Diese Einschränkung ist bereits in Gleichung (2.1) von Seite 13 verankert, denn h wird durch eine stückweise konstante Funktion $\alpha \mathbb{1}_A + \beta \mathbb{1}_{M \setminus A}$ approximiert. Abbildung 2.4 zeigt ein Beispiel für einen Fall, bei der das zu findende Objekt (ein achsenparalleles Rechteck) einen in x -Richtung inhomogenen Grauwertverlauf aufweist. Die Zwei-Niveau-Passung kann hier naturgemäß nicht das gewünschte Ergebnis, nämlich das gesamte Rechteck, liefern. Das in Abbildung 2.4 im rechten Bild eingeblendete weiße Rechteck entspricht der Funktion $\mathbb{1}_{A_0}$ der optimalen Menge A_0 , die Hintergrundmenge $M \setminus A_0$ wurde nicht einskizziert.

Eine Verallgemeinerung von (2.1) auf die Approximation eines Bildes $h \in L^2(M, \mathbb{R})$ durch nicht-konstante Funktionen wäre relativ einfach: Die Indikatorfunktionen der stückweise konstanten Funktion $\alpha \mathbb{1}_A + \beta \mathbb{1}_{M \setminus A}$ müssten nur durch allgemeinere Funktionen $e_A^{(1)}, e_A^{(2)} \in L^2(M, \mathbb{R})$ ersetzt werden. Allerdings ist kaum zu erwarten, daß Satz 2.1.2 dann noch in der vorliegenden Form gültig ist. Ob und unter welchen Voraussetzungen

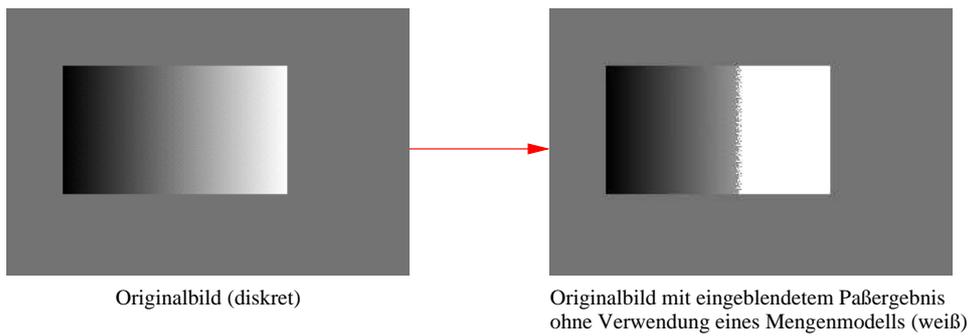


Abbildung 2.5: Paßergebnis des Bildes aus Abbildung 2.4, wenn $\mathcal{A}^* = \mathfrak{P}(M)$ gewählt wird (weiß eingeblendet in das rechte Bild). Auffallend ist hier der ausgefranste linke Rand im segmentierten Bereich.

an die Funktionen $e_A^{(i)}$ diese Idee trotzdem funktioniert, ist Thema von Kapitel 3.

Doch auch wenn die Passung in Abbildung 2.4 nicht das gewünschte Ergebnis lieferte, zeigt sie doch die Wichtigkeit des zugrundegelegten Mengenmodells auf. Als Mengenmodell in Abbildung 2.4 wurde die Menge aller achsenparallelen Rechtecke gewählt. Diese Voraussetzung scheint zunächst überzogen zu sein, da das Originalbild bereits eine sehr homogene Struktur besitzt. Auch ohne spezielles Mengenmodell kann man daher durchaus erwarten, als Paßergebnis ein Rechteck zu erhalten. Das Ergebnis dieser Überlegungen ist in Abbildung 2.5 dargestellt. Im Unterschied zur Passung aus Abbildung 2.4 wurde hier $\mathcal{A} = \mathfrak{P}(M)$ gewählt, wobei M die diskrete Pixelmenge bezeichnet die dem Bild zugrunde liegt. Da \mathcal{A} nicht weiter eingeschränkt wurde, entspricht diese Wahl dem gänzlichen Fehlen eines Mengenmodells. Der unregelmäßige Rand auf der linken Seite der in Abbildung 2.5 segmentierten Menge zeigt, daß der Grauwertverlauf im Originalbild dann doch nicht so gleichmäßig war, wie ursprünglich angenommen. Um aus der so segmentierten Menge nachträglich ein Rechteck zu extrahieren sind weitere Schritte wie Lückenfüllung und Kantenglättung nötig. Schränkt man dagegen \mathcal{A} gleich bei der Segmentierung entsprechend ein, sind diese Schritte überflüssig. Eine Erweiterung der Zwei-Niveau-Passung sollte daher möglichst nicht auf Kosten des Mengenmodells geschehen.

2.2 Funktionsquantisierungen

Die Zwei-Niveau-Passung reiht sich in die Klasse der *Funktionsquantisierungen* ein, wie sie in [Käm00] für Funktionen $f : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}$ beschrieben werden. Bei Quantisierungsproblemen dieser Art ist eine Menge \mathcal{S} von Funktionen $s : D \rightarrow \mathbb{R}$ gegeben. Für eine gegebene Funktion $f : D \rightarrow \mathbb{R}$ wird dann das Infimum

$$\inf_{s \in \mathcal{S}} \|f - s\|_p$$

gesucht. Mit $\|\cdot\|_p$ wird dabei die Standard- p -Norm bezeichnet ($p \in [1, \infty]$). Die Menge \mathcal{S} wird *Quantisierungsmenge* genannt und besteht normalerweise aus stückweise definierten

Funktionen mit $n > 0$ Stützstellen. In [Käm00] werden Funktionen

$$s_n := \sum_{i=1}^{n-1} y_i \mathbb{1}_{]x_i, x_{i+1}],} \quad n > 2$$

verwendet, für die $x_1 < \dots < x_n \in \mathbb{R}$ und $y_1, \dots, y_{n-1} \in \mathbb{R}$ gilt. Für $x < x_1$ und $x > x_n$ wird $s_n(x) = 0$ gesetzt. Statt dem allgemeinen \mathcal{S} werden nun die Mengen \mathcal{S}_n betrachtet, die nur Funktionen s_n mit höchstens n Stützstellen enthalten. Das so definierte Quantisierungsproblem lautet dann

$$\inf_{s_n \in \mathcal{S}_n} \|f - s_n\|_p.$$

Dieses Quantisierungsproblem wird in [Käm00] für $p \in \{1, 2, \infty\}$ und $f \in \mathcal{C}(D, \mathbb{R}) \cap L^p(D, \mathbb{R})$ ($D \in \{\mathbb{R}, \mathbb{R}_{\geq}, [a, b]\}$) betrachtet. Es werden Ergebnisse zur Existenz lokaler Optima gegeben und Algorithmen zur optimalen Quantisierung für den Fall $p = 2$ entwickelt. Allerdings beschränkt sich die Arbeit in [Käm00] auf die Betrachtung eindimensionaler Funktionen. Sollten die hier entwickelten Verfahren bei der Segmentierung eines zweidimensionalen Bildes angewendet werden, müßte das vorliegende Bild erst in eine eindimensionale Version umgewandelt werden. Üblicherweise geschieht dies durch Histogrammbildung. Es liegt auf der Hand, daß das Einbringen von a-priori-Informationen über die Form des zu findenden Objektes (eine zweidimensionale Mengenmodellierung) auf diese Weise kaum möglich ist. Auch eine eindimensionale Mengenmodellierung — denkbar wären zum Beispiel zusammenhängende Mengen — ist auf Histogrammfunktionen nicht sehr sinnvoll, da nicht unbedingt ein Zusammenhang zwischen zusammenhängenden Mengen im Bildhistogramm und dem originalen Bild bestehen muß. Natürlich ist es dennoch möglich, Segmentierungsalgorithmen auf Basis der Verfahren in [Käm00] zu entwickeln, wie die Arbeit in [KK98] zeigt. Im Gegensatz zu [Käm00] wird die Ausgangsfunktion f hier von vornherein als diskret angenommen. Aber auch in dieser Arbeit wird f als eindimensional vorausgesetzt, was bedeutet, daß wiederum als erster Schritt die Funktion f aus einem gegebenen Bild extrahiert werden muß. Damit enthält dieses Verfahren im Hinblick auf eine zweidimensionale Mengenmodellierung dieselben Nachteile wie die Arbeit in [Käm00].

In den Termini der Einleitung bedeutet dies: Für eine *Objektsegmentierung* sind die Arbeiten in [Käm00] und [KK98] eher weniger geeignet. Sie bieten aber gute Möglichkeiten in einer *Bildsegmentierung*, die kein weiteres Vorwissen über das zu findende Objekt in einem Bild verlangt.

2.3 Methoden der Variationsrechnung

Die Variationsrechnung ist eine in der Bildverarbeitung häufig verwendete Methode. Ihr Ziel ist es, die Ränder von Bildregionen möglichst genau durch stückweise stetige Funktionen zu approximieren. Die Approximation erfolgt dabei über die Minimierung

eines Energieterms, in den Eigenschaften wie Glätte und Krümmung der Bildfunktion einfließen. Dieser Energieterm wird in [MS85] oder auch in [BZ87] wie folgt definiert: Bezeichne $h \in L^2(M, \mathbb{R})$ das über einem kompakten Ausschnitt $M \subseteq \mathbb{R}^2$ gegebene Bild und $\mathcal{U} \subset \mathcal{L}^2(M, \mathbb{R})$ einen geeigneten Raum stückweise differenzierbarer Funktionen von M nach \mathbb{R} . Gesucht ist nun eine Funktion $u \in \mathcal{U}$, für die das folgende Funktional

$$E(u) = D(u) + S(u) + P(u)$$

minimal wird. Dabei ist

- das Funktional D (*Data*), definiert durch

$$D(u) = \int_M (u - f)^2 d\lambda^2,$$

ein Maß für die Bildtreue der gesuchten Funktion,

- das Funktional S (*Spine*), definiert durch

$$S(u) = \beta^2 \int_M \|\nabla u\|_2^2 d\lambda^2,$$

ein Maß für die Steigungen der gesuchten Funktion u und

- das Funktional P (*Penalty*) ein „Strafmaß“ für jede Diskontinuität der gesuchten Funktion.

Der Parameter β im Funktional S ist ein vom Benutzer zu wählender Parameter, der den Einfluß der Steigungen der zu suchenden Funktion auf das Optimum festlegt. Für den Strafterm P existieren mehrere Varianten. In [BZ87] wird er auf

$$P(u) = \alpha \text{length}(K)$$

gesetzt, wobei K die Menge aller Unstetigkeitsstellen bezeichnet und α ein weiterer vom Benutzer zu wählender Parameter ist. P ist nun ein Maß für die gesamte Konturlänge, d.h. eine Funktion u , die den Energieterm E minimiert, minimiert dann auch die Konturlänge. Von Mumford und Shah wurde ein anderer Strafterm vorgestellt ([MS85]), in den die Konturlänge K über ihr (eindimensionales) Hausdorffmaß \mathcal{H}^1 eingeht. Der gesamte Energieterm hat dann die Form

$$E(u, K) = \int_M (u - f)^2 d\lambda^2 + \beta^2 \int_M \|\nabla u\|_2^2 d\lambda^2 + \alpha \mathcal{H}^1(K)$$

Beiden Vorschlägen ist gemein, daß P — und damit die Diskontinuität der gesuchten Funktion u — über den Parameter α vom Benutzer gesteuert werden kann.

Dieser eigentlich kontinuierliche Ansatz muß für die Anwendung auf digitale Bilder durch eine geeignete diskrete Version approximiert werden. Es stellt sich dabei heraus, daß die

Optimierung der Energiefunktion zweistufig verlaufen kann: Zuerst werden die Unstetigkeitsstellen von u bestimmt, die das Bild in verschiedene Segmente aufteilen, und in einem zweiten Schritt die Funktion u innerhalb der Segmente eingepaßt. Die Sprunghöhen sowie die Glattheit von u innerhalb der verschiedenen Segmente werden dabei über die beiden Parameter α und β vom Benutzer vorgegeben.

Im Hinblick auf diese Arbeit ist der letzte Satz entscheidend: Zwar bieten die Parameter α und β eine gute Möglichkeit, die Eigenschaften von u zu steuern, allerdings kann darüber hinaus kein weiteres Vorwissen über eine mögliche Objektform übergeben werden. Für die Methoden der Variationsrechnung gelten damit die gleichen Überlegungen, die bereits zur Funktionsquantisierung angestellt wurden: Sie sind für eine *Bildsegmentierung* sehr gut, für eine *Objektsegmentierung* dagegen eher weniger geeignet.

2.4 Objektlokalisierung durch Adaption parametrischer Grauwertmodelle

Ein Verfahren zur subpixelgenauen Objektpassung — die *Objektlokalisierung durch Adaption parametrischer Grauwertmodelle*, hier im folgenden mit *Grauwertmodelladaption* bezeichnet — wird von Christian Drewniok in seiner Dissertation beschrieben ([Dre98]). Dieses Verfahren wurde von ihm primär dazu entwickelt, um Kanaldeckel aus Luftbildaufnahmen zu extrahieren, daher der Begriff der *Objektlokalisierung*. Wegen der groben Auflösung der Luftbilder ist eine Passung der Kanaldeckel auf diskreter Pixelbasis nicht möglich und es muß auf eine Subpixelpassung ausgewichen werden. Daher basiert die Grauwertmodelladaption auf kontinuierlichen Mengenmodellierungen, was in Bemerkung 2.1.4 als wünschenswerte Erweiterung in der Zwei-Niveau-Passung herausgestellt wurde. Der Algorithmus setzt auf Arbeiten in [Hue71], [NB86] und [Roh92] auf, die sich mit der Rekonstruktion von Grauwertkanten und Grauwertecken durch Minimierung eines quadratischen Fehlers befassen und wird in der Grauwertmodelladaption auf rechteckige und kreisförmige Objekte (sowie deren Überlagerungen) erweitert.

Die grundlegende Überlegung der Grauwertmodelladaption ist die folgende: In ein diskretes Luftbild I soll ein gegebenes Modell $M(\vec{p})$ so eingepaßt werden, daß der Fehler

$$\mathcal{E}(\vec{p}) = \sum_k (I_k - M_k(\vec{p}))^2$$

minimal wird. Der reelle Vektor \vec{p} beschreibt dabei die freien Parameter eines (noch zu spezifizierenden) Modells $M(\vec{p})$ und $I_k = I(\vec{x}_k)$ die an den Bildpositionen \vec{x}_k vorliegenden Grauwerte des Bildes I . Eine Objektlokalisierung besteht dann in der Minimierung von \mathcal{E} über die Parameter \vec{p} .

Da eine subpixelgenaue Passung angestrebt wird, muß das Modell $M(\vec{p})$ eine höhere Genauigkeit aufweisen als das gegebene diskrete Bild. Ein solches (kontinuierliches) Modell wird in [Dre98] als *parametrisches Grauwertmodell* bezeichnet. Zur Formulierung eines

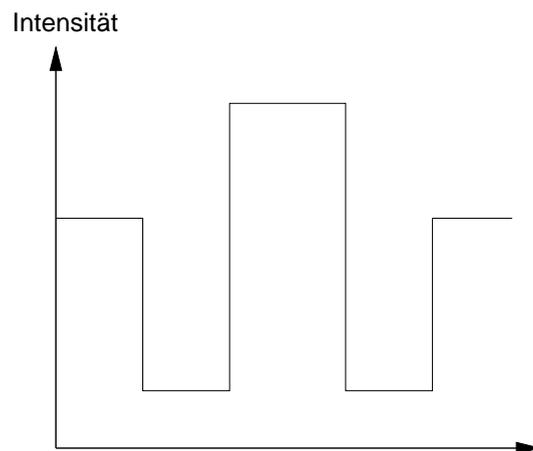


Abbildung 2.6: Mögliches ideales Intensitätsprofil \tilde{I} . Die Höhe der einzelnen Niveaus ist variabel, nur die Anzahl der Stufen ist entscheidend.

solchen Grauwertmodelles wird zunächst angenommen, daß es sich aus mehreren homogenen Teilflächen zusammensetzt. „Homogen“ bedeutet dabei, daß pro Teilfläche konstante Reflexionseigenschaften gelten, Texturen oder Spiegelungen werden nicht berücksichtigt. Jede Teilfläche reflektiert unter dieser Annahme einen konstanten Anteil der einfallenden Lichtstrahlung. Im Idealfall ergibt sich dadurch für jedes Objekt je nach Anzahl der Teilflächen ein stufenförmiges Intensitätsprofil \tilde{I} wie in Abbildung 2.6 skizziert.

Alle weiteren Effekte im Bildentstehungsprozeß wie zum Beispiel Streuung in der Atmosphäre, Verschmierungen durch Linsenoptik und Digitalisierungsprozeß (siehe zum Beispiel [Bur96] für eine ausführlichere Beschreibung), werden in der *point-spread function* (PSF) zusammengefaßt. Das resultierende Signal \hat{I} ergibt sich dann durch eine Faltung der PSF mit dem idealen Signal \tilde{I} . Die PSF selbst wird im allgemeinen als Gaußförmig (mit Mittelwert 0) angenommen, eine Anpassung an die jeweils aktuelle Situation erfolgt über deren Standardabweichung. Das so erhaltene verschmierte Signal wird schließlich an diskreten Stellen abgetastet (Abbildung 2.7), man erhält ein diskretes Bildsignal I .

Der letzte Schritt in der Bildentstehung nach [Dre98] ist die additive Überlagerung des diskreten Signals mit einem normalverteilten Rauschen. Dieser Schritt ist aber für die Modellbildung nicht wesentlich, so daß er hier nicht näher ausgeführt wird.

Die in [Dre98] entwickelten parametrischen Grauwertmodelle folgen dem oben beschriebenen Bildentstehungsprozeß: Ein kontinuierliches, zweidimensionales Signal wird durch eine Gaußsche PSF verschmiert und das Ergebnissignal diskret abgetastet. Die PSF ist dabei gegeben durch

$$G_\sigma(x, y) := g_\sigma(x) \cdot g_\sigma(y)$$

für

$$g_\sigma(x) := \frac{1}{\sigma} \cdot g\left(\frac{x}{\sigma}\right) \quad \text{und} \quad g(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

Der freie Parameter $\sigma > 0$ in G_σ (bzw. g_σ) gibt dabei den Grad der Verschmierung an.

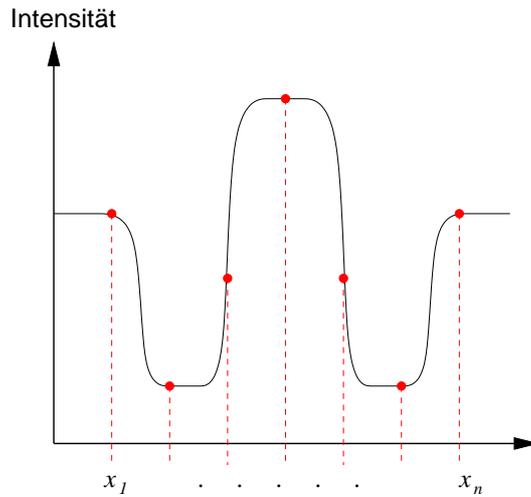


Abbildung 2.7: Verschmiertes (mit PSF gefaltetes) Intensitätssignal \hat{I} mit diskreten Abtastpunkten x_1, \dots, x_n .

Jedes Grauwertmodell läßt sich nun als Faltung von G_σ mit einer entsprechenden Funktion, die das ideale Signal darstellt, modellieren. Eine vertikale Stufe an Position $x_0 \in \mathbb{R}$ zum Beispiel läßt sich also mit Hilfe von

$$U_{x_0}(x) := u(x - x_0), \quad \text{wobei} \quad u(x) := \begin{cases} 0 & \text{für } x < 0 \\ 1 & \text{für } x \geq 0 \end{cases}$$

als $U_{x_0} * G_\sigma$ darstellen. U_{x_0} besitzt als freien Parameter die Position x_0 der Kante, so daß das Gesamtmodell zunächst zwei freie Parameter (die Verschmierung σ und die Position x_0) besitzt. Diese Modellfunktion $U_{x_0} * G_\sigma$ läßt sich vereinfachen zu

$$\begin{aligned} (U_{x_0} * G_\sigma)(x, y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G_\sigma(x - a, y - b) \cdot U_{x_0}(a) \, da \, db \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_\sigma(x - a) \cdot g_\sigma(y - b) \cdot u(a - x_0) \, da \, db \\ &= \int_{-\infty}^{\infty} g_\sigma(y - b) \, db \int_{-\infty}^{\infty} g_\sigma(x - a) \cdot u(a - x_0) \, da \\ &= \int_{x_0}^{\infty} g_\sigma(x - a) \, da \\ &= \int_{-\infty}^{\frac{x-x_0}{\sigma}} g(a) \, da =: \phi_\sigma(x - x_0). \end{aligned}$$

Für die Funktion ϕ_σ gilt

$$\begin{aligned}\phi_\sigma(z) &:= \frac{1}{\sigma} \int_{-\infty}^z g\left(\frac{a}{\sigma}\right) da \\ &= \int_{-\infty}^z g_\sigma(a) da =: \phi\left(\frac{z}{\sigma}\right).\end{aligned}$$

Die Funktion ϕ wird auch *Fehlerfunktion* genannt.

Die durch die Modellfunktion $U_{x_0} * G_\sigma$ definierte Stufe hat bis jetzt die Höhe 1. Um das zu korrigieren, werden zwei lineare Faktoren h_0 (für das Grundniveau) und h_1 (für die Höhenänderung zum Stufenniveau) eingeführt. Das Gesamtmodell

$$M_{\text{step}}(x, y) := h_0 + h_1 \cdot \phi\left(\frac{x - x_0}{\sigma}\right)$$

beschreibt dann eine (mit Standardabweichung σ) Gaußverschmierte vertikale Stufe im Punkt x_0 , die den Übergang zwischen zwei Grauwertplateaus der Höhen h_0 sowie $h_0 + h_1$ bildet. Die beiden Faktoren h_0 und h_1 sind zusätzliche freie Parameter des Gesamtmodells M_{step} , das insgesamt — zusammen mit den beiden bereits vorhandenen freien Parametern σ und x_0 — über vier freie Parameter verfügt. M_{step} entspricht dem Signal \hat{I} aus Abbildung 2.7.

Soll nun in einem gegebenen Bild — bzw. einem gegebenen Bildausschnitt — I eine vertikale Stufe gefunden werden, wird die zugehörige Fehlerfunktion

$$\mathcal{E} : \mathbb{R}_+ \times \mathbb{R}^3 \rightarrow \mathbb{R}, \quad \mathcal{E}(\sigma, x_0, h_0, h_1) := \sum_{(x,y)} (I(x, y) - M_{\text{step}}(x, y))^2$$

über ein nichtlineares Abstiegsverfahren (siehe dazu zum Beispiel [Spe93]) minimiert und dadurch die freien Modellparameter von M_{step} optimiert. In der Fehlerfunktion \mathcal{E} steckt implizit die Punktauswertung von M_{step} an diskreten Punkten (x, y) .

Mit Hilfe von M_{step} lassen sich einfache Rechteckstrukturen zusammensetzen. In [Dre98] werden außerdem Modelle für gedrehte Rechteckstrukturen angegeben. Alle diese Grundmodelle lassen sich additiv zu komplizierteren Strukturen überlagern. Die auf diese Weise entstandenen Modelle heißen *Superpositionen*.

Ein weiteres wichtiges Grundmodell neben dem eben beschriebenen Stufenmodell ist das Kreisscheibenmodell. Eine Kreisscheibe mit Radius $R > 0$ und Zentrum im Nullpunkt besitzt das ideale Signal

$$D_R : \mathbb{R}^2 \rightarrow \mathbb{R}, \quad D_R(x, y) := \begin{cases} 1, & x^2 + y^2 \leq R^2 \\ 0, & \text{sonst} \end{cases}$$

Zur Berechnung des Gaußverschmierten Bildmodells braucht man aus Symmetriegründen nur Punkte entlang der x -Achse zu berücksichtigen, denn für alle Punkte $(x, y) \in \mathbb{R}^2$ gilt

$$(D_R * G_\sigma)(x, y) = (D_R * G_\sigma)(\sqrt{x^2 + y^2}, 0). \quad (2.6)$$

Das mit G_σ gefaltete Signal hat dann die Form

$$\begin{aligned} (D_R * G_\sigma)(x, 0) &= \int_{-R}^R \int_{-\sqrt{R^2-a^2}}^{\sqrt{R^2-a^2}} G_\sigma(x-a, -b) db da \\ &\quad \vdots \\ &= 2 \int_{-R}^R g_\sigma(x-a) \phi_\sigma\left(\sqrt{R^2-a^2}\right) da - (\phi_\sigma(x+R) - \phi_\sigma(x-R)) \end{aligned}$$

([Dre98, Kap. 3.6]). Diese Modellfunktion kann ebenso wie das Stufenmodell zu einem Gesamtmodell mit Grundniveau h_0 und Höhenänderung h_1 zum Kreisniveau ausgebaut werden. Eine Superpositionsbildung mit anderen Modellen ist ebenfalls machbar.

Zum Schluß noch ein Wort zur Implementierung: Die Funktion ϕ ist analytisch nicht bestimmbar und muß daher numerisch approximiert werden. Das kann über eine Approximation der Funktion

$$\text{erf} : \mathbb{R}_+ \rightarrow \mathbb{R}, \quad \text{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt$$

erfolgen, denn es gilt der Zusammenhang

$$\phi(z) = \frac{1}{2} \left(1 + \text{erf} \left(\frac{z}{\sqrt{2}} \right) \right).$$

Zur Bestimmung des Teilterms

$$\int_{-R}^R g_\sigma(x-a) \phi_\sigma\left(\sqrt{R^2-a^2}\right) \tag{2.7}$$

aus dem Kreisscheibenmodell wird in [Dre98] eine Romberg-Integration verwendet. Einzelheiten zur Implementierung der Funktion erf bzw. der Integrationsmethode finden sich zum Beispiel in [PTVF96].

Im Gegensatz zu den in den Abschnitten 2.2 und 2.3 vorgestellten Verfahren legt die Grauwertmodelladaption viel Wert auf eine zugrundeliegende zweidimensionale Struktur. Diese Struktur ist in etwa dem Mengenmodell in der Zwei-Niveau-Passung vergleichbar. Allerdings werden hier keine diskreten Mengenmodelle verwendet, sondern jede Struktur wird durch reelle Parameter beschrieben, was zu kontinuierlichen Mengenmodellen und dadurch zu Subpixelpassungen führt.

Für inhomogen beleuchtete Objekte (wie in Abbildung 2.4) bietet die Grauwertmodelladaption jedoch keine Lösung. Das liegt an der Grundvoraussetzung an die Modellbildung, die besagt daß jede einzelne Teilfläche einen homogenen Grauwert besitzen muß. Eine Erweiterung auf nicht-homogene Teilflächen wäre prinzipiell möglich, dürfte aber durch die Faltung mit einer Gaußfunktion sehr schnell zu kaum noch handhabbaren Modellfunktionen führen.

Das folgende Kapitel stellt einen allgemeinen Ansatz zur mengenbasierten Objektsegmentierung vor.

3 Semilineare Approximation auf euklidischen Vektorräumen

Kapitel 2 beschäftigte sich mit verschiedenen Verfahren zur Bildsegmentierung, wobei der Schwerpunkt auf der Bereitstellung eines Mengenmodells zur Beschreibung der Objektform lag. Bei der Zwei-Niveau-Passung (Kapitel 2.1) basiert dieses Mengenmodell nach Voraussetzung auf dem gegebenen Bildraum. Als Folge davon ist die Genauigkeit des Mengenmodells direkt von der Auflösung dieses Bildraumes abhängig. So sind bei diskreten Bildern nur diskrete Mengenmodelle verfügbar, eine weitere Verfeinerung ist nicht möglich. Auch die Modellierung komplexerer Grauwertverläufe innerhalb eines Bildes (Abbildung 2.4) kann nicht realisiert werden. Die Grauwertmodelladaption (Kapitel 2.4) beseitigt zwar das Problem der Subpixelpassung, da sie explizit auf kontinuierlichen Mengenmodellen beruht. Eine Modellierung komplexerer Grauwertverläufe wurde aber auch hier nicht angedacht.

Im folgenden wird ein allgemeiner Ansatz zur mengenbasierten Objektsegmentierung — die semilineare Approximation — vorgestellt. Dieses Verfahren kann als theoretischer Überbau für mengenbasierte Segmentierungsalgorithmen angesehen werden und umfaßt je nach Spezifizierung auch die Modellierung inhomogener Grauwertverläufe sowie Subpixelpassungen. Zudem lassen sich Zusammenhänge zur Zwei-Niveau-Passung sowie (in Kapitel 6) zur Grauwertmodelladaption herstellen.

3.1 Semilineare Funktionen

Definition 3.1.1

Sei E ein \mathbb{R} -Vektorraum, \mathcal{A} eine Menge und $k \in \mathbb{N}$. Eine Funktion $g : \mathbb{R}^k \times \mathcal{A} \rightarrow E$ heißt dann semilinear, falls für alle $A \in \mathcal{A}$ die Abbildung $x \mapsto g(x, A)$ linear ist.

Ist g eine semilineare Funktion und $A \in \mathcal{A}$, heißt im folgenden die Abbildung

$$G_A : \mathbb{R}^k \rightarrow E, \quad G_A x := g(x, A)$$

die *lineare Teilfunktion* von g für $A \in \mathcal{A}$. Eine Reihe von Eigenschaften von g , die aus der Linearität von G_A folgen, sind in dem folgenden Korollar zusammengefaßt:

Korollar 3.1.2

Seien die Voraussetzungen von Definition 3.1.1 gegeben. Dann gilt:

- Jede semilineare Funktion g definiert eine Abbildung von \mathcal{A} in die Menge \mathcal{E}_k der höchstens k -dimensionalen Unterräume von E .
- Ist $A \in \mathcal{A}$ gegeben, existiert ein Erzeugendensystem $\{e_1^{(A)}, \dots, e_k^{(A)}\}$ von $\mathfrak{S}(G_A)$, so daß sich die lineare Teilfunktion G_A in der Form

$$G_A x = \sum_{i=1}^k x_i e_i^{(A)} \quad \text{für } x \in \mathbb{R}^k$$

schreiben läßt.

Beweis Mit $A \mapsto G_A = g(\cdot, A)$ definiert jede semilineare Funktion zunächst eine Abbildung von \mathcal{A} in die Menge $\text{Hom}(\mathbb{R}^k, E)$ der linearen Abbildungen von \mathbb{R}^k nach E . Da G_A linear ist, ist weiterhin der Bildraum $\mathfrak{S}(G_A) \subseteq E$ von G_A ein linearer Unterraum von E mit Dimension $l \leq \dim(\mathbb{R}^k) = k$. Wird nun mit \mathcal{E}_k die Menge der höchstens k -dimensionalen Unterräume von E bezeichnet, leistet die Abbildung

$$\mathfrak{G} : \mathcal{A} \rightarrow \mathcal{E}_k, \quad \mathfrak{G}(A) := \mathfrak{S}(G_A)$$

das Gewünschte.

Sei im folgenden $A \in \mathcal{A}$. Dann existiert wegen der Linearität von G_A ein Erzeugendensystem $\{e_1^{(A)}, \dots, e_k^{(A)}\}$ von $\mathfrak{S}(G_A)$, so daß sich G_A als

$$\forall x \in \mathbb{R}^k : G_A x = \sum_{i=1}^k x_i e_i^{(A)}$$

schreiben läßt. □

Man kann eine semilineare Funktion also im gewissen Sinne als zweigeteilt auffassen: Zuerst wird durch die Wahl von $A \in \mathcal{A}$ ein höchstens k -dimensionaler Unterraum von E bestimmt, in den dann die lineare Teilabbildung G_A abbildet. Da sich der so festgelegte Unterraum alleine durch die Wahl eines $x \in \mathbb{R}^k$ nicht mehr ändern kann, kann \mathcal{A} auch als eine Parametrisierung der höchstens k -dimensionalen Unterräume von E angesehen werden.

Der folgende Satz beschäftigt sich mit der Darstellung semilinearer Funktionen.

Satz 3.1.3

Unter den Voraussetzungen von Definition 3.1.1 existieren für jede semilineare Funktion g Abbildungen $e_1, \dots, e_k : \mathcal{A} \rightarrow E$, so daß sich g in der Form

$$g(x, A) = \sum_{i=1}^k x_i e_i(A)$$

schreiben läßt.

Beweis Nach Korollar 3.1.2 existiert für jedes $A \in \mathcal{A}$ ein Erzeugendensystem $\{e_1^{(A)}, \dots, e_k^{(A)}\}$ von $\mathfrak{S}(G_A)$ mit

$$\forall x \in \mathbb{R}^k : G_A x = \sum_{i=1}^k x_i e_i^{(A)}.$$

Setzt man nun für alle $i \in \{1, \dots, k\}$

$$e_i : \mathcal{A} \rightarrow E, \quad e_i(A) := e_i^{(A)}$$

erhält man das Gewünschte. □

Satz 3.1.3 ist die Basis, auf der die im folgenden auftretenden Formulierungen semilinear-er Funktionen beruhen. Statt direkt eine semilineare Funktion g zu definieren, werden zunächst k Vektoren $e_1, \dots, e_k \in E$ zusammen mit einer geeigneten Parametrisierung gesucht, die im nächsten Schritt zu *Erzeugendenfunktionen* $e_1, \dots, e_k : \mathcal{A} \rightarrow E$ ausgebaut werden. Diese Erzeugendenfunktionen beschreiben dann eine semilineare Funktion und nach Satz 3.1.3 existiert für jede semilineare Funktion eine derartige Darstellung. Wird zusätzlich mit G_A wieder die auf Seite 31 definierte lineare Teilfunktion einer semilinearen Funktion bezeichnet, gilt außerdem für den Bildraum $\mathfrak{S}(G_A)$:

$$\mathfrak{S}(G_A) = \text{linH}\{e_1(A), \dots, e_k(A)\}.$$

Das folgende Beispiel verdeutlicht, wie die semilinearen Funktionen als Verallgemeinerung der in der Zwei-Niveau-Passung verwendeten zweistufigen Funktionen aufgefaßt werden können.

Beispiel 3.1.4

Sei eine kompakte Menge $M \subseteq \mathbb{R}^2$ gegeben, die im folgenden der Aufnahme-raum genannt wird. Die Menge \mathcal{A} wird — wie bereits in der Zwei-Niveau-Passung in Kapitel 2.1 — als Quotientenraum

$$\mathcal{A} := \mathcal{B}(M)/\sim$$

definiert, wobei die Äquivalenzrelation \sim durch

$$\sim \subseteq \mathcal{A} \times \mathcal{A}, \quad A \sim B :\iff \lambda^2((A \setminus B) \cup (B \setminus A)) = 0$$

gegeben ist. λ^2 bezeichnet dabei das zweidimensionale Lebesgue-Maß.

Als Vektorraum E wird der Quotientenraum $L^2(M, \mathbb{R})$ gewählt. Setzt man dann zwei Erzeugendenfunktionen $e_1, e_2 : \mathcal{A} \rightarrow E$,

$$\begin{aligned} e_1(A) &:= \mathbb{1}_A && \text{und} \\ e_2(A) &:= \mathbb{1}_{M \setminus A}, \end{aligned}$$

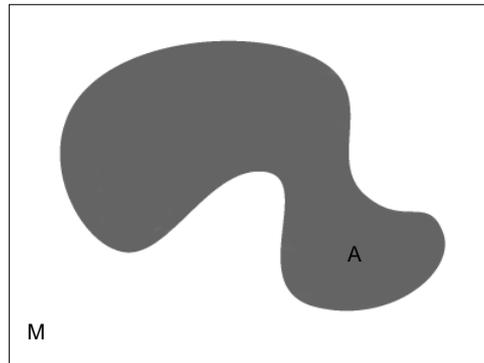


Abbildung 3.1: Skizzierter rechteckiger Aufnahmebereich M mit grauem Objekt A auf weißem Hintergrund $M \setminus A$.

ist die Funktion $g : \mathbb{R}^2 \times \mathcal{A} \rightarrow E$ mit

$$g(x, A) := x_1 e_1(A) + x_2 e_2(A) = x_1 \mathbb{1}_A + x_2 \mathbb{1}_{M \setminus A}$$

semilinear und g entspricht genau der in der Zwei-Niveau-Passung verwendeten Treppenfunktion in (2.1).

Neben g existiert aber noch eine andere Möglichkeit, eine semilineare Funktion in Analogie zur Zwei-Niveau-Passung zu definieren. Dazu wird eine neue Erzeugendenfunktion

$$e'_2(A) := \mathbb{1}_M$$

benötigt. Ersetzt man nun e_2 in g durch e'_2 , erhält man

$$g'(x, A) = x_1 \mathbb{1}_A + x_2 \mathbb{1}_M.$$

Der Unterschied in beiden Funktionen besteht nur in ihrer Interpretation: In g repräsentiert x_1 den Grauwert eines Objektes A , während x_1 in g' für die Grauwertänderung zwischen Objekt und Hintergrund steht.

Bemerkung 3.1.5

Die Formulierung der semilinearen Funktion g in Beispiel 3.1.4 ist auch möglich, ohne daß für \mathcal{A} und E Quotientenräume gewählt werden. An dieser Stelle der Arbeit existiert noch kein Grund, warum der Quotientenraum $\mathcal{A} = \mathcal{B}(M)/\sim$ der Menge $\mathcal{B}(M)$ vorzuziehen wäre. Ebenfalls kein Grund existiert, warum $L^2(M, \mathbb{R})$ eine bessere Wahl als $\mathcal{L}^2(M, \mathbb{R})$ sein sollte und natürlich sind auch

$$\tilde{g} : \mathbb{R}^2 \times \mathcal{B}(M) \rightarrow \mathcal{L}^2(M, \mathbb{R}), \quad \tilde{g}(x, A) = x_1 \mathbb{1}_A + x_2 \mathbb{1}_{M \setminus A}$$

bzw.

$$\tilde{g}' : \mathbb{R}^2 \times \mathcal{B}(M) \rightarrow \mathcal{L}^2(M, \mathbb{R}), \quad \tilde{g}'(x, A) = x_1 \mathbb{1}_A + x_2 \mathbb{1}_M$$

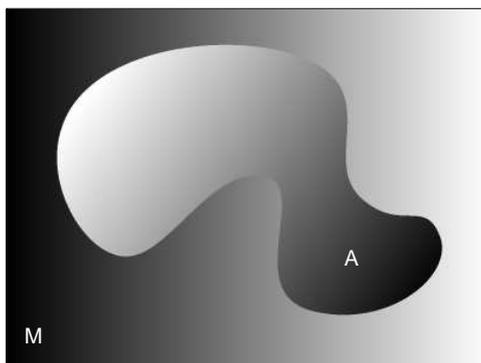


Abbildung 3.2: In x -Richtung inhomogener Aufnahmeraum M mit ebenfalls in x -Richtung inhomogenem Objekt A .

semilineare Funktionen. Gründe, warum diese Wahl nicht geschickt ist (Metrik auf \mathcal{A} bzw. Skalarprodukt auf E würden benötigt) sind zwar — zum Teil eher implizit — bereits in Kapitel 2 erwähnt worden, werden hier aber erst in Kapitel 3.2 und 3.3 für die Formulierung des Approximationsproblems bzw. für Aussagen über die Existenz lokaler Optima benötigt. Die Wahl von $\mathcal{B}(M)/\sim$ und $L^2(M, \mathbb{R})$ stellt in diesem Sinne einen Vorgriff auf die folgenden Abschnitte dar.

Abbildung 3.1 zeigt beispielhaft einen rechteckigen Aufnahmeraum M mit einem grauen Objekt A auf hellem Hintergrund $M \setminus A$. Will man Objekt und Hintergrund aus dieser Abbildung durch eine semilineare Funktion beschreiben, sind die in Beispiel 3.1.4 formulierten Erzeugendenfunktionen sowie die daraus resultierenden semilinearen Funktionen g bzw. g' sehr gut geeignet. Man ist bei der Formulierung semilinearere Funktionen aber weder in der Anzahl k der Erzeugendenfunktionen e_1, \dots, e_k noch in deren Form beschränkt, so daß sich auch kompliziertere Zusammenhänge formulieren lassen. Das folgende Beispiel zeigt, wie die in Abbildung 3.2 gezeigten Grauwertverläufe durch eine semilineare Funktion modelliert werden können.

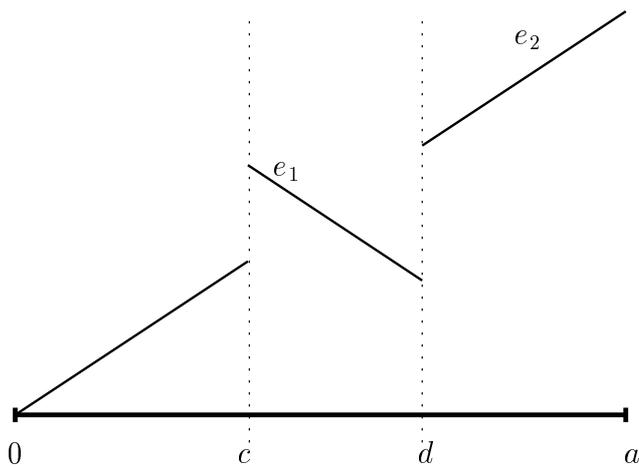
Beispiel 3.1.6

Abbildung 3.2 zeigt das gleiche Objekt wie Abbildung 3.1. Allerdings weisen nun sowohl Objekt als auch Hintergrund eine (in x -Richtung) monotone Änderung des Grauwertes auf, was zum Beispiel als inhomogene Beleuchtungssituation interpretiert werden kann. Zum einfacheren Verständnis wird die Modellierung dieser Situation im folgenden als eindimensionale Version für eine isolierte Bildzeile durchgeführt.

Sei also für $a > 0$ der Aufnahmeraum $M = [0, a]$ gegeben. \mathcal{A} wird wieder (wie in Beispiel 3.1.4) als Quotientenraum $\mathcal{A} = \mathcal{B}(M)/\sim$ definiert, wobei die Äquivalenzrelation $\sim \subseteq \mathcal{B}(M) \times \mathcal{B}(M)$ durch

$$A \sim B : \iff \lambda((A \setminus B) \cup (B \setminus A)) = 0$$

definiert ist. Der Vektorraum E sei die Menge $L^2(M, \mathbb{R})$.



Objekt:

$$e_1([c, d])(x) = \begin{cases} a - x, & x \in [c, d] \\ 0 & x \notin [c, d] \end{cases}$$

Hintergrund:

$$e_2([c, d])(x) = \begin{cases} 0, & x \in [c, d] \\ x, & x \notin [c, d] \end{cases}$$

Abbildung 3.3: Modellierung linearer Beleuchtung ohne konstanten Anteil.

Bis jetzt unterscheidet sich diese Modellierung nicht wesentlich gegenüber der aus Beispiel 3.1.4. Die einzige große Änderung erfolgt nun in der Wahl der Erzeugendenfunktionen e_i . Geht man davon aus, daß es sich bei Abbildung 3.2 um lineare Änderungen der Grauwerte handelt, eignet sich die Funktion

$$e_1 : \mathcal{A} \rightarrow L^2(M, \mathbb{R}), \quad e_1(A) := (a - \text{id})\mathbb{1}_A,$$

um das Objekt und die Funktion

$$e_2 : \mathcal{A} \rightarrow L^2(M, \mathbb{R}), \quad e_2(A) := \text{id} \cdot \mathbb{1}_{M \setminus A}$$

um den Hintergrund zu beschreiben. Abbildung 3.3 zeigt e_1 und e_2 , wobei das Objekt durch das Intervall $[c, d] \in \mathcal{A}$ und der Hintergrund durch die Restmenge $[0, c] \cup [d, a] \in \mathcal{A}$ repräsentiert wird. Die scheinbare Doppeldefinition in den Punkten $\{c\}$ und $\{d\}$ ist in Wirklichkeit kein Problem, da die Elemente in \mathcal{A} nur bis auf λ -Nullmengen eindeutig sind.

Die so definierte semilineare Funktion

$$g_1 : \mathbb{R}^2 \times \mathcal{B}(M)/\sim \rightarrow L^2(M, \mathbb{R}), \quad g_1(x, A) := x_1(a - \text{id})\mathbb{1}_A + x_2 \text{id} \cdot \mathbb{1}_{M \setminus A}$$

ist in den Randpunkten 0 und a noch sehr speziell, denn der Grauwert des so erzeugten Bildes ist beide Male auf 0 festgelegt. Eine allgemeinere Modellierung ohne feste Randpunkte erhält man, wenn die beiden „konstanten“ Erzeugendenfunktionen

$$\begin{aligned} e_3(A) &:= \mathbb{1}_A && \text{für das Objekt und} \\ e_4(A) &:= \mathbb{1}_{M \setminus A} && \text{für den Hintergrund} \end{aligned}$$

zur Modellierung hinzugefügt werden. Damit ändert die resultierende semilineare Funktion ihren Definitionsbereich. Sie hat nun die Form $g_2 : \mathbb{R}^4 \times \mathcal{B}(M)/\sim \rightarrow L^2(M, \mathbb{R})$,

$$g_2(x, A) := \underbrace{x_1(a - \text{id}_A) + x_3 \mathbb{1}_A}_{\text{Objekt}} + \underbrace{x_2 \text{id}_{M \setminus A} + x_4 \mathbb{1}_{M \setminus A}}_{\text{Hintergrund}}$$

Durch g_2 werden die in Abbildung 3.2 vorhandenen linearen Abhängigkeiten modelliert ohne daß zusätzliche Annahmen (wie Randpunktfestlegungen) getroffen werden müssen.

In Beispiel 3.1.6 wurden die Grauwertänderungen in Objekt und Hintergrund durch lineare Abhängigkeiten modelliert, da Abbildung 3.2 künstlich erzeugt wurde und daher bereits a-priori die bestmögliche Art der Modellierung bekannt war. In praktischen Situationen ist ein solches Vorwissen eher weniger vorhanden, die konkrete Art der Modellierung sollte daher bei jeder Anwendung sorgfältig abgewogen werden. Dies setzt eine möglichst genaue Analyse des vorliegenden Problems als zentralen Schritt vor die eigentliche Modellierung. Vernachlässigt man diese Analyse, kann die Anzahl der Erzeugendenfunktionen und damit die maximale Dimension k der auftretenden Unterräume sehr schnell zu groß werden. Die Modellierung wirkt dann umständlich und ist für die im folgenden beschriebene Approximation zu langsam. Eine Einschränkung auf nur wenige Erzeugendenfunktionen ist allerdings auch nicht immer die Lösung. Werden zu wenige oder zu spezielle Erzeugendenfunktionen ausgewählt, kann die vorliegende Situation meist nicht adäquat beschrieben werden. In diesen Kontext fällt die Funktion g_1 aus Beispiel 3.1.6, die in ihren Randpunkten festgelegt war ohne daß dazu ein Hinweis in der Ausgangssituation gegeben war. Auch Abbildung 2.4 von Seite 21 zeigt ein Beispiel für eine zu spezielle Wahl der Erzeugendenfunktionen. Hier wurde die im Ausgangsbild vorhandene lineare Grauwertänderung durch die „Erzeugendenfunktionen“ $\mathbb{1}_A$ und $\mathbb{1}_{M \setminus A}$ der Zwei-Niveau-Passung nicht nachgebildet, so daß der Paßprozeß an sich in einem unbefriedigenden Optimum endete.

3.2 Semilineare Approximation

Bis jetzt wurden semilineare Funktionen an sich betrachtet. Die Beispiele in 3.1.4 und 3.1.6 gaben dabei einen kleinen Einblick in die Möglichkeiten der Modellierung unterschiedlicher Ausgangssituationen. Der nächste Schritt ist der Vergleich eines gegebenen Elements $y \in E$ mit einem durch eine semilineare Funktion entstandenem Element $g(x, A) = \sum_{i=1}^k x_i e_i(A) \in E$ und natürlich die Frage, für welches Paar $(x, A) \in \mathbb{R}^k \times \mathcal{A}$ die Funktion $g(x, A)$ dem gegebenen $y \in E$ „am nächsten“ liegt. Diese Parameteroptimierung wird in der Bildverarbeitung als *Bildrekonstruktion* angesehen, wobei das rekonstruierte Bild hier durch die jeweilige semilineare Funktion bestimmt wird. Der Terminus „am nächsten“ erfordert einen Abstandsbegriff, so daß E in der folgenden Definition als euklidischer Vektorraum vorausgesetzt wird.

Definition 3.2.1

Sei E ein euklidischer Vektorraum, (\mathcal{A}, d) ein metrischer Raum und $k > 0$. Sei ferner $y \in E$ gegeben. Für jede semilineare Funktion $g : \mathbb{R}^k \times \mathcal{A} \rightarrow E$ heißt dann das Minimierungsproblem

$$\inf_{A \in \mathcal{A}} \inf_{x \in \mathbb{R}^k} \|y - g(x, A)\|$$

ein semilineares Approximationsproblem.

Eine direkte Folgerung, was die Darstellung semilinearer Approximationsprobleme betrifft, ergibt sich aus Satz 3.1.3: Da jede semilineare Funktion g als Linearkombination von Erzeugendenfunktionen $e_1, \dots, e_k : \mathcal{A} \rightarrow E$ dargestellt werden kann, läßt sich auch jedes semilineare Approximationsproblem in der Form

$$\inf_{A \in \mathcal{A}} \inf_{x \in \mathbb{R}^k} \left\| y - \sum_{i=1}^k x_i e_i(A) \right\|$$

schreiben.

Bemerkung 3.2.2

In Definition 3.2.1 ist E als euklidischer Vektorraum vorausgesetzt, was bedeutet, daß auf E ein Skalarprodukt $\langle \cdot | \cdot \rangle$ definiert ist und die Norm $\|\cdot\|$ auf E durch das Skalarprodukt induziert wird. In Definition 3.2.1 wird das Skalarprodukt selbst jedoch nicht benötigt. Prinzipiell ist Definition 3.2.1 also auch denkbar, wenn die Norm in E nicht über ein Skalarprodukt definiert wird. Im Rahmen dieser Arbeit war jedoch keine Notwendigkeit vorhanden, E von einem euklidischen Vektorraum auf einen allgemeinen normierten Raum zu verallgemeinern, so daß dieser allgemeinere Ansatz hier nicht weiter verfolgt wird.

Die Tatsache, daß jede semilineare Funktion einen linearen Anteil besitzt, läßt sich für die Approximation sehr gut nutzen. Das folgende Korollar zeigt den dabei wesentlichen Zusammenhang auf:

Korollar 3.2.3

Seien die Voraussetzungen von Definition 3.2.1 gegeben und sei $A_0 \in \mathcal{A}$ fest. Dann existiert ein eindeutig bestimmter Vektor $y_{A_0} \in \text{linH}\{e_1(A_0), \dots, e_k(A_0)\}$, für den gilt:

$$\inf_{x \in \mathbb{R}^k} \|y - g(x, A_0)\| = \min_{x \in \mathbb{R}^k} \|y - g(x, A_0)\| = \|y - y_{A_0}\|$$

Beweis Sei $G_{A_0} : \mathbb{R}^k \rightarrow E$ für festes $A_0 \in \mathcal{A}$ die auf Seite 31 definierte lineare Teilabbildung einer semilinearen Funktion g . Dann beschreibt das Teilproblem

$$\inf_{x \in \mathbb{R}^k} \|y - G_{A_0} x\|$$

ein lineares Ausgleichsproblem. Für diese Problemklasse existiert in euklidischen Vektorräumen stets ein eindeutig bestimmter Vektor $y_{A_0} \in \mathfrak{S}(G_{A_0})$ mit

$$\|y - y_{A_0}\| = \min_{x \in \mathbb{R}^k} \|y - G_{A_0} x\| = \text{dist}(y, \mathfrak{S}(G_{A_0})).$$

In diesen Fällen ist y_{A_0} ein Punkt minimaler Distanz von y an den Bildraum $\mathfrak{S}(G_{A_0}) = \text{linH}\{e_1(A_0), \dots, e_k(A_0)\}$:

$$y_{A_0} = \text{proj}_{\text{linH}\{e_1(A_0), \dots, e_k(A_0)\}} y.$$

□

```

1: Gegeben sei ein semilineares Approximationsproblem gemäß Def. 3.2.1
2: choose  $A_0 \in \mathcal{A}$ 
3: compute  $x_0 \in \mathbb{R}^k$  with  $\|y - G_{A_0}x_0\| = \min_{\xi \in \mathbb{R}^k} \|y - G_{A_0}\xi\|$ 
4:  $i := 0$ 
5: repeat
6:   choose  $A_{i+1} \in \mathcal{A}$ 
7:   find  $x_{i+1} \in \mathbb{R}^k$  with  $\|y - G_{A_{i+1}}x_{i+1}\| = \min_{\xi \in \mathbb{R}^k} \|y - G_{A_{i+1}}\xi\|$ 
8:   if  $\|y - G_{A_{i+1}}x_{i+1}\| < \|y - G_{A_i}x_i\|$  then
9:      $i := i + 1$ 
10: until (Abbruchkriterium erreicht)

```

Algorithmus 1: Lokale Minimierung eines semilinearen Approximationsproblems. Als Abbruchkriterium wird in der Regel $\|y - G_{A_{i+1}}x_{i+1}\| - \|y - G_{A_i}x_i\| < \varepsilon$ für $\varepsilon > 0$ verwendet.

Korollar 3.2.3 zeigt, daß zumindest für das lineare Teilproblem eines semilinearen Approximationsproblems sofort die Existenz eines Minimums garantiert werden kann. Da dieses Teilproblem ein lineares Ausgleichsproblem beschreibt, läßt sich darüber hinaus dieses Minimum auch leicht berechnen (siehe dazu zum Beispiel [LH79]). Algorithmus 1 zeigt schematisch, wie diese Tatsache in einer Implementierung aussehen könnte. Die Zeilen 3: und 7: beziehen sich dabei auf die Lösung dieses linearen Ausgleichsproblems. Die Minimierungsschleife von Zeile 5: bis Zeile 10: wird üblicherweise von einem nicht-linearen Abstiegsalgorithmus übernommen.

Durch Korollar 3.2.3 kann ein semilineares Approximationsproblem sehr viel einfacher formuliert werden, wie der folgende Satz zeigt:

Satz 3.2.4

Unter den Voraussetzungen von Definition 3.2.1 besitzt jedes semilineare Approximationsproblem die Form

$$\inf_{A \in \mathcal{A}} \|y - \text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}} y\|.$$

Beweis Der Beweis folgt direkt aus Korollar 3.2.3: Jedes semilineare Approximationsproblem läßt sich schreiben als

$$\inf_{A \in \mathcal{A}} \min_{x \in \mathbb{R}^k} \|y - g(x, A)\| = \inf_{A \in \mathcal{A}} \|y - y_A\|,$$

mit $y_A = \text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}} y$. □

Satz 3.2.4 führt endgültig zu einer neuen Betrachtungsweise semilinearer Approximationen, die bereits in Algorithmus 1 angedeutet wurde: Statt des ursprünglichen Problems von Definition 3.2.1 genügt es nun, die Funktion

$$\psi_{g,y} : \mathcal{A} \rightarrow \mathbb{R}, \quad \psi_{g,y}(A) := \|y - \text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}} y\| \tag{3.1}$$

zu minimieren¹. Der in Definition 3.2.1 zusätzlich vorhandene Parameter $x \in \mathbb{R}^k$ im Minimierungsprozeß fällt weg, übrig bleibt die alleinige Optimierung über \mathcal{A} .

Lemma 3.2.5

Die in (3.1) definierte Funktion $\psi_{g,y}$ ist durch 0 nach unten und $\|y\|$ nach oben beschränkt.

Beweis Offensichtlich ist $\psi_{g,y}$ durch 0 nach unten beschränkt. Weiterhin gilt der Zusammenhang

$$\psi_{g,y}^2(A) = \|y\|^2 - \|\text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}} y\|^2$$

und wegen $\|y\| \geq \|\text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}} y\|$ ist $\psi_{g,y}$ durch $\|y\|$ nach oben beschränkt. \square

Algorithmus 1 und Satz 3.2.4 lassen eine Frage außen vor, nämlich die ob überhaupt ein lokales Minimum existiert. Dieser Frage wird im Folgenden nachgegangen.

3.3 Existenz lokaler Optima

Um Aussagen über die Existenz lokaler Optima von $\psi_{g,y}$ treffen zu können, muß zuerst die Stetigkeit von g sichergestellt werden. Darüber lassen sich aber relativ einfach Aussagen treffen, wie der folgende Satz zeigt:

Satz 3.3.1

Sei ein semilineares Approximationsproblem nach Definition 3.2.1 gegeben. Läßt sich dann die semilineare Funktion $g : \mathbb{R}^k \times \mathcal{A} \rightarrow E$ durch stetige Erzeugendenfunktionen e_1, \dots, e_k darstellen, ist auch g stetig auf dem Produktraum $\mathbb{R}^k \times \mathcal{A}$.

Beweis Bezeichnet $d_{\mathcal{A}}$ die Metrik auf \mathcal{A} und wählt man auf \mathbb{R}^k zum Beispiel die euklidische Metrik d , bildet $\mathbb{R}^k \times \mathcal{A}$ zusammen mit der durch $\mathfrak{d} : (\mathbb{R}^k \times \mathcal{A}) \times (\mathbb{R}^k \times \mathcal{A}) \rightarrow \mathbb{R}$,

$$\mathfrak{d}((x_1, A_1), (x_2, A_2)) := \max(d(x_1, x_2), d_{\mathcal{A}}(A_1, A_2))$$

definierten Produktmetrik einen metrischen Raum. Die Projektionsabbildungen $\pi_{\mathbb{R}^k} : \mathbb{R}^k \times \mathcal{A} \rightarrow \mathbb{R}^k$ mit $\pi_{\mathbb{R}^k}(x, A) := x$ und $\pi_{\mathcal{A}} : \mathbb{R}^k \times \mathcal{A} \rightarrow \mathcal{A}$ mit $\pi_{\mathcal{A}}(x, A) := A$ sind dann stetig bzgl. \mathfrak{d} . Ebenso ist auch die Funktion $f : \mathbb{R}^k \times E^k \rightarrow E$ mit

$$f(x_1, \dots, x_k, y_1, \dots, y_k) := \sum_{i=1}^k x_i y_i$$

stetig bzgl. der entsprechenden Produktmetrik.

¹Der Index g in $\psi_{g,y}$ deutet an, daß sich $\psi_{g,y}$ auf eine semilineare Funktion bezieht, der Index y steht für das zu approximierende Element $y \in E$

Wird mit $\pi_i : \mathbb{R}^k \rightarrow \mathbb{R}$ die Projektion von \mathbb{R}^k auf die i -te Koordinate von \mathbb{R}^k bezeichnet und g als

$$\begin{aligned} g &= f \circ (\pi_1 \circ \pi_{\mathbb{R}^k}, \dots, \pi_k \circ \pi_{\mathbb{R}^k}, e_1 \circ \pi_{\mathcal{A}}, \dots, e_k \circ \pi_{\mathcal{A}}) \\ &= \sum_{i=1}^k (\pi_i \circ \pi_{\mathbb{R}^k}) \circ (e_i \circ \pi_{\mathcal{A}}) \end{aligned}$$

dargestellt, ist g als Verknüpfung stetiger Funktionen selbst stetig. \square

Wählt man also unter den Voraussetzungen von Satz 3.3.1 stetige Erzeugendenfunktionen $e_1, \dots, e_k : \mathcal{A} \rightarrow E$, ist die aus ihnen gebildete semilineare Funktion g ebenfalls stetig. Wegen der Stetigkeit von Skalarprodukten ist dann auch die Abbildung

$$(x, A) \mapsto \left\| y - \sum_{i=1}^k x_i e_i(A) \right\| = \|y - g(x, A)\|$$

stetig.

Bemerkung 3.3.2

Betrachtet man an dieser Stelle noch einmal die Beispiele 3.1.4 und 3.1.6, wird deutlich, was in Bemerkung 3.1.5 bereits angedeutet wurde. Durch die Festlegung von \mathcal{A} als Quotientenraum $\mathcal{B}(M)/\sim$ hat man Zugriff auf die durch

$$d(A, B) := \lambda^2((A \setminus B) \cup (B \setminus A))$$

auf $\mathcal{B}(M)/\sim$ definierte Metrik. Die Notwendigkeit dafür zeigte sich bereits in Kapitel 2.1. Ähnlich liegt der Fall bei der Wahl von $L^2(M, \mathbb{R})$ statt $\mathcal{L}^2(M, \mathbb{R})$ als Vektorraum E , auch wenn dies in Kapitel 2.1 nur implizit zur Sprache kam: Nach Definition 3.2.1 muß E ein euklidischer Vektorraum sein. Die Abbildung

$$\langle \cdot | \cdot \rangle : E^2 \rightarrow \mathbb{R}, \quad \langle f | g \rangle := \int_M f \cdot g$$

bildet aber nur dann ein Skalarprodukt, wenn sie auf $L^2(M, \mathbb{R})$ und nicht auf $\mathcal{L}^2(M, \mathbb{R})$ definiert ist.

Die in den Beispielen 3.1.4 und 3.1.6 definierten Erzeugendenfunktionen — und damit auch die daraus resultierenden semilinearen Funktionen — sind mit diesen Überlegungen stetig.

Die Stetigkeit von $(x, A) \mapsto \|y - \sum_{i=1}^k x_i e_i(A)\|$ impliziert allerdings nicht die Stetigkeit der in (3.1) von Seite 39 definierten Funktion

$$\psi_{g,y}(A) = \|y - \text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}} y\|.$$

Da dies die eigentliche Zielfunktion eines semilinearen Approximationsproblems ist, sind vor allem Aussagen über ihr Verhalten von großer Bedeutung. Von Interesse ist dabei vor allem, ob und wenn ja unter welchen Bedingungen sich $\psi_{g,y}$ stetig verhält und natürlich ob sich mögliche Unstetigkeitsstellen eingrenzen lassen.

Betrachtet man zunächst den Fall $k = 2$, stellt sich die Situation wie folgt dar: Jede stetige semilineare Funktion $g : \mathbb{R}^2 \times \mathcal{A} \rightarrow E$ hat die Form

$$g(x, A) = x_1 e_1(A) + x_2 e_2(A),$$

wobei $e_1, e_2 : \mathcal{A} \rightarrow E$ stetige Erzeugendenfunktionen bezeichnen. Ist weiterhin ein Element $y \in E$ gegeben, stellt sich das entsprechende semilineare Approximationsproblem als

$$\inf_{A \in \mathcal{A}} \inf_{x \in \mathbb{R}^2} \|y - x_1 e_1(A) + x_2 e_2(A)\|,$$

bzw. mit Satz 3.2.4 als

$$\inf_{A \in \mathcal{A}} \psi_{g,y}(A) = \inf_{A \in \mathcal{A}} \|y - \text{proj}_{\text{linH}\{e_1(A), e_2(A)\}} y\|$$

dar.

Sei nun $\mathcal{A}_2 \subseteq \mathcal{A}$ so gewählt, daß \mathcal{A}_2 alle Elemente enthält, für die der Bildraum $\text{linH}\{e_1(A), e_2(A)\}$ volle Dimension besitzt:

$$\forall A \in \mathcal{A} : A \in \mathcal{A}_2 \Leftrightarrow \dim(\text{linH}\{e_1(A), e_2(A)\}) = 2.$$

Für jedes $A \in \mathcal{A}_2$ läßt sich dann eine Orthonormalbasis $\{e'_1(A), e'_2(A)\}$ des Raumes $\text{linH}\{e_1(A), e_2(A)\}$ konstruieren, mit deren Hilfe sich $\psi_{g,y}$ als

$$\psi_{g,y}(A) = \|y - \langle y | e'_1(A) \rangle e'_1(A) - \langle y | e'_2(A) \rangle e'_2(A)\|$$

darstellen läßt. Daraus folgt direkt, daß $\psi_{g,y}$ stetig auf \mathcal{A}_2 ist, denn für alle $A \in \mathcal{A}_2$ lassen sich die Vektoren $e'_1(A)$ und $e'_2(A)$ über stetige Transformationen aus $e_1(A)$ und $e_2(A)$ konstruieren (Schmidtsches Orthogonalisierungsverfahren).

Für $A \in \mathcal{A} \setminus \mathcal{A}_2$ reduziert sich die Dimension des Bildraumes $\text{linH}\{e_1(A), e_2(A)\}$ auf

$$\forall A \in \mathcal{A} \setminus \mathcal{A}_2 : \dim(\text{linH}\{e_1(A), e_2(A)\}) \leq 1.$$

Für $A \in \mathcal{A} \setminus \mathcal{A}_2$ mit $e_1(A) \neq 0$ gilt

$$\psi_{g,y}(A) = \|y - \text{proj}_{\text{linH}\{e_1(A)\}} y\| = \left\| y - \frac{\langle y | e_1(A) \rangle}{\|e_1(A)\|^2} e_1(A) \right\|.$$

und für $A \in \mathcal{A} \setminus \mathcal{A}_2$ mit $e_1(A) = 0$ entsprechend

$$\psi_{g,y}(A) = \begin{cases} \left\| y - \frac{\langle y | e_2(A) \rangle}{\|e_2(A)\|^2} e_2(A) \right\|, & \text{falls } e_2(A) \neq 0 \\ 0 & \text{sonst.} \end{cases}$$

Man kann nun erkennen, daß die Stetigkeit der Funktion $\psi_{g,y}$ nur von der Stetigkeit ihrer Teilabbildung $A \mapsto \text{proj}_{\text{linH}\{e_1(A), e_2(A)\}} y$ abhängt. Insgesamt stellt sich diese Projektionsabbildung als stückweise definierte Funktion wie folgt dar:

$$A \mapsto \text{proj}_{\text{linH}\{e_1(A), e_2(A)\}} y = \begin{cases} \langle y | e'_1(A) \rangle e'_1(A) \\ + \langle y | e'_2(A) \rangle e'_2(A), & \text{falls } A \in \mathcal{A}_2 \\ \frac{\langle y | e_1(A) \rangle}{\|e_1(A)\|^2} e_1(A), & \text{falls } A \in \mathcal{A} \setminus \mathcal{A}_2 \text{ und } e_1(A) \neq 0 \\ \frac{\langle y | e_2(A) \rangle}{\|e_2(A)\|^2} e_2(A), & \text{falls } A \in \mathcal{A} \setminus \mathcal{A}_2, e_1(A) = 0 \\ & \text{und } e_2(A) \neq 0 \\ 0 & \text{sonst} \end{cases}$$

$e'_1(A)$ und $e'_2(A)$ werden dabei aus $e_1(A)$ und $e_2(A)$ über das Schmidtsche Orthogonalisierungsverfahren bestimmt. Da die einzelnen Teilabbildungen von $A \mapsto \text{proj}_{\text{linH}\{e_1(A), e_2(A)\}} y$ stetig sind, bleiben nur die Übergänge zwischen den einzelnen Teilabbildungen als mögliche Unstetigkeitsstellen übrig.

Bemerkung 3.3.3

Der Fall $k = 2$ läßt den Zusammenhang zur Zwei-Niveau-Passung besonders deutlich werden. Gilt nämlich $\langle e_1(A) | e_2(A) \rangle = 0$ für alle $A \in \mathcal{A}_2$, ist für $A \in \mathcal{A}_2$

$$\psi_{g,y}(A) = \left\| y - \frac{\langle y | e_1(A) \rangle}{\|e_1(A)\|^2} e_1(A) - \frac{\langle y | e_2(A) \rangle}{\|e_2(A)\|^2} e_2(A) \right\|.$$

In diesem Spezialfall hat $\psi_{g,y}$ einen engen Bezug zu ψ aus Satz 2.1.2. Zunächst läßt sich $\psi_{g,y}^2$ umformen zu

$$\begin{aligned} \psi_{g,y}^2(A) &= \left\| y - \frac{\langle y | e_1(A) \rangle}{\|e_1(A)\|^2} e_1(A) - \frac{\langle y | e_2(A) \rangle}{\|e_2(A)\|^2} e_2(A) \right\|^2 \\ &= \|y\|^2 - 2 \left\langle y \left| \frac{\langle y | e_1(A) \rangle}{\|e_1(A)\|} e_1(A) + \frac{\langle y | e_2(A) \rangle}{\|e_2(A)\|} e_2(A) \right\rangle \\ &\quad + \left\| \frac{\langle y | e_1(A) \rangle}{\|e_1(A)\|} e_1(A) + \frac{\langle y | e_2(A) \rangle}{\|e_2(A)\|} e_2(A) \right\|^2 \\ &= \|y\|^2 - \frac{\langle y | e_1(A) \rangle^2}{\|e_1(A)\|} - \frac{\langle y | e_2(A) \rangle^2}{\|e_2(A)\|}. \end{aligned}$$

Im Fall der Zwei-Niveau-Passung von Kapitel 2.1 mit $M \subseteq \mathbb{R}^2$ wird als euklidischer Vektorraum E der Raum $L^2(M, \mathbb{R})$ zusammen mit dem (nicht explizit definierten) Integralskalarprodukt $\langle f | g \rangle := \int f \cdot g$ verwendet. Die Erzeugendenfunktionen ergeben sich zu $e_1(A) = \mathbb{1}_A$ und $e_2(A) = \mathbb{1}_{M \setminus A}$ und daher gilt der folgende Zusammenhang:

$$\psi_{g,y}^2(A) = \int_M y - \frac{\left(\int_A y\right)^2}{\lambda^2(A)} - \frac{\left(\int_{M \setminus A} y\right)^2}{\lambda^2(M \setminus A)} = \int_M y^2 - \psi(A)$$

wobei λ^2 das zweidimensionale Lebesgue-Maß und ψ die in Satz 2.1.2 definierte Funktion bezeichnet.

Für beliebiges k stellt sich die Projektionsabbildung analog zum Fall $k = 2$ dar. Hier existiert für jeden Unterraum, der durch die Erzeugendenfunktionen e_1, \dots, e_k erzeugt werden kann, eine eigene Projektions-Teilabbildung. Entsprechend häufiger können dann auch potentielle Unstetigkeitsstellen auftreten.

Eine einfache Möglichkeit, die Stetigkeit von $\psi_{g,y}$ zu garantieren wird in [Han01] verwendet und besteht darin, \mathcal{A} so weit einzuschränken, daß keine potentiellen Unstetigkeitsstellen mehr vorhanden sind. Dies ist Thema des folgenden Korollars:

Korollar 3.3.4

Seien die Voraussetzungen von Definition 3.2.1 gegeben und seien zusätzlich die Erzeugendenfunktionen $e_i : \mathcal{A} \rightarrow E$ stetig für alle $i = 1, \dots, k$. Sei $\mathcal{A}' \subseteq \mathcal{A}$. Gilt dann

$$\forall A' \in \mathcal{A}' : \dim(\text{linH}\{e_1(A'), \dots, e_k(A')\}) = k,$$

ist $\psi_{g,y}$ stetig auf \mathcal{A}' für alle $y \in E$. Ist \mathcal{A}' zusätzlich kompakt, existiert ein (sogar globales) Minimum von $\psi_{g,y}$ über \mathcal{A}' .

Beweis Wird mit G_A wieder die auf Seite 31 definierte lineare Teilabbildung von g bezeichnet, gilt: Für alle $A' \in \mathcal{A}'$ bildet nach Voraussetzung die Menge $\{e_1(A'), \dots, e_k(A')\}$ ein minimales Erzeugendensystem des k -dimensionalen Raumes $\mathfrak{S}(G_A)$. Aus den Vektoren $e_1(A'), \dots, e_k(A')$ kann daher eine Orthonormalbasis $\{e'_1(A'), \dots, e'_k(A')\}$ des Raumes $\text{linH}\{e_1(A'), \dots, e_k(A')\}$ konstruiert werden. Da dies für alle $A' \in \mathcal{A}'$ möglich ist, sind auch die entsprechenden Abbildungen $e'_1, \dots, e'_k : \mathcal{A}' \rightarrow E$, wohldefiniert. Sie sind weiterhin stetig, da die e_i bereits nach Voraussetzung stetig sind und das Schmidtsche ON-Verfahren die Stetigkeit nicht beeinträchtigt. Damit ist für jedes $y \in E$ auch

$$\psi_{g,y}(A') = \left\| y - \sum_{i=1}^k \langle y | e'_i(A') \rangle e'_i(A') \right\|$$

stetig. Ist \mathcal{A}' zusätzlich kompakt, existiert ein (sogar globales) Minimum von $\psi_{g,y}$ über \mathcal{A}' , da stetige Funktionen über kompakten Definitionsmengen ihre Extrempunkte annehmen. □

Eine zweite einfache Möglichkeit zur Sicherstellung der Stetigkeit von $\psi_{g,y}$ bietet das folgende Korollar 3.3.6, für das zunächst eine Hilfsdefinition benötigt wird.

Definition 3.3.5 (Hilfsdefinition)

Sei E ein \mathbb{R} -Vektorraum und $U \subseteq E$ ein Unterraum von E sowie $n \in \mathbb{N}$. Dann heißt die Menge $\{e_1, \dots, e_n\} \subseteq E$ eine erweiterte Orthonormalbasis von U , falls

- $\text{linH}\{e_1, \dots, e_n\} = U$, sowie
- eine Permutation $\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\}$ und ein $n' \leq n$ existiert so daß $\{e_{\pi(1)}, \dots, e_{\pi(n')}\}$ eine Orthonormalbasis von U bildet und $e_{\pi(i)} = 0$ für alle $i > n'$ ist.

Korollar 3.3.6

Seien die Voraussetzungen von Definition 3.2.1 gegeben und sei $\mathcal{A}' \subseteq \mathcal{A}$. Existieren dann stetige Funktionen $e'_1, \dots, e'_k : \mathcal{A}' \rightarrow E$ derart daß für alle $A' \in \mathcal{A}'$ die Menge $\{e'_1(A'), \dots, e'_k(A')\}$ eine erweiterte Orthonormalbasis von $\text{linH}\{e_1(A'), \dots, e_k(A')\}$ bildet, ist auch $\psi_{g,y}$ stetig auf \mathcal{A}' für alle $y \in E$.

Beweis Da E ein euklidischer Vektorraum ist und $\{e'_1(A'), \dots, e'_k(A')\}$ nach Voraussetzung eine erweiterte Orthonormalbasis bildet, läßt sich für alle $y \in E$ die Teilabbildung $A' \mapsto \text{proj}_{\text{linH}\{e_1(A'), \dots, e_k(A')\}} y$ von $\psi_{g,y}$ durch

$$\text{proj}_{\text{linH}\{e_1(A'), \dots, e_k(A')\}} y = \sum_{i=1}^k \langle y | e'_i(A') \rangle e'_i(A')$$

darstellen. Diese Abbildung ist stetig, da bereits die e'_i nach Voraussetzung stetig auf \mathcal{A}' sind. Damit ist auch $\psi_{g,y}$ stetig auf \mathcal{A}' . \square

Sowohl Korollar 3.3.4 als auch Korollar 3.3.6 basieren auf der Stetigkeit der Erzeugendenfunktionen e_i bzw. ihrer normalisierten Versionen. Viele Anwendungsfälle können durch diese beiden Korollare abgedeckt werden; so wäre eine einfache Maßnahme zur Anwendung von Korollar 3.3.4 im konkreten Fall die Einschränkung von \mathcal{A} auf einen geeigneten kompakten Ausschnitt. Allerdings kann eine derart radikale Vorgehensweise auch nachteilig sein, denn gleichzeitig muß sichergestellt werden, daß das zu erwartende Optimum auch innerhalb dieses eingeschränkten Bereiches zu finden ist. Ist dies nicht der Fall, erhält man Ergebnisse, die zwar optimal bzgl. der Zielfunktion sind die aber dennoch „unerwünscht“ sind in dem Sinne daß das eigentlich erwartete Optimum nicht erreicht wurde.

Das folgende Beispiel zeigt einen Fall, bei dem dieses Problem auftritt.

Beispiel 3.3.7

Sei für $a > 0$ der euklidische Vektorraum E als $L^2([0, a], \mathbb{R})$ zusammen mit dem Integralskalarprodukt $\langle f | g \rangle := \int f \cdot g$ gegeben und sei $\mathcal{A} := [0, a]$.

Als Erzeugendenfunktionen seien weiterhin

$$\begin{aligned} e_1(r) &:= \mathbf{1}_{[0,r]} && \text{und} \\ e_2(r) &:= \text{id}_{[0,a] \setminus [0,r]} = \text{id}_{]r,a]} \end{aligned}$$

gegeben. Für diese Erzeugendenfunktionen ist die Bedingung

$$\dim(\text{linH}\{\mathbf{1}_{[0,r]}, \text{id}_{]r,a]}\}) = 2$$

aus Korollar 3.3.4 nur für $r \in \mathcal{A}' :=]0, a[$ erfüllt. \mathcal{A}' ist allerdings nicht mehr kompakt. Schränkt man sich jetzt weiter auf eine kompakte Teilmenge $\mathcal{A}'' \subseteq \mathcal{A}'$ ein, tritt für

$$y := \mathbf{1}_{[0,a]}$$

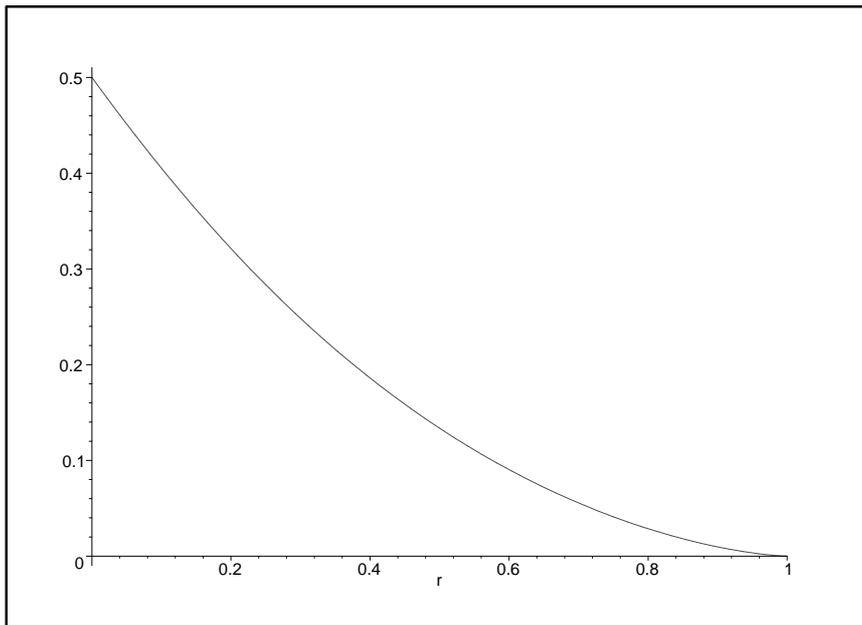


Abbildung 3.4: Verlauf der Zielfunktion $\psi_{g,y}$ aus Beispiel 3.3.7 für $y = \mathbb{1}_{[0,a]}$ und $a = 1$

genau der Fall ein, daß das eigentlich „erwünschte“ Optimum $r = a$ nicht in \mathcal{A}' enthalten ist. Wie leicht zu sehen ist, ist der Punkt a in diesem Fall das einzige (lokale und globale) Optimum von $\psi_{g,y}$ über $[0, a]$, denn es gilt

$$\begin{aligned} e_1(a) &= \mathbb{1}_{[0,a]} = y && \text{und} \\ e_2(a) &= 0, \end{aligned}$$

woraus $\psi_{g,y}(a) = 0$ folgt. Weiterhin ist aber $e_2(r) \neq 0$ für alle $r \in [0, a[$. In diesen Fällen wird durch $e_2(r)$ der Wert von $\psi_{g,y}$ erhöht, woraus $\psi_{g,y}(r) > 0$ für alle $r \in [0, a[$ folgt.

Abbildung 3.4 zeigt den Verlauf von $\psi_{g,y}$ für $a = 1$.

Das eigentlich erwünschte Optimum der Zielfunktion $\psi_{g,y}$ aus Beispiel 3.3.7 wird also genau dann angenommen, wenn die Dimension des Bildraumes $\text{linH}\{e_1(r), e_2(r)\}$ um eins zusammenfällt. Jede Einschränkung des ursprünglichen Intervalls $\mathcal{A} = [0, a]$, um (gemäß Korollar 3.3.4) die Dimensionsproblematik zu umgehen, führt in diesem Beispiel zwingend zu einem unbefriedigenden Optimum. Die Optimierung muß daher über das volle Intervall $[0, a]$ laufen können. Allerdings ist auch Korollar 3.3.6 für dieses Beispiel nicht anwendbar, denn für die Erzeugendenfunktion e_1 gilt

$$\|e_1(r)\| = \sqrt{r} \quad \text{für } r \in]0, a],$$

d.h. die Norm von e_1 geht für $r \rightarrow 0$ gegen Null und nimmt nur für $r = 0$ den Wert Null an. Die normalisierte Version e'_1 von e_1 muß daher wie folgt definiert werden:

$$e'_1(r) = \begin{cases} \frac{1}{\sqrt{r}} \mathbb{1}_{[0,r]} & , r \neq 0 \\ 0 & , r = 0. \end{cases}$$

Man beachte aber, daß der Anschluß im Punkt $r = 0$ nicht stetig ist. Der einzige Grenzwert der als stetiger Anschluß in Frage käme, wäre eben die Nullfunktion, aber da für alle $\varepsilon > 0$

$$\|e'_1(\varepsilon) - 0\|^2 = 1$$

unabhängig von ε ist, fällt auch die Nullfunktion als Limes weg. Damit läßt sich $e'_1(r)$ im Punkt $r = 0$ nicht stetig fortsetzen.

Die Frage nach der Stetigkeit von $\psi_{g,y}$ im Randpunkt $r = 0$ muß daher auf direktem Weg geklärt werden. Der Vollständigkeit halber geschieht dies im Folgenden für beide Anschlußstellen $r = 0$ und $r = a$.

Beispiel 3.3.7 (Fortsetzung)

Für $r \in]0, a[$ gilt zunächst

$$\begin{aligned} \psi_{g,y}^2(r) &= \left\| \mathbf{1}_{[0,a]} - \mathbf{1}_{[0,r]} - \frac{\frac{1}{2}(a^2 - r^2)}{\frac{1}{3}(a^3 - r^3)} \text{id}_{]r,a]} \right\|^2 \\ &= \left\| \mathbf{1}_{]r,a]} - \frac{3(a^2 - r^2)}{2(a^3 - r^3)} \text{id}_{]r,a]} \right\|^2 \\ &= \int_r^a \left(1 - \frac{3(a^2 - r^2)}{2(a^3 - r^3)} x \right)^2 dx \\ &= \int_r^a 1 - 3 \frac{(a^2 - r^2)}{(a^3 - r^3)} x + \frac{9(a^2 - r^2)^2}{4(a^3 - r^3)^2} x^2 dx \\ &= (a - r) - 3 \frac{(a^2 - r^2)}{(a^3 - r^3)} \cdot \frac{1}{2}(a^2 - r^2) + \frac{9(a^2 - r^2)^2}{4(a^3 - r^3)^2} \cdot \frac{1}{3}(a^3 - r^3) \\ &= a - r - \frac{3(a^2 - r^2)^2}{4(a^3 - r^3)} \end{aligned}$$

Weiterhin hat $\psi_{g,y}$ in den Randpunkten 0 und a die Form

$$\begin{aligned} \psi_{g,y}(0) &= \left\| \mathbf{1}_{[0,a]} - \text{proj}_{\text{linH}\{\text{id}_{]0,a]}\} \mathbf{1}_{[0,a]} \right\| \\ &= \left\| \mathbf{1}_{[0,a]} - \frac{\frac{1}{2}a^2}{\frac{1}{3}a^3} \text{id}_{]0,a]} \right\| \\ &= \sqrt{a - \frac{3}{2}a + \frac{3}{4}a} \\ &= \frac{1}{2}\sqrt{a}. \end{aligned}$$

und

$$\begin{aligned} \psi_{g,y}(a) &= \left\| \mathbf{1}_{[0,a]} - \text{proj}_{\text{linH}\{\mathbf{1}_{[0,a]}\} \mathbf{1}_{[0,a]} \right\| \\ &= \left\| \mathbf{1}_{[0,a]} - \mathbf{1}_{[0,a]} \right\| \\ &= 0 \end{aligned}$$

Es fehlt nun noch die Untersuchung der Übergänge in die Randpunkte. Für $r \rightarrow a$ gilt dabei

$$\begin{aligned} \lim_{r \rightarrow a} \psi_{g,y}^2(r) &= \lim_{r \rightarrow a} a - r - \frac{3(a^2 - r^2)^2}{4(a^3 - r^3)} \\ &= \lim_{r \rightarrow a} \frac{-6(a^2 - r^2) \cdot 2r}{12r^2} \\ &= \lim_{r \rightarrow a} \frac{r^2 - a^2}{r} \\ &= 0 \end{aligned}$$

bzw. für $r \rightarrow 0$

$$\lim_{r \rightarrow 0} \psi_{g,y}^2(r) = a - \frac{3a^4}{4a^3} = \frac{1}{4}a.$$

Damit ist $\psi_{g,y}$ stetig in beiden Randpunkten (wie Abbildung 3.4 bereits vermuten ließ).

Natürlich kann in der Praxis nicht für jedes vorliegende $y \in E$ zuerst die Stetigkeit der Zielfunktion überprüft werden. Aber dieses eine positive Stetigkeitsergebnis läßt hoffen, daß sich die Zielfunktion $\psi_{g,y}$ auch für andere Funktionen aus E gutartig an den Randpunkten verhält. Der folgende Absatz zeigt, daß das auch tatsächlich der Fall ist:

Ersetzt man das bisherige $y = \mathbb{1}_{[0,a]}$ durch eine beliebige Funktion $h \in L^2([0, a], \mathbb{R})$, führt das für $r \in]0, a[$ zu der Zielfunktion

$$\begin{aligned} \psi_{g,h}(r) &= \left\| h - \text{proj}_{\text{linH}\{\mathbb{1}_{[0,r]}, \text{id}_{]r,a]}\} h \right\| \\ &= \left\| h - \frac{1}{r} \langle \mathbb{1}_{[0,r]} | h \rangle \mathbb{1}_{[0,r]} - \frac{3}{a^3 - r^3} \langle \text{id}_{]r,a]} | h \rangle \text{id}_{]r,a]} \right\|. \end{aligned}$$

Im Übergang $r \rightarrow 0$ gilt dann für den ersten Koeffizienten $\frac{1}{r} \langle \mathbb{1}_{[0,r]} | h \rangle$ der Projektionsabbildung:

$$\begin{aligned} \lim_{r \rightarrow 0} \frac{1}{r} \langle \mathbb{1}_{[0,r]} | h \rangle &= \lim_{r \rightarrow 0} \frac{1}{r} \int_0^r h(x) d\lambda(x) \\ &= \lim_{r \rightarrow 0} \frac{H(r) - H(0)}{r} \quad (\text{Stammfunktion}) \\ &= \lim_{r \rightarrow 0} \frac{h(r)}{1} \\ &= 0, \end{aligned}$$

was insgesamt zu

$$\lim_{r \rightarrow 0} \psi_{g,h}(r) = \left\| h - \text{proj}_{\text{linH}\{\text{id}_{]r,a]}\} h \right\| = \psi_{g,h}(0)$$

führt. Damit ist $\psi_{g,h}$ im Punkt 0 stetig für alle $h \in L^2([0, a], \mathbb{R})$. Die Stetigkeit im Punkt a zeigt man analog.

Die Zielfunktion in Beispiel 3.3.7 verhält sich also völlig unproblematisch bzgl. ihrer Stetigkeit im Nullpunkt, obwohl die normalisierte Erzeugendenfunktion e'_1 unstetig in eben diesem Nullpunkt ist. Das folgende Beispiel dagegen zeigt eine Modellierung, bei der im Dimensionsübergang keine Stetigkeit mehr vorhanden ist.

Beispiel 3.3.8

Sei der Vektorraum $E = \mathbb{R}^2$ zusammen mit dem Standardskalarprodukt gegeben und sei $\mathcal{A} = [0, \frac{\pi}{2}]$. Es sei nur eine einzige Erzeugendenfunktion

$$e(\alpha) := \left(\frac{\pi}{2} - \alpha\right) \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}$$

vorhanden. Die zugehörige semilineare Funktion $g : \mathbb{R} \times [0, \frac{\pi}{2}] \rightarrow \mathbb{R}^2$ hat dann die Form

$$g(x, \alpha) = x \begin{pmatrix} \frac{\pi}{2} - \alpha \\ \cos \alpha \\ \sin \alpha \end{pmatrix}.$$

Wird als Metrik auf \mathcal{A} etwa die Absolutbetragsmetrik gewählt, ist e stetig und daraus folgt nach Satz 3.3.1 die Stetigkeit von g .

Für $y = (0, 1)' \in \mathbb{R}^2$ und $\alpha \neq \frac{\pi}{2}$ gilt dann

$$\text{proj}_{\text{linH}\{e(\alpha)\}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \sin \alpha \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}$$

und damit

$$\begin{aligned} \psi_{g,y}(\alpha) &= \left\| \begin{pmatrix} 0 \\ 1 \end{pmatrix} - \text{proj}_{\text{linH}\{e(\alpha)\}} \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\| \\ &= \sqrt{\sin^2 \alpha \cos^2 \alpha + (1 - \sin^2 \alpha)^2} \end{aligned}$$

Interessant an diesem Beispiel ist der Randpunkt $\alpha = \frac{\pi}{2}$. Für $\alpha \in [0, \frac{\pi}{2}[$ gilt zunächst

$$\lim_{\alpha \rightarrow \frac{\pi}{2}} \psi_{g,y}(\alpha) = \lim_{\alpha \rightarrow \frac{\pi}{2}} \sqrt{\sin^2 \alpha \cos^2 \alpha + (1 - \sin^2 \alpha)^2} = 0,$$

d.h. $\psi_{g,y}$ nähert sich für $\alpha \rightarrow \frac{\pi}{2}$ dem Wert 0 an. Andererseits ist aber nach Voraussetzung $e(\frac{\pi}{2}) = 0$ und daraus folgt für $\psi_{g,y}$ im Punkt $\frac{\pi}{2}$

$$\psi_{g,y} \left(\frac{\pi}{2}\right) = \|y\| = 1 \neq 0.$$

$\psi_{g,y}$ ist also unstetig im Punkt $\alpha = \frac{\pi}{2}$. Diese Unstetigkeit resultiert aus der in $\psi_{g,y}$ enthaltenen Projektionsabbildung $\alpha \mapsto \text{proj}_{\text{linH}\{(\frac{\pi}{2}-\alpha)\begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}\}} y$ der Form

$$\text{proj}_{\text{linH}\{(\frac{\pi}{2}-\alpha)\begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}\}} y = \begin{cases} \sin \alpha \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix}, & \alpha \neq \frac{\pi}{2} \\ 0, & \alpha = \frac{\pi}{2}, \end{cases}$$

die einen unstetigen Anschluß im Punkt $\alpha = \frac{\pi}{2}$ aufweist.

Im Gegensatz zu Beispiel 3.3.7 ist in Beispiel 3.3.8 die Projektionsabbildung an den Anschlußstellen nicht stetig. Das zeigt, wie leicht Unstetigkeitsstellen in der Zielfunktion entstehen können, selbst wenn bei der Modellformulierung stetige Erzeugendenfunktionen gewählt wurden.

Beispiel 3.3.8 (Fortsetzung)

Bisher wurde die Unstetigkeit der Zielfunktion $\psi_{g,y}$ nur für ein bestimmtes $y \in \mathbb{R}^2$ nachgewiesen. Im folgenden geht es um das Verhalten von $\psi_{g,y}$ für allgemeines $y \in E$. Zunächst steht fest, daß die Stetigkeit dieser Funktion von der Stetigkeit ihrer Teilabbildung

$$\alpha \mapsto \text{proj}_{\text{linH}\left\{\begin{pmatrix} \frac{\pi}{2}-\alpha \\ \sin \alpha \end{pmatrix}\right\}} y$$

abhängt. Um die Notation zu verkürzen, wird der darin enthaltene lineare Projektionsoperator im folgenden mit

$$\mathbf{P}_\alpha := \text{proj}_{\text{linH}\left\{\begin{pmatrix} \frac{\pi}{2}-\alpha \\ \sin \alpha \end{pmatrix}\right\}} \in \mathbb{R}^{2 \times 2}$$

abgekürzt. Für $y \in \mathbb{R}^2$ und $\alpha \in [0, \frac{\pi}{2}[$ gilt dann

$$\mathbf{P}_\alpha y = \left\langle \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} \middle| y \right\rangle \begin{pmatrix} \cos \alpha \\ \sin \alpha \end{pmatrix} = \dots = \begin{pmatrix} \cos^2 \alpha & \sin \alpha \cos \alpha \\ \sin \alpha \cos \alpha & \sin^2 \alpha \end{pmatrix} y,$$

d.h.

$$\mathbf{P}_\alpha = \begin{pmatrix} \cos^2 \alpha & \sin \alpha \cos \alpha \\ \sin \alpha \cos \alpha & \sin^2 \alpha \end{pmatrix}.$$

Die Abbildung $\alpha \mapsto \mathbf{P}_\alpha$ ist also stetig für alle $\alpha \in [0, \frac{\pi}{2}[$ und es gilt

$$\lim_{\alpha \rightarrow \frac{\pi}{2}} \mathbf{P}_\alpha = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Für $\alpha = \frac{\pi}{2}$ dagegen folgt aus $e(\frac{\pi}{2}) = 0$ direkt

$$\mathbf{P}_{\frac{\pi}{2}} = 0,$$

was bedeutet, daß die Abbildung $\alpha \mapsto \mathbf{P}_\alpha$ unstetig im Punkt $\frac{\pi}{2}$ ist. Damit ist $\psi_{g,y}$ unstetig im Punkt $\frac{\pi}{2}$ für $y \in \mathbb{R}^2$, $y \neq \begin{pmatrix} \gamma \\ 0 \end{pmatrix}$, $\gamma \in \mathbb{R}$.

Statt der Projektionsabbildung $y \mapsto \text{proj}_{\text{linH}\{e(\alpha)\}}$ wurde im obigen Beispiel zum Schluß der Projektionsoperator \mathbf{P}_α betrachtet. Dies führt zu einer anderen Sichtweise der semilinearen Approximation, die statt der Erzeugendenfunktion e_i die von ihr erzeugten linearen Operatoren betrachtet.

Satz 3.3.9

Sei E ein euklidischer Vektorraum, (\mathcal{A}, d) ein metrischer Raum sowie $y \in E$. Bezeichne weiterhin

$$\{\mathbf{P}_A \in \text{Hom}(E, E) \mid A \in \mathcal{A} \text{ und } \mathbf{P}_A \text{ ist Projektionsoperator}\}$$

eine Menge linearer Projektionsoperatoren. Dann sind die folgenden beiden Aussagen äquivalent:

1. Durch $\inf_{A \in \mathcal{A}} \|y - \mathbf{P}_A y\|$ wird ein semilineares Approximationsproblem mit höchstens k Erzeugendenfunktionen beschrieben
2. $\forall A \in \mathcal{A} : \dim \mathfrak{S}(\mathbf{P}_A) \leq k$.

Beweis Wird durch $\inf_{A \in \mathcal{A}} \|y - \mathbf{P}_A y\|$ ein semilineares Approximationsproblem mit höchstens k Erzeugendenfunktionen beschrieben, existieren k Erzeugendenfunktionen $e_1, \dots, e_k : \mathcal{A} \rightarrow E$ mit der Eigenschaft

$$\forall A \in \mathcal{A} \exists x_1, \dots, x_k \in \mathbb{R} : \mathbf{P}_A y = \sum_{i=1}^k x_i e_i(A).$$

Für jedes $A \in \mathcal{A}$ bilden diese Erzeugendenfunktionen ein Erzeugendensystem von $\mathfrak{S}(\mathbf{P}_A)$. Daraus folgt direkt $\dim \mathfrak{S}(\mathbf{P}_A) \leq k$.

Gilt andererseits $\dim \mathfrak{S}(\mathbf{P}_A) \leq k$, dann existiert eine k -elementige erweiterte Orthonormalbasis $\{v_{1,A}, \dots, v_{k,A}\}$ von $\mathfrak{S}(\mathbf{P}_A)$ im Sinne von Definition 3.3.5. Setzt man nun $e_i(A) := v_{i,A}$ für $i = 1, \dots, k$, erhält man k Erzeugendenfunktionen $e_i : \mathcal{A} \rightarrow E$. Da die \mathbf{P}_A nach Voraussetzung lineare Projektionsoperatoren sind, gilt weiterhin

$$\mathbf{P}_A y = \text{proj}_{\mathfrak{S}(\mathbf{P}_A)} y = \text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}} y = \sum_{i=1}^k \langle v_{i,A} | y \rangle v_{i,A}$$

und daraus folgt $\|y - \mathbf{P}_A y\| = \left\| y - \sum_{i=1}^k \langle e_i(A) | y \rangle e_i(A) \right\|$ für alle $y \in E$. □

Satz 3.3.10

Sei ein semilineares Approximationsproblem gemäß Definition 3.2.1 gegeben. Werden dann für $A \in \mathcal{A}$ mit

$$\mathbf{P}_A := \text{proj}_{\text{linH}\{e_1(A), \dots, e_k(A)\}}$$

die in der Zielfunktion $\psi_{g,y}$ enthaltenen linearen Projektionsoperatoren bezeichnet, gilt:

1. Ist die Abbildung $A \mapsto \mathbf{P}_A$ unstetig im Punkt A_0 und $y \in E$ derart, daß $\lim_{A \rightarrow A_0} \mathbf{P}_A y \notin \mathfrak{S}(\mathbf{P}_{A_0})$, so ist auch $\psi_{g,y}$ unstetig im Punkt A_0 .
2. Existiert $\mathcal{A}' \subseteq \mathcal{A}$ derart, daß $\psi_{g,y}$ stetig über \mathcal{A}' und die Menge $\{\mathbf{P}_A \mid A \in \mathcal{A}'\}$ aller Projektionsoperatoren über \mathcal{A}' kompakt² ist, existiert ein (sogar globales) Minimum von $\psi_{g,y}$ über $\{\mathbf{P}_A \mid A \in \mathcal{A}'\}$.

²Um $\{\mathbf{P}_A \mid A \in \mathcal{A}'\}$ zu einem metrischen Raum zu machen, wähle man hierfür die norminduzierte Metrik für eine geeignete Operatornorm.

Beweis Aus den Voraussetzungen des ersten Unterpunktes folgt direkt die Unstetigkeit der Abbildung $A \mapsto \mathbf{P}_A y$ im Punkt A_0 . Damit ist auch die Gesamtabbildung $\psi_{g,y}$ unstetig in A_0 .

Für den Beweis des zweiten Unterpunktes betrachte man die Funktion

$$\tilde{\psi}_{g,y} : \{\mathbf{P}_A \mid A \in \mathcal{A}'\} \rightarrow \mathbb{R}, \quad \tilde{\psi}_{g,y}(\mathbf{P}_A) := \|y - \mathbf{P}_A y\|.$$

Wählt man auf $\{\mathbf{P}_A \mid A \in \mathcal{A}'\}$ eine geeignete Operatornorm, ist $\tilde{\psi}_{g,y}$ zusammen mit der norminduzierten Metrik auf $\{\mathbf{P}_A \mid A \in \mathcal{A}'\}$ stetig und es gilt

$$\inf_{A \in \mathcal{A}} \psi_{g,y}(A) = \inf_{A \in \mathcal{A}} \|y - \mathbf{P}_A y\| = \inf_{\{\mathbf{P}_A \mid A \in \mathcal{A}'\}} \|y - \mathbf{P}_A y\| = \inf_{\{\mathbf{P}_A \mid A \in \mathcal{A}'\}} \tilde{\psi}_{g,y}.$$

Ist nun $\{\mathbf{P}_A \mid A \in \mathcal{A}'\}$ kompakt, existiert ein sogar globales Minimum von $\tilde{\psi}_{g,y}$ (und damit von $\psi_{g,y}$) über $\{\mathbf{P}_A \mid A \in \mathcal{A}'\}$. \square

Dieser Satz beinhaltet eine Kernaussage über die semilineare Approximation: Zwar ist $\psi_{g,y}$ die zu minimierende Zielfunktion, aber von eigentlichem Interesse ist nur die Funktion $\tilde{\psi}_{g,y}$ aus dem obigen Beweis zusammen mit der Menge $\{\mathbf{P}_A \mid A \in \mathcal{A}\}$ der Projektionsoperatoren. Bereits Beispiel 3.3.7 zeigte, daß die Form dieser Menge nicht von der Form der Erzeugendenfunktionen abhängen muß, denn die Zielfunktion aus diesem Beispiel verhielt sich gutartig obwohl die normalisierte Erzeugendenfunktion e'_1 unstetig im Nullpunkt war. In Beispiel 3.3.8 wird die Situation noch deutlicher: Die Menge $\{\mathbf{P}_\alpha \mid \alpha \in [0, \frac{\pi}{2}]\}$ wäre nur dann kompakt, wenn $\mathbf{P}_{\frac{\pi}{2}}$ die Form

$$\mathbf{P}_{\frac{\pi}{2}} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

besäße. Dies ist nicht der Fall, wie auf Seite 50 gezeigt wurde.

Man beachte, daß für den zweiten Unterpunkt von Satz 3.3.10 weder die Stetigkeit der Abbildung $A \mapsto \mathbf{P}_A$ noch die Kompaktheit von \mathcal{A} verlangt wird. Es ist also nicht sichergestellt, daß das im Satz garantierte Minimum auch existiert, wenn über \mathcal{A} und nicht über $\{\mathbf{P}_A \mid A \in \mathcal{A}\}$ optimiert wird. Da aber die Minimierung der Zielfunktion $\psi_{g,y}$ eben nicht über $\{\mathbf{P}_A \mid A \in \mathcal{A}\}$, sondern über \mathcal{A} erfolgt, wird in den Anwendungen in Kapitel 6 der 1. Unterpunkt von Satz 3.3.10 (die Stetigkeit von $A \mapsto \mathbf{P}_A$) zusammen mit der Kompaktheit von \mathcal{A} als Kriterium für die Zielfunktion herangezogen.

Bemerkung 3.3.11

Zum Schluß soll noch eine informelle aber dennoch sehr anschauliche Erklärung gegeben werden, wann ein unstetiges Verhalten in der Zielfunktion $\psi_{g,y}$ erwartet werden kann: Dazu werden nicht die Erzeugendenfunktionen e_1, \dots, e_k bzw. die Projektionsoperatoren $\{\mathbf{P}_A \mid A \in \mathcal{A}\}$ an sich, sondern nur die von ihnen erzeugten Bildräume $\mathfrak{S}(\mathbf{P}_A) = \text{linH}\{e_1(A), \dots, e_k(A)\}$ betrachtet: In Beispiel 3.3.8 nähert sich der von $e(\alpha)$ erzeugte Unterraum von der x -Achse aus für $\alpha \rightarrow \frac{\pi}{2}$ immer weiter der y -Achse an,

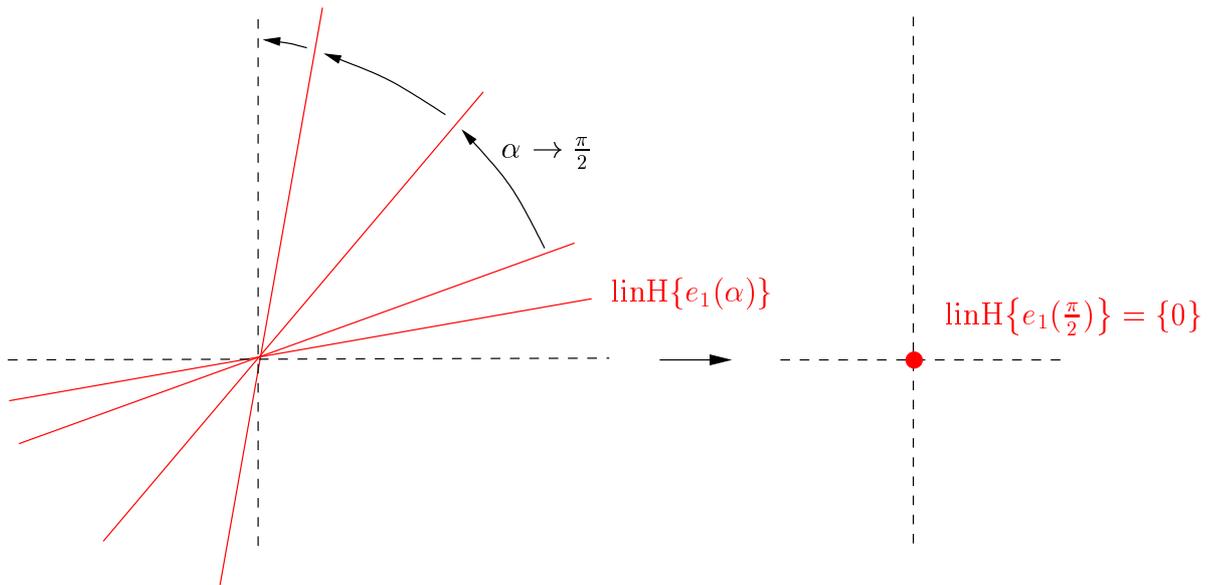


Abbildung 3.5: Skizze der von e_1 aus Beispiel 3.3.8 erzeugten Unterräume. Für $\alpha \rightarrow 0$ nähern sie sich immer mehr der y -Achse an. Die y -Achse wird aber nie erreicht, denn genau im Punkt $\frac{\pi}{2}$ fällt der von $e_2(\frac{\pi}{2})$ erzeugte Unterraum in sich zusammen.

fällt aber dann für $\alpha = \frac{\pi}{2}$ plötzlich in sich zusammen (Abbildung 3.5). Auch ohne eine spezielle Unterraummetrik definiert zu haben, kann man hier informell von einem „unstetigen“ Verhalten der von $e(\alpha)$ erzeugten Unterräume sprechen. In Beispiel 3.3.7 dagegen nähert sich der von $e_1(r)$ erzeugte Unterraum $\text{linH}\{\mathbb{1}_{[0,r]}\}$ für $r \rightarrow 0$ kontinuierlich dem für $r = 0$ erzeugten Nullraum an. Im Gegensatz zu Beispiel 3.3.8 liegt hier ein „stetiges“ Verhalten der Unterräume für $r \rightarrow 0$ vor. Die Stetigkeit der Teilfunktion $A \mapsto \mathbf{P}_A$ der Zielfunktion $\psi_{g,y}$ hängt also anschaulich gesehen von der „Stetigkeit“ der Abbildung

$$A \mapsto \text{linH}\{e_1(A), \dots, e_k(A)\}$$

ab, die jedem Element $A \in \mathcal{A}$ den Unterraum $\text{linH}\{e_1(A), \dots, e_k(A)\}$ von E zuordnet.

Diese Überlegung ist nur informeller Natur und basiert natürlich auf dem Zusammenhang zwischen den Projektionsoperatoren \mathbf{P}_A und ihren Bildräumen. Sie kann aber dennoch eine Hilfestellung dafür darstellen, wann Unstetigkeiten in der Zielfunktion eines semilinearen Approximationsproblems zu erwarten sind. Eine genauere Untersuchung des hier beschriebenen Zusammenhangs könnte interessant sein, dürfte sich aber in aller Allgemeinheit schwierig gestalten und würde den Rahmen dieser Arbeit sprengen.

Mit der semilinearen Approximation wird eine sehr allgemeine Möglichkeit zur mengenbasierten Passung bereitgestellt. Im folgenden Kapitel wird beschrieben, wie dieses Verfahren geeignet in bestehende Aufnahmeprozesse integriert werden kann.

4 Sensoren und Sensoraufnahmen

Dieses Kapitel befaßt sich mit der Einbindung der semilinearen Approximation in Probleme der Bildverarbeitung. Dazu ist es vor allem nötig, den *Bildaufnahmeprozess* näher zu beleuchten.

Die folgenden Betrachtungen beziehen sich nur auf die Entstehung von einkanaligen Bildern (Graustufenbildern). Der Hauptgrund dafür ist, daß im Rahmen dieser Arbeit nur einkanalige Bilder untersucht wurden. Eine Erweiterung auf mehrkanalige Bilder wäre aber durch eine kanalweise Verallgemeinerung durchaus möglich. In Kapitel 8 findet sich dazu eine kurze Betrachtung.

Im allgemeinen wird für den Bildentstehungsprozeß das u.a. in [Haa99] beschriebene Modell verwendet, das dem prinzipiellen Aufbau einer CCD- bzw. CMOS-Kamera Rechnung trägt: Ein reales Objekt, das Lichtstrahlen reflektiert oder auch selbst aussendet, wird durch ein Sensorarray in einem diskreten Bild festgehalten. Auf dem Weg vom Objekt zum Sensorarray passieren die Lichtstrahlen ein Linsensystem, das für die Fokussierung des Objektes zuständig ist. Die Menge der einfallenden Lichtstrahlen kann durch ein Blendensystem geregelt werden.

Zunächst gilt, daß das beobachtete Objekt zeitabhängig — z.B. durch Eigenbewegung oder durch veränderten Lichteinfall — sein Aussehen ändern kann. Daher sollte das beobachtete Szenario (vor seiner Digitalisierung durch das Sensorarray) korrekterweise durch eine zeitabhängige Funktion $f(x, y, t)$ modelliert werden. In der Regel geht man aber davon aus, daß sich \tilde{f} im Zeitraum Δt einer Aufnahme nur unwesentlich ändert. In diesem Fall kann statt \tilde{f} die vereinfachte zeitunabhängige Version $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ verwendet werden. f wird in der Literatur auch als *Sensorinputfunktion* bezeichnet (mehr dazu auf Seite 57).

4.1 Das Sensorarray

Im Fall einer digitalen Kamera passieren die Lichtstrahlen auf ihrem Weg zum Sensorarray zunächst ein Linsensystem, das für die Fokussierung eines Objektes zuständig ist. Dieses Linsensystem wird normalerweise in die Modellierung des Bildaufnahmeprozesses mit einbezogen. Es existieren aber auch Sensorsysteme, die kein solches Linsensystem besitzen. Die wohl bekanntesten Vertreter solcher Sensorsysteme sind (digitale) Röntgensysteme, wie z.B. die Durchleuchtungssysteme für Handgepäck auf Flughäfen. Im Gegensatz zu analogen Röntgensystemen, bei denen die Bilderzeugung durch einen

belichteten Film geschieht, besitzen diese Systeme Detektoren, die Röntgenstrahlung absorbieren und in digitaler Form weiterleiten.

Jedes digitale Röntgensystem besitzt den folgenden Grobaufbau: Von einer Quelle aus werden Röntgenstrahlen ausgesendet, die dann das zu untersuchende Objekt passieren, innerhalb dessen sie — je nach Ausgangsenergie der Strahlung und je nach Objektdichte — geschwächt werden. Mit der so entstandenen Endenergie trifft die Röntgenstrahlung schließlich auf die Detektoren, die das Signal aufnehmen und an den Rechner weitergeben (Abbildung 4.1). Handelt es sich um ein CT-System, werden Quelle und Detektor rotiert, um weitere Aufnahmen des Objektes zu erhalten ([BS96] oder Kapitel 7).

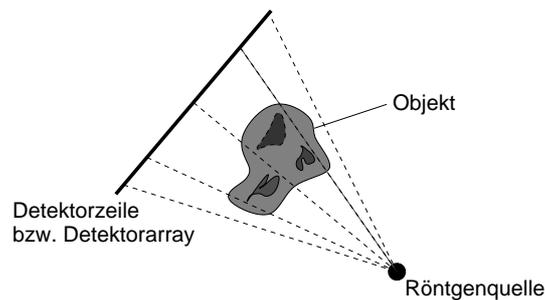


Abbildung 4.1: Schematische Darstellung eines Röntgensystems.

Bei den Detektoren handelt es sich (aus Kostengründen) meist um eine einzige Zeile. Ein solches Röntgensystem kann auch als Zeilenkamera aufgefaßt werden. Ein flächiges Bild wird erzeugt, indem das Objekt senkrecht zur Zeile transportiert wird und in regelmäßigen Abständen eine weitere Zeile aufgenommen wird. Meist abstrahiert man von diesem Aufnahme-prozeß und modelliert das System durch ein zweidimensionales Sensorarray.

Abgesehen von der Art der Strahlung unterscheidet sich der prinzipielle Aufbau der Röntgensensorik also wenig von dem Sensorsystem einer digitalen Grauwertkamera. Läßt man die Art der Strahlung außer acht, bleibt das in Abbildung 4.2 skizzierte einfache Modell eines Sensorsystems übrig: Die einfallende Strahlung trifft auf ein Sensorarray, das für die Weitergabe der Information zuständig ist.

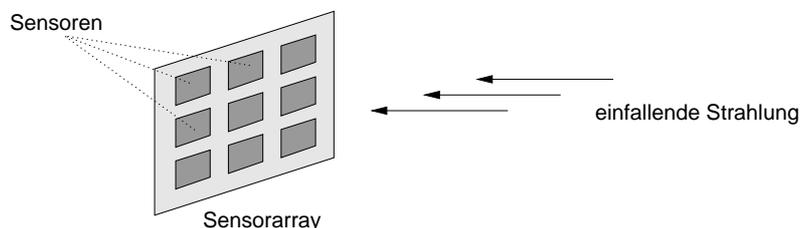


Abbildung 4.2: Schematische Darstellung eines Sensorarrays

Gewünscht ist nun eine Modellierung der Sensorauswertung. Dies kann als Punktauswertung der Sensorinputfunktion f geschehen, was für viele Anwendungen ausreichend ist.

Eine allgemeingültigere Modellierung ergibt sich jedoch, wenn eine Sensorauswertung als eine Linearform aufgefaßt wird, die auf die Sensorinputfunktion angewendet werden kann. In diesem Fall können charakteristische Eigenschaften jedes Sensorelements mit berücksichtigt werden.

Der Begriff der *Sensorinputfunktion* wurde bis jetzt eher intuitiv benutzt. Um diesen Begriff genauer fassen zu können wird zuerst der Begriff des *Sensors* benötigt:

Definition 4.1.1 (Sensorelement, Sensor)

Für eine kompakte Menge $A \in \mathfrak{K}(\mathbb{R}^2)$ sei ein endliches Maß $\mu : \mathcal{B}(A) \rightarrow \mathbb{R}$ gegeben. Dann heißt das Paar (A, μ) ein Sensorelement. A heißt die Apertur (aktive Oberfläche) und μ die Sensorcharakteristik des Sensorelements.

Ein Sensor(array) ist eine endliche Familie $(A_i, \mu_i)_{i \in I}$ von Sensorelementen.

Die Anordnung der Sensorelemente, d.h. die Geometrie des Sensors ist in dieser Definition nicht wesentlich. Theoretisch könnten die Sensorelemente eines so definierten Sensors nebeneinander aber auch übereinander oder nur teilweise überlappend angeordnet sein. Die Anordnung der Aperturen innerhalb einer Fläche wird durch Lage und Größe der Aperturen $A_i \subseteq \mathbb{R}^2$ festgesetzt.

Da die Art der Strahlung, die die Sensorelemente auffangen, durch Definition 4.1.1 nicht festgesetzt ist, erlaubt diese Definition einen leichten Zugang zu Multisensorsystemen. Multisensorsysteme zeichnen sich dadurch aus, daß von einem beobachteten Szenario mehrere verschiedenartige Bildinformationen verarbeitet werden, also ein vektorwertiges (bzw. mehrkanaliges) Ergebnis vorliegt. Diese unterschiedlichen Bildinformationen können dabei verschiedene Farbkanäle, aber auch aus Laserlichtschnitt- bzw. Röntenaufnahmen gewonnene Daten ein- und desselben Objektes darstellen. In diesem Sinne kann auch eine Farbkamera als Multisensorsystem aufgefaßt werden. Ist also eine Apertur A nach Definition 4.1.1 gegeben, kann ein n -kanaliges Sensorelement auf A durch n einzelne Sensorelemente $(A, \mu_1), \dots, (A, \mu_n)$ modelliert werden, wobei μ_1, \dots, μ_n die jeweiligen Sensorcharakteristiken bezeichnen. Die Sensorcharakteristik eines Röntgendetektors wird sich im Vergleich zu der eines Laserlichtschnittdetektors sicherlich signifikant unterscheiden. Werden dagegen durch $(A_1, \mu_1), (A_2, \mu_2), (A_3, \mu_3)$ die drei Sensorelemente der Farbkanäle Rot, Grün und Blau einer Farbkamera beschrieben, liegt die Annahme nahe daß die Sensorcharakteristiken μ_1, μ_2 und μ_3 im wesentlichen identisch sind. Identifiziert man nun die Aperturen A_1, A_2 und A_3 miteinander, werden auch die Sensorelemente an sich identisch. Eine Unterscheidung dieser drei Sensorelemente kann also nicht im Kontext von Definition 4.1.1 geschehen. Viel leichter (und selbsterklärender) funktioniert eine Unterscheidung über die Sensorinputfunktion:

Definition 4.1.2 (Sensorinputfunktion)

Sei ein Sensorelement (A, μ) laut Definition 4.1.1 gegeben. Dann heißt jede Funktion $f \in L^1_\mu(A, \mathbb{R})$ eine Sensorinputfunktion (von (A, μ)) und der Wert $l(f)$ der Linearform

$$l : L^1_\mu(A, \mathbb{R}) \rightarrow \mathbb{R}, \quad l(f) := \int_A f d\mu$$

die Aufnahme oder Digitalisierung (von f bzgl. (A, μ)).

Definition 4.1.2 setzt die Sensorinputfunktion nur für einzelne Sensorelemente fest. Ist dagegen ein ganzer Sensor $(A_i, \mu_i)_{i \in I}$ gegeben, kann man jedem Element (A_i, μ_i) dieses Sensors eine eigene Sensorinputfunktion f_i zuzuordnen. Man erhält damit eine Familie $(f_i)_{i \in I}$ von Sensorinputfunktionen für die gilt

$$\forall i \in I : f_i \in L^1_{\mu_i}(A_i, \mathbb{R}).$$

Die Aufnahme eines Sensors $(A_i, \mu_i)_{i \in I}$ besteht dann aus den Einzelaufnahmen $(l_i(f_i))_{i \in I}$ der Sensorelemente.

Abbildung 4.3 zeigt anhand eines ideellen Bildaufnahmeprozesses, wie zwei verschiedene Sensorinputfunktionen eines Objektes aussehen könnten. Das vorliegende Objekt wird auf einen Tisch gelegt, über dem (parallel zum Tisch) ein Sensorarray montiert ist. Das Bild, wie es von diesem Sensorarray „gesehen“ wird, bevor es in digitaler Form weitergeleitet wird, entspricht der Sensorinputfunktion. Die Form der Sensorinputfunktion ist abhängig davon, auf welche Strahlungsart die Sensorelemente reagieren. Zwei mögliche Sensorinputfunktionen sind in Abbildung 4.3 unten dargestellt: Links ist das Objekt nicht-transparent auf dem Hintergrund zu sehen, wie es zum Beispiel von einer Grauwertkamera wahrgenommen werden könnte. Rechts dagegen scheint der Hintergrund durch das Objekt durch, wie es etwa bei der Verwendung einer Röntgensensorik der Fall sein könnte. Der Hintergrund besitzt in beiden Fällen ein periodisches Muster. Natürlich ist dieses Beispiel etwas akademisch, denn ein Hintergrund, der von Grauwert- und Röntgensensor identisch wahrgenommen wird, tritt in der Praxis kaum auf. Zudem sind in beiden Fällen die Seitenflächen nicht mehr erkennbar, was eine ideale Anordnung der Sensoren senkrecht zur Objektfläche voraussetzt. Derartige Aufnahmen sind in der Praxis kaum realisierbar. Im vorliegenden Fall verdeutlichen sie aber die Art der Sensorinputfunktion, die dem jeweiligen Bild entspricht wie es von den Sensorelementen „gesehen“ wird, bevor die Bildinformation in aufintegrierter bzw. digitalisierter Form weitergeleitet wird.

Bemerkung 4.1.3

Die Sensorinputfunktion stellt ein eigentlich nicht visualisierbares Konstrukt dar. So sind auch die Beispielbilder der Sensorinputfunktion aus Abbildung 4.3 durch den Druckprozeß digitalisiert worden und stellen daher genaugenommen keine Sensorinputfunktionen im eigentlichen Sinne dar. Als Gedankenstütze ist Abbildung 4.3 jedoch sehr hilfreich. Insgesamt gesehen ist die Sensorinputfunktion aber häufig nur von ideellem Wert. Sie wird als Zwischenschritt meist übersprungen und vom zu untersuchenden Szenario direkt zur digitalen Aufnahme weitergegangen.

Für die in Abbildung 4.3 skizzierten Beispiele wird jeweils nur eine einzige Sensorinputfunktion verwendet, die das gesamte Sensorarray (und damit auch das gesamte Objekt) abdeckt. Dies wird aber in Definition 4.1.2 so gar nicht gefordert. Eine Sensorinputfunktion kann prinzipiell von Sensorelement zu Sensorelement verschieden sein — je nach

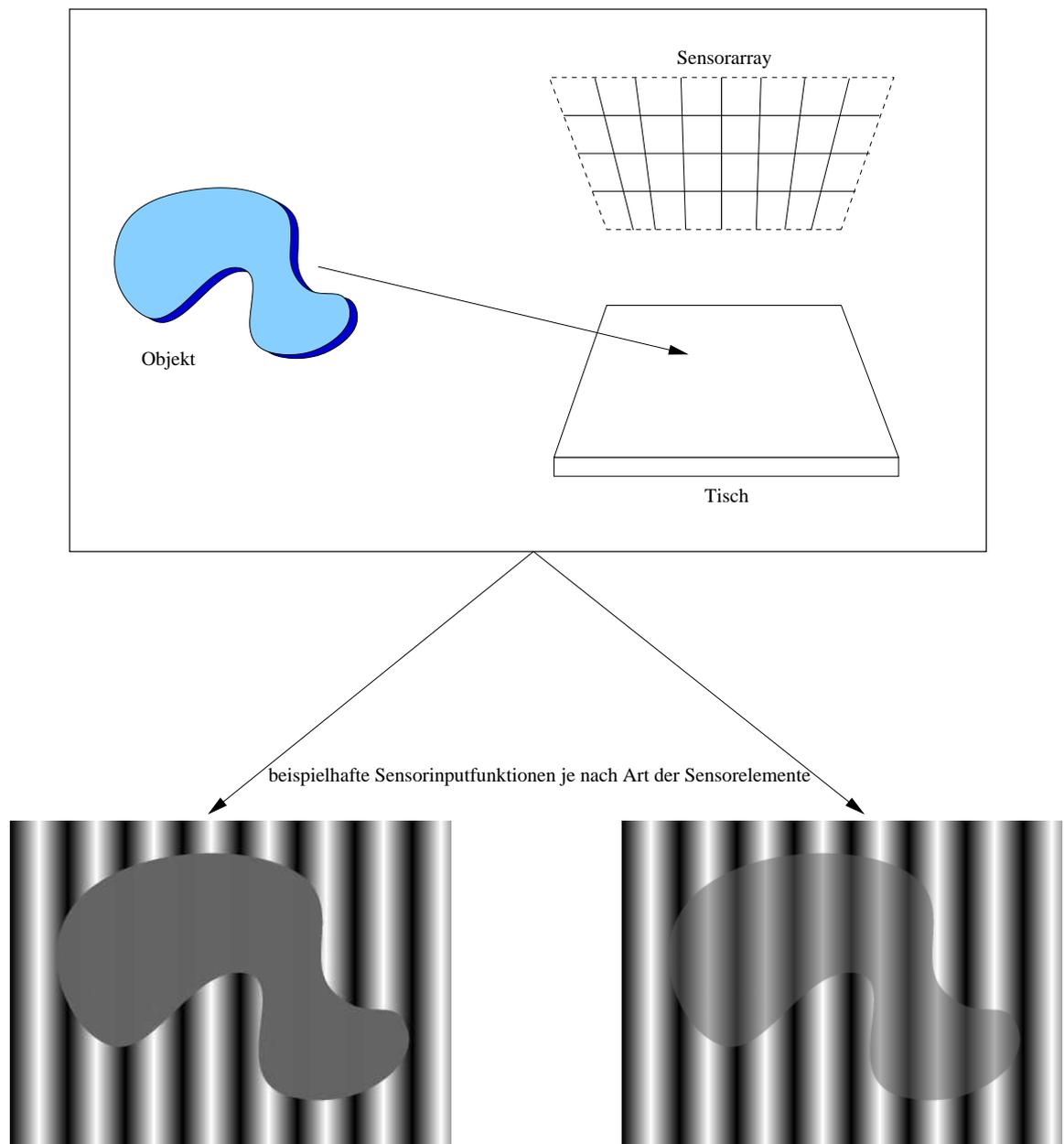


Abbildung 4.3: Beispiel für zwei verschiedene Sensorinputfunktionen eines Objekts. Das eigentliche Objekt besitzt eine gewisse Dicke, die Seitenflächen sind aber in beiden Aufnahmen nicht mehr zu erkennen (vor allem — aber natürlich nicht nur — daran erkennt man, daß das Objekt sowie seine Sensorinputfunktionen künstlich generiert sind).

Art des Sensorelements. Die Sensorelemente aus Abbildung 4.3 sind also in der Hinsicht speziell, als daß sie alle auf die gleiche Art von Strahlung reagieren. Sensoren, deren Sensorelemente diese Eigenschaft besitzen, können *bildhomogen* genannt werden. Üblicherweise sind die meisten Sensoren bildhomogen, eine willkürliche Mischung verschiedener Typen von Sensorelementen kommt in der Praxis am ehesten in Multisensoranlagen vor. Aber auch da werden gleichartige Sensorelemente zu Arrays zusammengefaßt.

Eine andere Art der Homogenität beschreibt die folgende Definition:

Definition 4.1.4

Ein Sensor $(A_i, \mu_i)_{i \in I}$ heißt (sensor-)homogen, falls eine Apertur A und eine Familie $(t_i)_{i \in I}$ mit $t_i \in \mathbb{R}^2$ sowie ein Maß ν auf $\mathcal{B}(A)$ existiert, so daß

$$\forall i \in I : A_i = A + t_i \wedge \mu_i = \nu(\cdot - t_i).$$

Sensorhomogenität bedeutet, daß die Sensorelemente eines Sensors (bis auf Translation) die gleiche Sensorcharakteristik und die gleiche Aperturform besitzen. In der Regel ist das bei allen gängigen Systemen der Fall.

Bemerkung 4.1.5

Im Allgemeinen kann man für die Sensorcharakteristiken μ_i eines Sensors $(A_i, \mu_i)_{i \in I}$ voraussetzen, daß sie eine Dichte bzgl. λ^2 besitzen, d.h. es existiert eine nichtnegative $\mathcal{B}(A_i)$ -meßbare Funktion $p_i : A_i \rightarrow \mathbb{R}$ so daß $\mu_i = p_i \lambda^2$ für alle $i \in I$ ist. Es gilt also

$$l_i(f) = \int_{A_i} f d\mu_i = \int_{A_i} f \cdot p_i d\lambda^2.$$

Für homogene Sensoren nach Definition 4.1.4 läßt sich dies weiter zu

$$l_i(f) = \int_{A_i} f \cdot p_i d\lambda^2 = \int_{A_i} f \cdot p(\cdot - t_i) d\lambda^2$$

vereinfachen, wobei $p : A \rightarrow \mathbb{R}$ wiederum eine nichtnegative $\mathcal{B}(A_i)$ -meßbare Funktion bezeichnet. Dieser Zusammenhang

$$l_i(f) = \int_{A_i} f \cdot p(\cdot - t_i) d\lambda^2$$

wird normalerweise statt Definition 4.1.2 als Definition einer Aufnahme verwendet. In dieser Arbeit werden in Kapitel 6.2 und Kapitel 7 aber auch Punktmaße als Sensorcharakteristiken zugelassen, was durch eine derartige Definition nicht abgedeckt wäre.

In der Regel sind die Aperturen von homogenen Sensoren paarweise disjunkt und im Rechteck neben- und untereinander angeordnet, wie es in Abbildung 4.2 bereits skizziert wurde. Ein Beispiel dafür, wie sich Aufnahmen derselben Sensorinputfunktion bei

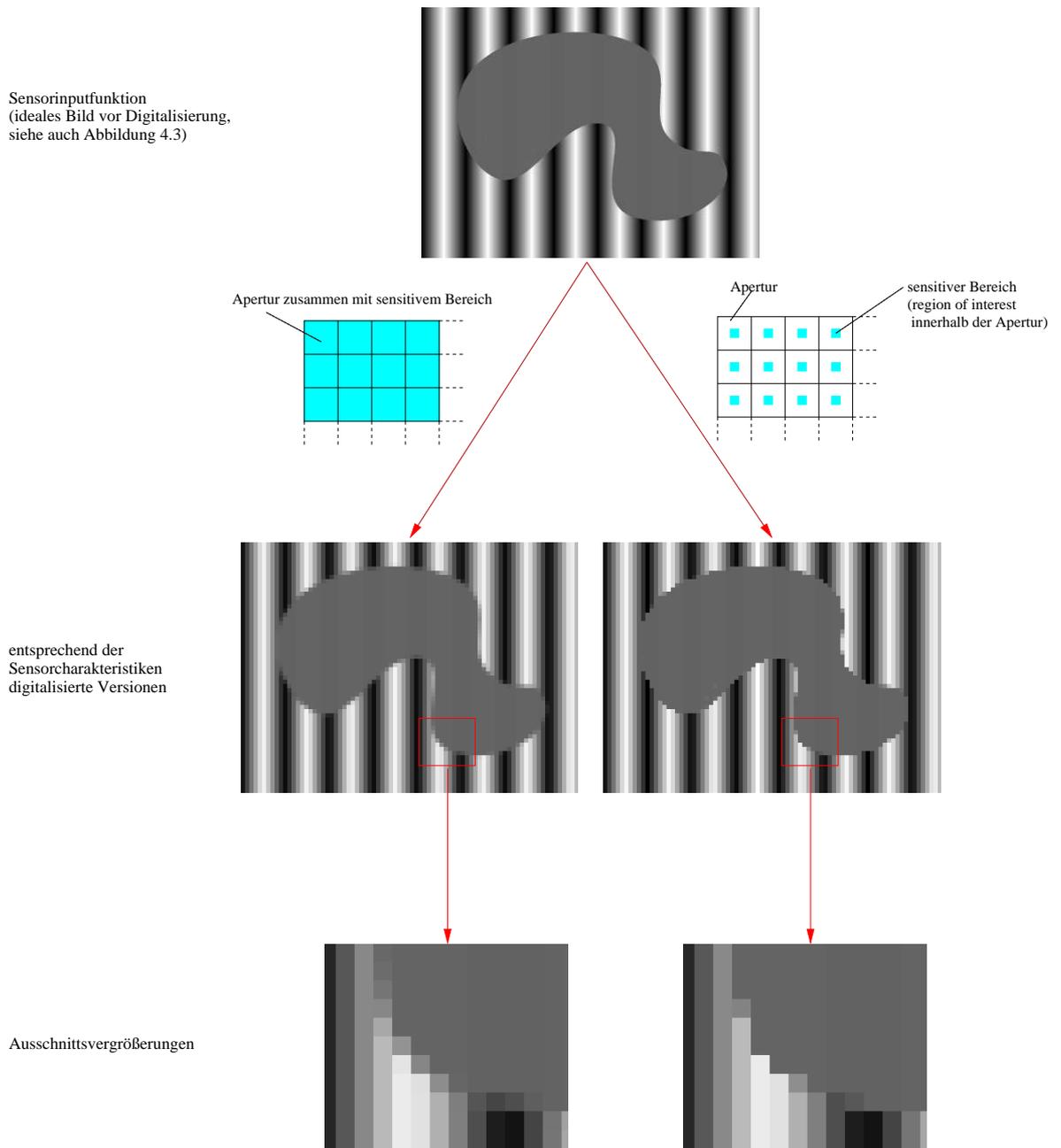


Abbildung 4.4: Verschiedene Sensorkonfigurationen mit entsprechender Digitalisierung des Originalbildes. Links wird durch $\text{sup}(\mu) = A$ (hier als *sensitiver Bereich* bezeichnet) für alle Aperturen A die gesamte Sensorinputfunktion abgetastet. Rechts dagegen stellt der sensitive Bereich eine echte Teilmenge der Aperturen dar. Es wird daher nur ein signifikant kleinerer Teil des Bildes von den Sensorelementen „gesehen“. Die Unterschiede in beiden Aufnahmen sind vor allem im Randbereich des Objektes sichtbar (Ausschnittsvergrößerungen unten).

der Verwendung von Sensoren mit unterschiedlichen Sensorcharakteristiken unterscheiden können, zeigt Abbildung 4.4. Die beiden Digitalisierungen der Sensorinputfunktion in Abbildung 4.4 unterscheiden sich nur in ihren Sensorcharakteristiken. Links wurde für jede Apertur A_i des Sensors $(A_i, \mu_i)_{i \in I}$ als zugehörige Sensorcharakteristik $\mu_i = \lambda_{A_i}^2$ gewählt. Damit gilt $\text{sup}(\mu_i) = A_i$, d.h. durch den Sensor wird das gesamte Bild abgetastet. Bezeichnet nun $(A'_i, \mu'_i)_{i \in I}$ den in der rechten Digitalisierung verwendeten Sensor, gilt für die Sensorcharakteristik einer Apertur A'_i jetzt $\mu'_i = \mathbb{1}_{B_i} \lambda^2$, wobei B_i eine echte Teilmenge von A'_i darstellt. In diesem Fall wird durch den vorliegenden Sensor nur ein Teil jeder Apertur abgetastet. Eine solche Modellierung entspricht eher der Realität einer CCD-Kamera, bei der die lichtempfindliche Fläche ebenfalls nur einen Teil der Apertur abdeckt (siehe dazu auch [Haa99]). Die unterschiedlichen Digitalisierungsergebnisse in Abbildung 4.4 sind vor allem in den Vergrößerungen in Abbildung 4.4 unten deutlich sichtbar. Die Unterschiede treten hauptsächlich dort auf, wo große Grauwertunterschiede im Originalbild existieren — also vor allem im Randbereich des Objektes.

Derartige Überlegungen über die Art der Sensorcharakteristiken können natürlich noch weiter getrieben werden: Eine naheliegende Idee wäre etwa, jeder Sensorcharakteristik μ_i eines Sensorelementes eine Gewichtung über seine Apertur A_i — zum Beispiel in Form einer Normalverteilung — zuzuweisen. Man sollte sich allerdings darüber im Klaren sein, daß in der Regel keine genauere Kenntnis darüber vorhanden ist, wie die in einer Kamera vorliegenden Sensorcharakteristiken ein einfallendes Bildsignal wirklich gewichten. Allenfalls kann man den Träger $\text{sup}(\mu_i)$ jeder Apertur A_i über den Bauplan der Kamera bestimmen. Eine zusätzliche Gewichtung der Sensorcharakteristiken ist daher zwar ein faszinierendes Thema, wird aber aus den eben genannten Gründen hier nicht weiter verfolgt.

4.2 Das Linsensystem

In der Modellierung von CMOS- oder CCD-Kameras wird das Blenden- und Linsensystem in der Regel mit berücksichtigt: Die Lichtstrahlen passieren auf dem Weg vom Objekt zum Sensorarray ein Linsensystem, das für die korrekte Fokussierung des Objektes benötigt wird und treffen erst dann auf das Sensorarray (Abbildung 4.5). Die Anwendungen in den Kapiteln 6 und 7 beinhalten jedoch kein Linsensystem, so daß hier nur grundsätzliche Überlegungen zum Zusammenhang von Sensormodellierung und Linsensystem vorgestellt werden.

Typische Störungen, die durch das Linsensystem hervorgerufen werden können (Kissen- oder Tonnenverzerrungen), könnten zunächst durch eine geeignete Kamerakalibrierung eliminiert werden (siehe dazu zum Beispiel [Tsa87] oder [SA96]). Auf dem so kalibrierten Bild kann dann mit einem einfachen Sensormodell weitergearbeitet werden.

Denkbar ist im Rahmen dieser Arbeit auch eine andere Vorgehensweise: Statt eine Kalibrierung durchzuführen kann die Kameraverzerrung direkt in die Erzeugendenfunktionen der verwendeten semilinearen Funktion mit eingearbeitet werden. Dies führt zu

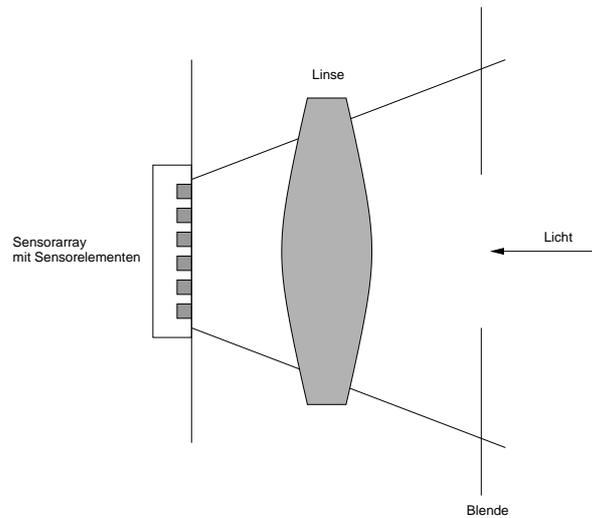


Abbildung 4.5: Schematische Darstellung des Linsensystems einer Kamera zur Verdeutlichung des Bildentstehungsprozesses

Modellierungen, in denen jedes ursprünglich vorliegende Objekt (jedes Element $A \in \mathcal{A}$ gemäß Def. 3.1.1) durch die Erzeugendenfunktionen gemäß der vorliegenden Linsenoptik verzerrt wird. Für jedes Paar $(x, A) \in \mathbb{R}^k \times \mathcal{A}$ weist dann das Bild $g(x, A)$ dieselbe Verzerrung wie das vorliegende (mit unkalibrierter Optik aufgenommene) Bild auf. Diese Vorgehensweise läßt das ursprüngliche Bild unverändert, was in manchen Fällen von Vorteil sein kann. Ein Beispiel für eine solche Vorgehensweise ist die *Grauwertmodelladaption* von Kapitel 2.4, in der sämtliche Effekte im Bildentstehungsprozeß in der point-spread function (PSF) zusammengefaßt werden. Für eine Einbindung dieser PSF in eine Erzeugendenfunktion sei auf Kapitel 6.2 verwiesen.

Ob nun eine Kamerakalibrierung oder eine Einbindung der Verzerrung in die Erzeugendenfunktionen gewählt wird, beide Möglichkeiten hängen von der Güte der Kalibrier- bzw. Verzerrungsfunktion ab. Nur wenn die Kameraverzerrungen ausreichend gut korrigiert wurden (bzw. entsprechende Verzerrungsfunktionen gefunden wurden), ist ein weiteres Arbeiten überhaupt möglich.

4.3 Einbeziehung von Sensorarrays in semilineare Funktionen

Die in Kapitel 4.1 definierten Sensoren lassen sich gut in die semilineare Theorie von Kapitel 3 integrieren: Sei dazu $(A_i, \mu_i)_{i \in I}$ ein Sensor gemäß Definition 4.1.1. Beobachtet wird dann eine Aufnahme $(y_i)_{i \in I}$, die als Digitalisierung $(l_i(f_i))_{i \in I}$ entsprechender Sensorinputfunktionen $(f_i)_{i \in I}$ interpretiert wird. Eine erste informelle Formulierung eines Approximationsproblems — zunächst ohne semilinearem Kontext — lautet daher:

Gesucht wird eine Familie $(f_i)_{i \in I}$ von Sensorinputfunktionen, so daß die Digitalisierung $(l_i(f_i))_{i \in I}$ der $(f_i)_{i \in I}$ durch den Sensor $(A_i, \mu_i)_{i \in I}$ sich „möglichst wenig“ von der gegebenen Aufnahme $(y_i)_{i \in I}$ unterscheidet.

Sei nun $n = \#I$ die Anzahl der Sensorelemente. Für eine Formulierung des vorliegenden Problems als semilineares Approximationsproblem wird zunächst der euklidische Vektorraum E aus Definition 3.2.1 als $E = \mathbb{R}^n$ zusammen mit dem Standardskalarprodukt gewählt. E entspricht dem Beobachtungsraum, der die Menge der Digitalisierungen möglicher Sensorinputfunktionen enthält. In diesem Sinne wird im folgenden auch eine Aufnahme $(y_i)_{i \in I}$ als Vektor $y = (y_1, \dots, y_n)' \in \mathbb{R}^n = E$ und eine Familie $(f_i)_{i \in I}$ von Sensorinputfunktionen als Funktionenvektor

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix} \in L_{\mu_1}^1(A_1, \mathbb{R}) \times \dots \times L_{\mu_n}^1(A_n, \mathbb{R})$$

aufgefaßt. Dieser Funktionenvektor wird ebenfalls als Sensorinputfunktion bezeichnet. Statt

$$\begin{pmatrix} l_1(f_1) \\ \vdots \\ l_n(f_n) \end{pmatrix}$$

wird auch $l(f)$ geschrieben.

Bemerkung 4.3.1

Die lineare Anreihung der Elemente in der Indexmenge I wirkt in der Praxis etwas künstlich, wenn reale (zweidimensionale) Bilder betrachtet werden. Gemäß dieser Indexmenge müssen die Pixel dieser Bilder als Einträge in einem Vektor aufgefaßt werden, was ihrer Bildstruktur eher abträglich ist. Allerdings ist diese lineare Anordnung nur in der Approximationsphase von Bedeutung. Daher kann in der Modellierungsphase eine passendere Indexmenge gewählt werden, solange die eigentliche lineare Anreihung im Hinterkopf behalten wird. Dies geschieht auch in Kapitel 6, wo zwar Bilder aus $\mathbb{R}^{m \times n}$ betrachtet werden und die Indexmenge auch entsprechend gewählt wurde, aber eigentlich $\mathbb{R}^{m \cdot n}$ verwendet wird.

Gesucht wird zunächst der minimale Abstand einer gegebenen Aufnahme $y \in E$ an eine Digitalisierung $(l_1(f_1), \dots, l_n(f_n))' \in E$. Die Sensorinputfunktionen f_1, \dots, f_n , sind dabei unbekannt. Es soll also der Abstand

$$\|y - l(f)\| = \sqrt{\sum_{i=1}^n (y_i - l_i(f_i))^2} \quad (4.1)$$

minimiert werden.

(4.1) stellt zwar ein Approximationsproblem, aber bis jetzt noch kein *semilineares* Approximationsproblem dar. Dieser Zusammenhang wird erst durch die Formulierung der

Sensorinputfunktionen über semilineare Abbildungen hergestellt, was Thema der folgenden Annahme ist:

Annahme 4.3.2

Jeder Vektor $(f_1, \dots, f_n)' \in L^1_{\mu_1}(A_1, \mathbb{R}) \times \dots \times L^1_{\mu_n}(A_n, \mathbb{R})$ von Sensorinputfunktionen eines gegebenen Sensors $(A_i, \mu_i)_{i \in I}$, $\#I = n$, besteht aus mehreren Teilfunktionen, die sich additiv überlagern:

$$f := \sum_{\ell=1}^k x_\ell g^{(\ell)} = \sum_{\ell=1}^k x_\ell \begin{pmatrix} g_1^{(\ell)} \\ \vdots \\ g_n^{(\ell)} \end{pmatrix},$$

wobei die Teilfunktionen $g^{(\ell)}$ für $\ell = 1, \dots, k$ ebenfalls Sensorinputfunktionen des Sensors $(A_i, \mu_i)_{i \in I}$ sind.

Da die Funktionale $(l_i)_{i \in I}$ linear sind, folgt aus Annahme 4.3.2 für die Sensorinputfunktion f :

$$l(f) = l\left(\sum_{\ell=1}^k x_\ell g^{(\ell)}\right) = \sum_{\ell=1}^k x_\ell l(g^{(\ell)})$$

und damit aus (4.1)

$$\|y - l(f)\| = \left\| y - \sum_{\ell=1}^k x_\ell l(g^{(\ell)}) \right\| = \sqrt{\sum_{i=1}^n \left(y_i - \sum_{\ell=1}^k x_\ell l_i(g_i^{(\ell)}) \right)^2}.$$

Für $\mathcal{F} \subseteq L^1_{\mu_1}(A_1, \mathbb{R}) \times \dots \times L^1_{\mu_n}(A_n, \mathbb{R})$ ist nun die Funktion $g : \mathbb{R}^k \times \mathcal{F}^k \rightarrow \mathbb{R}^n$,

$$g(x, g^{(1)}, \dots, g^{(k)}) := \sum_{\ell=1}^k x_\ell l(g^{(\ell)})$$

semilinear nach Definition 3.1.1. Damit werden für ein $y \in E = \mathbb{R}^n$ alle Voraussetzungen von Definition 3.2.1 erfüllt und

$$\inf_{(g^{(1)}, \dots, g^{(k)})' \in \mathcal{F}^k} \min_{x \in \mathbb{R}^k} \left\| y - \sum_{\ell=1}^k x_\ell l(g^{(\ell)}) \right\| \tag{4.2}$$

stellt ein semilineares Approximationsproblem dar.

Durch (4.2) wird jetzt ein semilineares Approximationsproblem beschrieben, wobei $E = \mathbb{R}^n$ sowie $\mathcal{A} = \mathcal{F}^k$ gilt und über die k Erzeugendenfunktionen $g^{(i)}$ optimiert wird. Zusätzlich basiert (4.2) auf einem Sensormodell. Allerdings ist die in (4.2) verwendete Menge \mathcal{F} (bzw. \mathcal{F}^k) der Sensorinputfunktionen ein zu komplexes Gebilde, um in einer Approximation sinnvoll eingesetzt werden zu können. Leichter handhabbar und daher besser geeignet wären abstraktere Mengensysteme wie sie etwa in den Beispielen in Kapitel

3 verwendet wurden. Soll eine derartige Menge \mathcal{A} verwendet werden, werden statt der Sensorinputfunktionen $g^{(1)}, \dots, g^{(k)} \in \mathcal{F}$ nun k Einbettungen $g_e^{(\ell)}$ benötigt, die von \mathcal{A} in die Menge \mathcal{F} der Sensorinputfunktionen abbilden. Zusammen mit diesen Einbettungen $g_e^{(\ell)} : \mathcal{A} \rightarrow \mathcal{F}$ sehen die nun verwendeten Erzeugendenfunktionen $e^{(\ell)}$ schließlich wie folgt aus:

$$e^{(\ell)} : \mathcal{A} \rightarrow \mathbb{R}^n, \quad e^{(\ell)} := l \circ g_e^{(\ell)}, \quad \text{für } \ell = 1, \dots, k.$$

Damit sind die Sensorinputfunktionen im Approximationsprozeß zu einem — mehr oder weniger impliziten — Zwischenschritt auf dem Weg von $A \in \mathcal{A}$ zum betrachteten digitalisierten Bild von A in E geworden. Dies entspricht dem, was bereits in Bemerkung 4.1.3 über die Sensorinputfunktionen gesagt wurde: Ausgehend von einer Menge $A \in \mathcal{A}$ wird letztendlich nur noch die bereits digitalisierte Version (das digitale Bild eines Szenarios) $l(g^{(\ell)}(A)) \in \mathbb{R}^n$ betrachtet. Das semilineare Approximationsproblem unter Einbeziehung eines Sensormodells hat nun für ein gegebenes $y \in E$ und unter Verwendung eines abstrakten Mengensystems die Form

$$\inf_{A \in \mathcal{A}} \min_{x \in \mathbb{R}^k} \left\| y - \sum_{i=1}^k x_i \left(\ell_i \circ g_i^{(\ell)} \right) (A) \right\|.$$

Das folgende Kapitel beschäftigt sich mit einer möglichen Wahl dieses abstrakten Mengensystems als Teilmenge von \mathbb{R}^2 zusammen mit einer geeigneten Metrik.

5 Ein mögliches Mengenmodell

Die Beispiele aus Kapitel 3 lassen bereits ahnen, daß bei der Modellierung eines gegebenen Problems durch semilineare Funktionen ein breites Spektrum an Möglichkeiten zur Verfügung steht. Für den Einsatz in der Praxis dürfen diese Modellierungen aber nicht allzu komplex werden, damit der zum Approximationsproblem gehörige Optimierungsprozeß in einer nicht allzu langen Zeitspanne zu einem (lokalen) Optimum führen kann (vgl. Algorithmus 1 von Seite 39). Natürlich sollte außerdem das so gefundene lokale Optimum nicht „allzu weit“ von dem gesuchten globalen Optimum entfernt sein, denn nur so kann ein realitätsnahes Ergebnis garantiert werden. Diese Ziele können — und sollten — zum einen durch die Wahl eines günstigen Startwertes erreicht werden, zum anderen sollte Vorwissen über die Form des erwarteten Optimums genutzt werden, um aus der Menge \mathcal{A} aus Definition 3.1.1 „unerwünschte“ Elemente so weit wie möglich auszuschließen.

Ein günstiger Startwert $A_0 \in \mathcal{A}$, von dem aus ein semilinearer Optimierungsprozeß starten kann, kann zum Beispiel durch eine Vorsegmentierung durch einen anderen Algorithmus gefunden werden. So ist etwa die ungefähre Größe und Position einer Kreisscheibe in einem Bild über eine pixelbasierte Kantenextraktion oder eine Zwei-Niveau-Passung segmentierbar. Derartige pixelbasierte Verfahren benötigen in der Regel keine lange Rechenzeit und die so gewonnenen Resultate können meist gut als Ausgangsbasis verwendet werden.

Die zweite oben erwähnte Option ist die konkrete Wahl von \mathcal{A} . Natürlich sind die „erwünschten“ bzw. „unerwünschten“ Elemente von \mathcal{A} je nach vorliegendem Problem verschieden, daher fließt in die Wahl von \mathcal{A} mehr oder weniger automatisch Vorwissen über die konkrete Art des Problems mit ein. In diesem Kapitel wird $\mathcal{A} \subseteq \mathfrak{A}(\mathbb{R}^d)$, $d \in \mathbb{N}$ als mögliches Mengenmodell¹ vorgeschlagen.

5.1 Die lokale Hausdorff-Topologie

In praktischen Anwendungen wird normalerweise ein kompakter Ausschnitt M des \mathbb{R}^2 bzw. des \mathbb{R}^3 betrachtet. Das dabei verwendete Mengensystem — der metrische Raum (\mathcal{A}, d) in der semilinearen Approximation — kann dann über M definiert werden. Dies ist in den Beispielen aus Kapitel 3.3 auch so geschehen. Ist also ein kompakter Ausschnitt

¹Im folgenden wird \mathcal{A} auch *Mengenmodell* oder *Mengensystem* genannt, wenn \mathcal{A} Teilmengen des \mathbb{R}^d bezeichnet.

$M \subseteq \mathbb{R}^d$ für $d \in \mathbb{N}$ gegeben, ist die Wahl von $\mathcal{A} \subseteq \mathcal{B}(M)/\sim$ zusammen mit der durch das Lebesgue-Maß induzierten Metrik $d(A, B) = \lambda^n((A \setminus B) \cup (B \setminus A))$ auf \mathcal{A} eine Möglichkeit. Eine alternative Möglichkeit stellt die Hausdorff-Metrik dar:

Sei (X, δ) ein nicht-leerer metrischer Raum und sei $\mathfrak{K}(X)^* := \mathfrak{K}(X) \setminus \{\emptyset\}$ das System aller kompakten nicht-leeren Teilmengen von X . Dann bildet die Abbildung

$$\mathfrak{h} : \mathfrak{K}(X)^* \times \mathfrak{K}(X)^* \rightarrow \mathbb{R}, \quad \mathfrak{h}(A, B) := \max \left\{ \sup_{x \in A} \text{dist}(x, B), \sup_{y \in B} \text{dist}(A, y) \right\}$$

eine Metrik auf $\mathfrak{K}(X)^*$, die *Hausdorff-Metrik*. Ist X kompakt, gilt für alle $A, B \in \mathfrak{K}(X)^*$

$$\mathfrak{h}(A, B) \leq \text{diam}(X)$$

wobei $\text{diam}(X) := \sup\{\delta(x, y) \mid x, y \in X\}$ den Durchmesser von X bezeichnet. Die Hausdorff-Metrik kann in diesem Fall von $\mathfrak{K}(X)^*$ auf ganz $\mathfrak{K}(X)$ durch

$$\mathfrak{h}'(A, B) := \begin{cases} \mathfrak{h}(A, B) & A \neq \emptyset \text{ und } B \neq \emptyset \\ 0 & A = \emptyset \text{ und } B = \emptyset \\ \text{diam}(X) & \text{sonst} \end{cases}$$

erweitert werden.

Je nach Art der konkreten Anwendung eignet sich eine der beiden Metriken eher als die andere. Eine dritte Möglichkeit zur Wahl des metrischen Raumes wird im folgenden vorgestellt. Sie eignet sich vor allem dann, wenn die zu findenden Objekte auch aus dem Aufnahmeraum M herausragen können. In diesem Fall sind die auftretenden Mengen in ihrer Größe nicht mehr durch M nach oben beschränkt.

Als Beispiel dazu betrachte man das in Abbildung 5.1 skizzierte System: Sei für $r > 0$

$$C'_r := \{(x, y) \mid (x - r)^2 + y^2 \leq r^2\}$$

ein (in \mathbb{R}^2 abgeschlossener) Kreis mit Radius r und Zentrum $(r, 0)$ und sei weiterhin

$$\mathcal{A}_{circ} := \{C'_r \mid r \geq 0\}$$

die Menge aller Kreise dieser Art. Dann ist \mathcal{A}_{circ} offen in $\mathfrak{K}(\mathbb{R}^2)$ (bzgl. \mathfrak{h}). Der für den Abschluß benötigte Halbraum

$$H := \{(x, y) \mid x \geq 0\}$$

kann jedoch als nicht-kompakte Teilmenge des \mathbb{R}^2 nicht von der Hausdorff-Metrik erfaßt werden.

Es ist zwar auch in diesem Fall möglich, ein zu \mathcal{A}_{circ} äquivalentes Mengensystem als Teilmenge von $\mathfrak{K}(M)$ zu definieren, indem man den Durchschnitt

$$\mathcal{A}_{circ} \cap M := \{A \cap M \mid A \in \mathcal{A}_{circ}\}$$

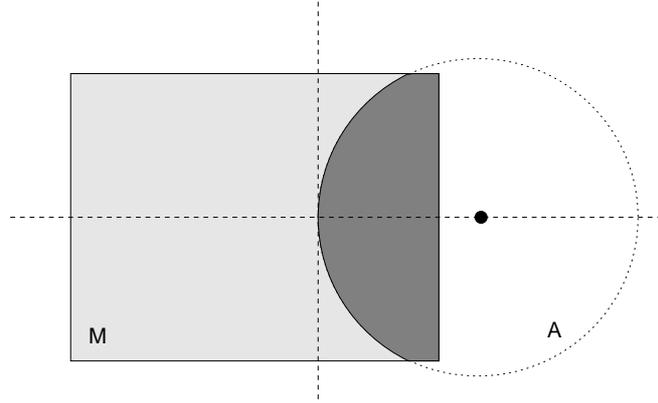


Abbildung 5.1: Aufnahmeraum (oder Kameraausschnitt) M mit Kreisscheibe A , die nur teilweise in M sichtbar ist.

betrachtet. Die folgende Definition der in [Don97] entwickelten lokalen Hausdorff-Topologie — einer metrisierbaren Topologie auf $\mathfrak{A}(\mathbb{R}^d)$ — zeigt aber, daß dieser Schritt unnötig ist und man sich auf die Betrachtung alleine von $\mathcal{A}_{\text{circ}}$ ohne Berücksichtigung von M zurückziehen kann.

Sei dazu (X, δ) ein lokal-kompakter, σ -kompakter, metrischer Raum und $(K_i)_{i \in \mathbb{N}}$ eine Folge kompakter Teilmengen von X , für die

$$\forall i \in \mathbb{N} : K_i \subseteq \overset{\circ}{K}_{i+1} \quad \text{und} \quad \bigcup_{i \in \mathbb{N}} \overset{\circ}{K}_i = X \quad (5.1)$$

gilt. Die Existenz einer solchen Folge $(K_i)_{i \in \mathbb{N}}$ ist durch die σ -Kompaktheit von X sichergestellt.

Sei nun $\mathfrak{A}(X)^* := \mathfrak{A}(X) \setminus \{\emptyset\}$ das System aller abgeschlossenen nicht-leeren Teilmengen von X . Dann läßt sich für jedes $i \in \mathbb{N}$ die Abbildung $\mathfrak{d}_i : \mathfrak{A}(X)^* \times \mathfrak{A}(X)^* \rightarrow \mathbb{R}$ durch

$$\forall A, B \in \mathfrak{A}(X)^* : \mathfrak{d}_i(A, B) := \max \left\{ \sup_{x \in A} \text{dist} \left(x, B \cup \overline{X \setminus K_i} \right), \sup_{y \in B} \text{dist} \left(y, A \cup \overline{X \setminus K_i} \right) \right\}$$

definieren. Die Distanzfunktion ist dabei durch

$$\forall A, B \subseteq X : \text{dist}(A, B) = \begin{cases} \inf_{x \in A} \inf_{y \in B} \delta(x, y), & \text{falls } A \neq \emptyset \text{ und } B \neq \emptyset \\ \infty, & \text{sonst} \end{cases}$$

gegeben. Nach [Don97] gilt dann für die \mathfrak{d}_i :

- $\mathfrak{d}_i(A, B) = \mathfrak{d}_i(B, A)$
- $\mathfrak{d}_i(A, A) = 0$
- $\mathfrak{d}_i(A, C) \leq \mathfrak{d}_i(A, B) + \mathfrak{d}_i(B, C)$

für alle $A, B, C \in \mathfrak{A}(X)^*$ und für beliebiges $i \in \mathbb{N}$. Weil aber für $A, B \in \mathfrak{A}(X)^*$ bereits $\mathfrak{d}_i(A, B) = 0$ ist, wenn nur $A \cap K_i = B \cap K_i$ zutrifft, ist die für Metriken zusätzlich geforderte Bedingung

$$\forall A, B \in \mathfrak{A}(X)^* : \mathfrak{d}_i(A, B) = 0 \Leftrightarrow A = B$$

hier nicht erfüllt. Bei den Funktionen \mathfrak{d}_i handelt es sich daher nur um sogenannte *Quasi-metriken*.

Allerdings definiert die Folge $(\mathfrak{d}_i)_{i \in \mathbb{N}}$ gemäß [Don97] eine lokal-kompakte Topologie auf $\mathfrak{A}(X)^*$, die mit Hilfe der Einpunkt-Kompaktifizierung nach Alexandroff zu einer Topologie auf ganz $\mathfrak{A}(X)$ ausgebaut werden kann, bzgl. derer $\mathfrak{A}(X)$ kompakt ist. Ist X (wie hier nach Voraussetzung) lokal-kompakt, ist diese Topologie metrisierbar und $\mathfrak{A}(X)$ wird zu einem kompakten metrischen Raum. Diese Topologie wird *lokale Hausdorff-Topologie* (LH-Topologie) genannt.

Für die durch

$$\forall A, B \in \mathfrak{A}(X) : \tau_i(A, B) = \max \left\{ \sup_{x \in A \cap K_i} \text{dist}(x, B), \sup_{y \in B \cap K_i} \text{dist}(y, A) \right\}$$

definierten Abbildungen $\tau_i : \mathfrak{A}(X) \times \mathfrak{A}(X) \rightarrow \mathbb{R} \cup \{\infty, -\infty\}$ ($i \in \mathbb{N}$) wird in [Don97] außerdem nachgewiesen, daß mit

$$\mathfrak{V}(A) := \left\{ \left\{ B \in \mathfrak{A}(X)^* \mid \tau_i(A, B) \leq \varepsilon \right\} \mid i \in \mathbb{N}, \varepsilon > 0 \right\}$$

ein fundamentales Umgebungssystem für $A \in \mathfrak{A}(X)^*$ bzgl. der LH-Topologie gegeben ist. $\mathfrak{V}(A)$ enthält also nur Umgebungen von A und jede Umgebung $\mathfrak{U} \subseteq \mathfrak{A}(X)^*$ von A ist die Obermenge eines Elements von $\mathfrak{V}(A)$.

Wird für jedes $i \in \mathbb{N}$ die Menge $U_i \subseteq \mathfrak{A}(X)$ als

$$U_i = \{A \in \mathfrak{A}(X) \mid A \cap K_i = \emptyset\}$$

gewählt, ist $\{U_i \mid i \in \mathbb{N}\}$ darüber hinaus ein fundamentales Umgebungssystem der leeren Menge bzgl. der LH-Topologie.

Im in dieser Arbeit interessanten Fall $X = \mathbb{R}^d$ ($d \in \mathbb{N}$) zusammen mit der euklidischen Metrik δ ist X lokal-kompakt und σ -kompakt, und damit ist jede abgeschlossene Teilmenge $\mathcal{A} \subseteq \mathfrak{A}(\mathbb{R}^d)$ ebenfalls kompakt (bzgl. der LH-Topologie). Im Folgenden werden dafür die Abkürzungen LH-abgeschlossen bzw. LH-kompakt verwendet.

Bezüglich der lokalen Hausdorff-Topologie ist dann die Menge

$$\bar{\mathcal{A}}_{circ} = \{C'_r \mid r \geq 0\} \cup H$$

abgeschlossen in $\mathfrak{A}(\mathbb{R}^2)$. Als Beweis betrachte man die Abbildung $h : \mathcal{A}_{circ} \cup H \rightarrow \mathbb{R} \cup \{\infty\}$, die durch

$$h(A) := \begin{cases} r, & \text{falls } A = C'_r \text{ für ein } r \geq 0 \\ \infty, & \text{falls } A = H \end{cases}$$

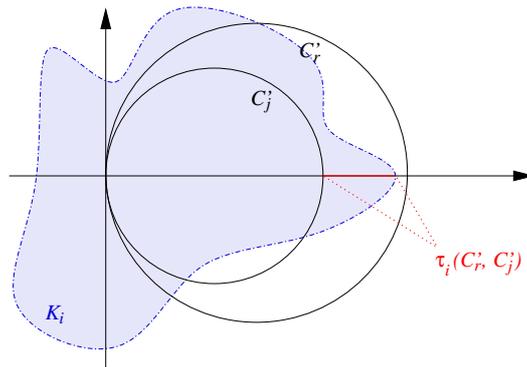


Abbildung 5.2: Wert von $\tau_i(C'_r, C'_j)$ für $j < r$

definiert ist. Wie leicht zu sehen ist, ist $\bar{\mathcal{A}}_{circ}$ das Urbild von $[0, \infty]$ unter h . Das bedeutet, bei stetigem h ist $\bar{\mathcal{A}}_{circ}$ ebenfalls abgeschlossen.

Die Stetigkeit von h zeigt man, indem man die Stetigkeit an jedem Punkt durch die Verwendung fundamentaler Umgebungssysteme sicherstellt: Für $r > 0$ ist

$$\{[r - \varepsilon, r + \varepsilon] \mid \varepsilon > 0\}$$

ein fundamentales Umgebungssystem von r . Sei also $\varepsilon > 0$ und sei $[r - \varepsilon, r + \varepsilon]$ eine Umgebung von r . Dann setzt man zunächst

$$V_{r,\varepsilon} := \left\{ C'_j \mid j \in \left[r - \frac{\varepsilon}{2}, r + \frac{\varepsilon}{2} \right] \right\}.$$

$V_{r,\varepsilon}$ ist Element von $\mathfrak{B}(K'_r)$, da für alle $i \in \mathbb{N}$

$$\tau_i(C'_r, C'_j) \leq |2r - 2j| \leq \varepsilon$$

gilt, Abbildung 5.2 verdeutlicht dies. Für das Bild von $V_{r,\varepsilon}$ unter h gilt weiterhin

$$h(V_{r,\varepsilon}) = \left[r - \frac{\varepsilon}{2}, r + \frac{\varepsilon}{2} \right] \subseteq [r - \varepsilon, r + \varepsilon],$$

was die Stetigkeit von h an jedem Punkt C'_r , $r > 0$ impliziert. Für $r = 0$ wählt man die Umgebungen $[0, \varepsilon]$ und

$$\left\{ K'_j \mid j \in \left[0, \frac{\varepsilon}{2} \right] \right\} \in \mathfrak{B}(K'_0).$$

Die Stetigkeit von h folgt dann mit analogen Argumenten.

Bis jetzt konnte der Stetigkeitsbeweis von h geführt werden, ohne daß auf die Folge $(K_i)_{i \in \mathbb{N}}$ kompakter Teilmengen von \mathbb{R}^2 zurückgegriffen werden mußte. H ist jedoch nicht beschränkt in \mathbb{R}^2 und daher sind die Abbildungen τ_i abhängig von der konkreten Wahl der Folge $(K_i)_{i \in \mathbb{N}}$.

Im folgenden wird daher

$$K_i := [-i, i]^2$$

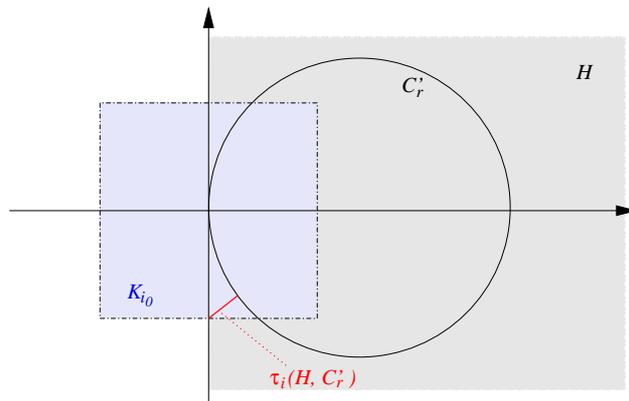


Abbildung 5.3: Berechnung von $\tau_i(H, C'_r)$ für $K_i \subseteq [-1, 1] \times [-1, 1]$.

gewählt. Die so definierte Folge $([-i, i]^2)_{i \in \mathbb{N}}$ erfüllt die Voraussetzungen von (5.1) auf Seite 69: Für jedes $i \in \mathbb{N}$ ist $[-i, i]^2$ eine kompakte Teilmenge des \mathbb{R}^2 und es gilt

$$[-i, i]^2 \subseteq]-i-1, i+1[^2 \quad \text{sowie} \quad \bigcup_{i \in \mathbb{N}}]-i, i[^2 = \mathbb{R}^2.$$

Sei nun $[r_0, \infty]$ für $r_0 \in \mathbb{R}$ eine Umgebung von ∞ . Dann existiert für jedes $i_0 \in \mathbb{N}$ ein $\varepsilon_0 > 0$, so daß

$$\tau_{i_0}(H, C'_{r_0}) \leq \varepsilon_0$$

gilt. Aus $\tau_{i_0}(H, C'_{r_0}) \leq \varepsilon_0$ folgt aber automatisch

$$\tau_{i_0}(H, C'_j) \leq \varepsilon_0 \quad \text{für alle } j \geq r_0.$$

Damit ist

$$\{C'_j \mid j \geq r_0\}$$

ein Element von $\mathfrak{V}(H)$ und das Bild von $\{C'_j \mid j \geq r_0\}$ unter h ist eine Teilmenge von $[r_0, \infty]$. Die Abbildung h ist also auch im Punkt H stetig.

Der Beweis zur Stetigkeit von h im Punkt H verdeutlicht zusammen mit Abbildung 5.3 einen wichtigen Aspekt der lokalen Hausdorff-Topologie, der auch intuitiv leicht zu erfassen ist: Ist eine gegebene Menge „zu groß“ um sie ganz erfassen zu können, werden statt der Gesamtmenge nur kompakte Ausschnitte von ihr betrachtet. Diese Sichtweise kommt der Modellierung eines Kameraausschnittes sehr entgegen, denn auch hier kann ein gegebenes Objekt größer als der aktuell betrachtete Ausschnitt sein, aber die Betrachtung reduziert sich auf den innerhalb des Ausschnittes liegenden Bereichs.

Ein LH-kompaktes Mengensystem wird in Kapitel 6.1 verwendet.

6 Subpixelgenaue Objektpassung auf Grauwertbildern

Ein aktuelles Problem in der MR-Angiographie ist die Segmentierung von Blutgefäßen aus den gemessenen Angiogrammen. Probleme machen dabei vor allem die Gefäße, deren Querschnitt nur wenig größer als die Bildauflösung ist. Eine diskrete Segmentierung solcher Gefäße ist zwar möglich, jedoch schwankt die Anzahl von derart segmentierten Gefäßpixeln von Aufnahme zu Aufnahme in der Regel viel zu stark, um eine verlässliche Aussage über den wahren Gefäßquerschnitt zu erhalten. Ein Beispiel dafür, wie unterschiedlich diskrete Segmentierungen des quasi gleichen Ausgangssachverhalts ausfallen können, zeigt Abbildung 6.1: Eine Kreisscheibe mit festem Radius wird zweimal digitalisiert. Dabei wird angenommen, daß der Kreis einen weißen, der Hintergrund dagegen einen schwarzen Grauwert besitzt. Als der Digitalisierung zugrundeliegender Sensor wird $(A_{ij}, \lambda_{A_{ij}}^2)_{i,j \in \{1, \dots, 9\}}$ verwendet. A_{ij} bezeichnet dabei das Bildquadrat in Zeile i und Spalte j . Im ersten Fall (links in Abbildung 6.1) liegt das Kreiszentrum direkt in einer Pixelmitte, was zu einer recht regelmäßigen Diskretisierung führt. Im zweiten Fall dagegen wurde das Kreiszentrum etwas aus der Pixelmitte verschoben und dementsprechend unregelmäßiger ist auch die dazugehörige Digitalisierung. Insgesamt sind die Unterschiede in beiden Digitalisierungen bis jetzt aber nicht sonderlich signifikant. Ganz anders sieht die Lage aber aus, wenn man eine diskrete Segmentierung auf den beiden digitalisierten Bildern durchführt. In Abbildung 6.1 ist dies durch eine Zwei-Niveau-Passung mit zugrundeliegendem allgemeinem Mengenmodell $\mathcal{A} = \mathfrak{P}(M)$ geschehen, wobei $M = \{A_{ij} \subseteq \mathbb{R}^2 \mid i, j = 1 \dots, 9\}$ die Menge der Bildpixel bezeichnet. Die beiden Paßergebnisse links bzw. rechts unten zeigen nicht nur signifikant unterschiedliche Strukturen, auch die Anzahl der gefundenen Pixel variiert von der linken Passung (25 Pixel) zur rechten Passung (28 Pixel) erheblich. Stellt man sich beide Ergebnisse als aufeinanderfolgende Segmentierungen in einem Gefäßverlauf vor, ist klar, daß sie so in der Praxis kaum verwendet werden können.

Geometrische Verfahren zur Kreispassung, wie sie zum Beispiel in [ARR99], [AR99] oder auch [FPF99] beschrieben werden, lassen sich für derartige Probleme nicht verwenden, da diese stets auf einer Punktmenge $(p_i)_{i \in I}$, $p_i \in \mathbb{R}^2$ aufsetzen. Diese Punkte $(p_i)_{i \in I}$ werden als mögliche Randpunkte einer geometrischen Struktur (wie zum Beispiel einer Kreisscheibe) aufgefaßt. Bei dem hier vorgestellten Problem kann ein Pixel aber einen ganzen Randbereich beinhalten, außerdem ist a-priori keine Information darüber vorhanden, welche Pixel überhaupt als Randpixel dienen könnten.

In Kapitel 6.1 wird ein Ansatz zur subpixelgenauen Objektpassung über eine semilineare Approximation vorgestellt. Dieses Verfahren basiert auf einer Sensorintegration gemäß

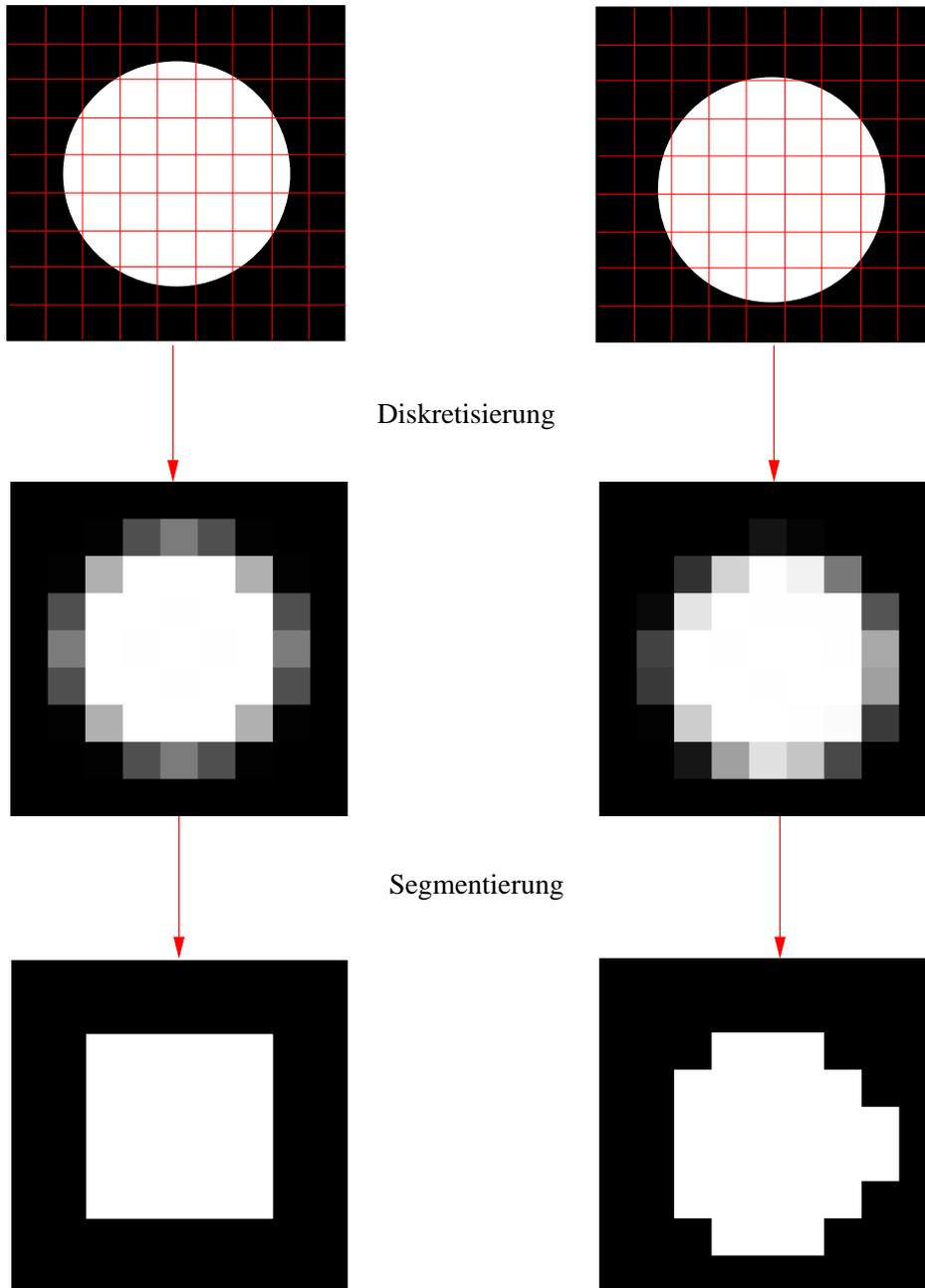


Abbildung 6.1: Zwei Diskretisierungen von weißen Kreisen auf schwarzem Hintergrund. Der Radius beider Kreise ist gleich, das Kreiszentrum im rechten oberen Bild gegenüber dem Bild links oben etwas verschoben. Die Segmentierung der Kreispixel wurde mit einer Zwei-Niveau-Passung mit $\mathcal{A} = \mathfrak{P}(M)$ durchgeführt, wobei M die diskrete Pixelmenge bezeichnet. Obwohl die Ausgangssituation nahezu identisch ist, zeigen die beiden diskreten Paßergebnisse signifikante Abweichungen.

Kapitel 4 und benutzt ein einfach handhabbares Sensormodell. Ein weiterer Ansatz zur subpixelgenauen Objektpassung auf Grauwertbildern ist die in Kapitel 2.4 vorgestellte Grauwertmodelladaption. Dieses Verfahren ist zwar primär zur Erkennung von Kanaldeckeln aus Luftbildern entwickelt worden, aber Kanaldeckel und Gefäßquerschnitte unterscheiden sich in der gegebenen Ansicht (helle Kreisscheibe auf dunklem Hintergrund in grober Auflösung) nur wenig voneinander. Kapitel 6.2 befaßt sich daher mit diesem Verfahren und stellt eine Weiterentwicklung vor, mit deren Hilfe sich die Grauwertmodelladaption ebenfalls in den semilinearen Kontext einordnet. Beide Verfahren werden in Kapitel 6.3 an MR-Bildern von Schlauchquerschnitten getestet. Die Segmentierungsergebnisse bilden den Schluß dieses Kapitels.

6.1 Ein semilinearer Ansatz zur subpixelgenauen Objektpassung

Im folgenden wird ein Verfahren zur subpixelgenauen Objektpassung vorgestellt, bei dem die Subpixelstruktur über eine Sensorintegration gemäß Kapitel 4 erreicht wird. Die Modellierung folgt dabei der Beschreibung aus Kapitel 4.3. Die wesentliche Idee dieses Verfahrens wurde bereits in [LP04] vorab vorgestellt.

Gegeben sei ein Sensor $(A_{kl}, \mu_{kl})_{(k,l) \in I}$, wobei die Indexmenge I die Form

$$I = \{(k, l) \in \mathbb{Z} \times \mathbb{Z} \mid 1 \leq k \leq m, 1 \leq l \leq n\}$$

für vorgegebene Bilddimensionen $m, n \in \mathbb{N}$ besitzt. Jedes Bild $(a_{kl})_{(k,l) \in I}$ kann dann mit den Bezeichnungen von Kapitel 4.3 als Matrix

$$a = \begin{pmatrix} l_{11}(f_{11}) & \cdots & l_{1n}(f_{1n}) \\ \vdots & \ddots & \vdots \\ l_{m1}(f_{m1}) & \cdots & l_{mn}(f_{mn}) \end{pmatrix} \in \mathbb{R}^{m \times n}$$

aufgefaßt werden, wobei die einzelnen Einträge die Grauwerte der entsprechenden Bildpixel wiedergeben. Gemäß Kapitel 4.3 bezeichnet f_{kl} die auf der Apertur A_{kl} definierte Sensorinputfunktion und $l_{kl}(f_{kl})$ die Digitalisierung bzw. Aufnahme von f_{kl} wiederum auf A_{kl} .

Als Skalarprodukt auf $\mathbb{R}^{m \times n}$ wird das Standardskalarprodukt auf $\mathbb{R}^{m \cdot n}$

$$\langle a|b \rangle := \sum_{(k,l) \in I} a_{kl} b_{kl}$$

verwendet. Ein zweidimensionales Bild wird damit nicht als Matrix sondern als eindimensionaler Vektor in $\mathbb{R}^{m \cdot n}$ aufgefaßt, nur seine Darstellung erfolgt in Matrizenform (siehe dazu auch Bemerkung 4.3.1 auf Seite 64).

Die genaue Form der Aperturen A_{kl} des Sensors ist an dieser Stelle noch nicht wichtig. In der Regel wird davon ausgegangen, daß sie eine Partition eines kompakten Aufnahme-
 raumes $M \subseteq \mathbb{R}^2$ bilden wie es typischerweise in einem CCD-Array der Fall ist. In
 den meisten Fällen sind die Aperturen zusätzlich rechteckig bzw. sogar quadratisch und
 damit identisch zu dem entsprechenden Bildpixel.

Gesucht wird zunächst eine passende semilineare Funktion $g : \mathbb{R}^\ell \times \mathcal{A} \rightarrow \mathbb{R}^{m \times n}$ für ein
 geeignetes $\ell \in \mathbb{N}$. Mit \mathcal{A} wird dabei eine LH-abgeschlossene Teilmenge von $\mathfrak{A}(\mathbb{R}^2)$ be-
 zeichnet, die konkrete Spezifikation von \mathcal{A} beschreibt dann die gewünschten zu passenden
 Mengen.

Im Sinne von Kapitel 4.3 werden zunächst ℓ Sensorinputfunktionen $g^{(1)}, \dots, g^{(\ell)}$ benö-
 tigt. Da nur einfache Abhängigkeiten (Pixel liegt in Menge $A \in \mathcal{A}$ vs. Pixel liegt nicht
 in A) von Bedeutung sind, reichen die durch

$$\begin{aligned} g_{kl}^{(1)}(A) &:= \mathbb{1}_{A_{kl} \cap A} \quad \text{und} \\ g_{kl}^{(2)}(A) &:= \mathbb{1}_{A_{kl} \setminus A} \quad \text{für } k = 1, \dots, m \text{ sowie } l = 1, \dots, n \end{aligned}$$

definierten Sensorinputfunktionen $g^{(1)}$ und $g^{(2)}$ für die Modellierung aus. $g^{(1)}$ repräsen-
 tiert die gegebene Menge (in Kapitel 6.3 dann die Kreisscheibe) und $g^{(2)}$ den sichtbaren
 Resthintergrund.

Ist nun $A \in \mathcal{A}$ gegeben, hat die Digitalisierung von $g_{kl}^{(1)}(A)$ auf der Apertur A_{kl} die Form

$$l_{kl}(g_{kl}^{(1)}(A)) = l_{kl}(\mathbb{1}_{A_{kl} \cap A}) = \int_{A_{kl}} \mathbb{1}_{A_{kl} \cap A} d\mu_{kl} = \mu_{kl}(A \cap A_{kl}).$$

Analog gilt für die Digitalisierung von $g_{kl}^{(2)}(A)$

$$l_{kl}(g_{kl}^{(2)}(A)) = \int_{A_{kl}} \mathbb{1}_{A_{kl} \setminus A} d\mu_{kl} = \mu_{kl}(A \setminus A_{kl}).$$

Auf diese Weise erhält man die folgenden Erzeugendenfunktionen $e_1, e_2 : \mathcal{A} \rightarrow \mathbb{R}^{m \times n}$:

$$\begin{aligned} e_1(A)_{kl} &= \mu_{kl}(A_{kl} \cap A) \quad (\text{Anteil von } A \text{ in } A_{kl}) \\ e_2(A)_{kl} &= \mu_{kl}(A_{kl} \setminus A) \quad (\text{Anteil des Komplements von } A \text{ in } A_{kl}) \end{aligned}$$

für alle $(k, l) \in I$. Die zugehörige semilineare Funktion $g : \mathbb{R}^2 \times \mathcal{A} \rightarrow E$,

$$\begin{aligned} g(x, A) &= x_1 \begin{pmatrix} \mu_{11}(A \cap A_{11}) & \cdots & \mu_{1n}(A \cap A_{1n}) \\ \vdots & \ddots & \vdots \\ \mu_{m1}(A \cap A_{m1}) & \cdots & \mu_{mn}(A \cap A_{mn}) \end{pmatrix} \\ &+ x_2 \begin{pmatrix} \mu_{11}(A \setminus A_{11}) & \cdots & \mu_{1n}(A \setminus A_{1n}) \\ \vdots & \ddots & \vdots \\ \mu_{m1}(A \setminus A_{m1}) & \cdots & \mu_{mn}(A \setminus A_{mn}) \end{pmatrix} \end{aligned}$$

beschreibt dann Linearkombinationen von Digitalisierungen der Funktionen $g^{(1)}$ und $g^{(2)}$ entsprechend dem gegebenen Sensor $(\mu_i, A_i)_{i \in I}$.

Insgesamt ergibt sich für ein gegebenes Bild $(a_{kl})_{(k,l) \in I}$ damit das folgende Approximationsproblem

$$\inf_{A \in \mathcal{A}} \min_{x \in \mathbb{R}^2} \sum_{(k,l) \in I} (a_{kl} - x_1 \cdot \mu_{kl}(A_{kl} \cap A) - x_2 \cdot \mu_{kl}(A_{kl} \setminus A))^2,$$

bzw. äquivalent dazu

$$\inf_{A \in \mathcal{A}} \|a - \text{proj}_{\text{linH}\{e_1(A), e_2(A)\}}\| = \inf_{A \in \mathcal{A}} \psi_{g,a}. \quad (6.1)$$

Dabei bezeichnet $\psi_{g,a} : \mathcal{A} \rightarrow E$ die in (3.1) von Seite 39 definierte Zielfunktion eines semilinearen Approximationsproblems.

Diese Modellierung ist allgemein in dem Sinne, daß die konkrete Objektform nur durch \mathcal{A} bestimmt wird. So kann auf einfache Weise die Objektform ausgetauscht werden ohne daß die gesamte Modellierung verändert werden muß. Allerdings werden sowohl Objekt als auch Hintergrund in dieser Modellierung durch einen konstanten Grauwert repräsentiert. Kompliziertere Grauwertabhängigkeiten, wie sie etwa in Beispiel 3.1.6 verwendet wurden, bedürfen einer Modellerweiterung. (6.1) kann als *subpixelgenaue Objektpassung* bezeichnet werden.

Bemerkung 6.1.1

Sensorinput- und Erzeugendenfunktionen sollten bei dieser Anwendung nicht durcheinander geworfen werden. Zwar werden Objekt und Hintergrund zunächst durch Indikatorfunktionen modelliert, aber daraus folgt nicht, daß auch die Erzeugendenfunktionen konstant sind. Die Wahl von konstanten Erzeugendenfunktionen — also Indikatorfunktionen — würde im wesentlichen in einer Variante der Zwei-Niveau-Passung (Kapitel 2.1) resultieren. Dies ist hier jedoch nicht der Fall, denn die Erzeugendenfunktionen in der subpixelgenauen Objektpassung repräsentieren nicht den Grauwert einer gegebenen Menge A im Bild, sondern den Anteil dieser Menge innerhalb eines Pixels und besitzen damit einen völlig anderen Kontext als die Indikatorfunktionen der Zwei-Niveau-Passung.

Um die subpixelgenaue Objektpassung auf die vorhandenen Testbilder anwenden zu können, müssen konkrete Sensor- und Mengenmodelle bestimmt werden, wobei beide nicht unnötig kompliziert sein sollten. Das folgende Sensormodell, das die Aperturen mit den Bildpixeln identifiziert, erwies sich als geeignet. Sei dazu für $(k, l) \in I$

$$\llbracket k, l \rrbracket := \left[k - \frac{1}{2}, k + \frac{1}{2} \right] \times \left[l - \frac{1}{2}, l + \frac{1}{2} \right]$$

das quadratische Bildpixel mit Mittelpunkt (k, l) und

$$\lambda_{kl}^2 := \lambda^2(\cdot \cap \llbracket k, l \rrbracket)$$

das zweidimensionale Lebesgue-Maß auf dem Pixel $\llbracket k, l \rrbracket$. Das Sensormodell hat dann die Form

$$(\llbracket k, l \rrbracket, \lambda_{k,l}^2)_{(k,l)=(0,0),\dots,(m,n)}.$$

Dieses Sensormodell kann auch als Standardsensormodell angesehen werden und wurde in dieser Arbeit bereits mehrfach zur Bildgenerierung verwendet (Abbildung 6.1 auf Seite 74 oder Abbildung 2.3 auf Seite 20).

Als Mengenmodell bietet sich der Abschluß $\bar{\mathcal{A}}$ bzgl. der lokalen Hausdorff-Topologie der Menge

$$\mathcal{A} := \{A \subseteq \mathbb{R}^2 \mid \exists x, y \in \mathbb{R}, r \geq 0 : A = C_r^l(x, y)\}$$

an. \mathcal{A} enthält alle (bzgl. der euklid. Metrik) abgeschlossenen Kreisscheiben im \mathbb{R}^2 . $\bar{\mathcal{A}}$ ist (nach Definition) eine LH-abgeschlossene Teilmenge von $\mathfrak{A}(\mathbb{R}^2)$ und damit LH-kompakt. Dieses Mengensystem deckt auch die Fälle ab, in denen die zu passende Kreisscheibe nur ausschnittsweise im Bild sichtbar ist und/oder in denen die Kreisscheibe zum Halbraum entartet. Implementationstechnisch müssen die auftretenden Halbräume natürlich gesondert behandelt werden. Zusammen mit diesem Sensor- und Mengenmodell wird die subpixelgenaue Objektpassung im folgenden als *Kreisscheibenpassung* bezeichnet.

Die Erzeugendenfunktionen von Seite 6.1 haben für $(k, l) \in I$ und $A \in \bar{\mathcal{A}}$ die Form

$$\begin{aligned} e_1(A)_{kl} &:= \lambda^2(\llbracket k, l \rrbracket \cap A) && \text{(Flächenanteil von } A \text{ im Pixel } \llbracket k, l \rrbracket) \\ e_2(A)_{kl} &:= \lambda^2(\llbracket k, l \rrbracket \setminus A) && \text{(Flächenanteil des Hintergrundes ohne } A \text{ in } \llbracket k, l \rrbracket). \end{aligned}$$

e_1 und e_2 sind stetig. Da jedes Pixel Fläche 1 besitzt, gilt zusätzlich

$$e_2(A)_{kl} = 1 - e_1(A)_{kl}$$

für alle $(k, l) \in I$.

Es bleibt nun, die Stetigkeit der Zielfunktion $\psi_{g,a}$ aus (6.1) zu überprüfen. Durch die Form der Erzeugendenfunktionen gilt $\text{linH}\{e_1(A), e_2(A)\} = 2$ für alle Mengen (Kreise oder Halbräume) $A \in \bar{\mathcal{A}}$, für die $\lambda^2(A) \neq 0$ sowie $\lambda^2(\llbracket i, j \rrbracket \cap A) \neq 0$ für mindestens ein $(i, j) \in I$ gilt. In diesen Fällen ist $\psi_{g,a}$ stetig. Durch die spezielle Form der beiden Erzeugendenfunktionen gilt in diesen Fällen zusätzlich

$$0 < \sum_{(k,l) \in I} e_1(A)_{kl} < m \cdot n.$$

Bleiben die Mengen $A \in \bar{\mathcal{A}}$ übrig, für die $\text{linH}\{e_1(A), e_2(A)\} = 1$ gilt. Diese Fälle führen im rekonstruierten Bild $g(x, A)$ zu gleichen Werten für alle $(i, j) \in I$ unabhängig von $x \in \mathbb{R}^2$. Das Verhalten von $\psi_{g,a}$ an einer derartigen Übergangsstelle hin zu einem konstanten Bild kann am besten anhand des folgenden Szenarios verdeutlicht werden¹:

¹Läßt man die λ^2 -Nullmengen außer acht, wird anschaulich in einem solchen Fall der gesamte Bildausschnitt durch A entweder vollständig überdeckt oder aber überhaupt nicht berührt.

Sei eine Kreisscheibe $C'_r(x, y) \in \mathcal{A}$ gegeben. Das Kreiszentrum (x, y) liege dabei im Inneren eines Pixels $\llbracket k, l \rrbracket$. Für $r \rightarrow 0$ zeigt dann die normierte Version

$$e'_1(C'_r(x, y)) := \frac{e_1(C'_r(x, y))}{\|e_1(C'_r(x, y))\|}$$

von e_1 das folgende Verhalten: Da das Kreiszentrum im Inneren von $\llbracket k, l \rrbracket$ liegt, existiert ein $\varepsilon_0 > 0$, so daß $C'_\varepsilon(x, y) \subseteq \llbracket k, l \rrbracket$ für alle $\varepsilon < \varepsilon_0$ gilt. Da die Erzeugendenfunktion e_1 den Kreisanteil in den Bildpixeln repräsentiert, folgt daraus

$$e_1(C'_\varepsilon(x, y))_{ij} = 0$$

für alle $(i, j) \in I$, $(i, j) \neq (k, l)$, Für die normierte Version e'_1 gilt zusätzlich

$$e'_1(C'_\varepsilon(x, y))_{kl} = 1.$$

Daraus folgt daß e'_1 konstant ist für alle $\varepsilon > 0$, $\varepsilon < \varepsilon_0$. Ist dagegen $\varepsilon = 0$ folgt daraus $\lambda^2(C'_0(x, y)) = 0$ und somit auch $e'_1(C'_0(x, y))_{ij} = 0$ für alle $(i, j) \in I$. Das bedeutet, daß e'_1 an dieser Stelle unstetig ist. Diese Situation ist im Prinzip identisch zu der in Beispiel 3.3.8 beschriebenen Situation und daher ist auch hier — wie in Beispiel 3.3.8 — eine Unstetigkeitsstelle in der Zielfunktion zu erwarten. Die folgenden Überlegungen bestätigen dies: Laut Satz 3.3.9 kann ein semilineares Approximationsproblem statt über die Erzeugendenfunktionen auch über seine Projektionsoperatoren beschrieben werden. Für $0 < \varepsilon < \varepsilon_0$ hat die zu $C'_\varepsilon(x, y)$ gehörende Projektionsmatrix die Form

$$\mathbf{P}_{C'_\varepsilon(x, y)} = \begin{pmatrix} & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ & & & 0 & & & \\ & & & \vdots & & & \\ & & & 0 & & & \end{pmatrix} \in \mathbb{R}^{m \cdot n \times m \cdot n},$$

wobei die 1 in der Mitte den (k, l) -ten Eintrag der Matrix bezeichnet. Diese Matrix bleibt konstant für alle $\varepsilon > 0$, $\varepsilon < \varepsilon_0$. Die einpunktige Menge $C'_0(x, y)$ dagegen entspricht der Matrix \mathbf{P}_\emptyset , die in allen Einträgen den Wert 1 besitzt. Damit ist die Abbildung $A \mapsto \mathbf{P}_A$ (und mit Satz 3.3.10 dann auch die Zielfunktion $\psi_{g, a}$) im Punkt $A = \emptyset$ tatsächlich unstetig².

Analoge Argumente lassen sich für sämtliche Übergänge hin zu einem konstanten Bild finden.

Die hier vorgestellte Kreisscheibenpassung ist ein Verfahren, das ohne großen Aufwand implementiert werden kann. Die Berechnung der Zielfunktion $\psi_{g, a}$ benötigt zudem nicht

²Die Voraussetzungen von Satz 3.3.10(1) sind erfüllt, solange das Bild $(a_{ij})_{ij \in I}$ selber nicht-konstant ist, d.h. gleiche Werte in allen a_{ij} , $(i, j) \in I$ besitzt.

viel Rechenzeit (mehr dazu am Ende dieses Kapitels). Die einzige Einschränkung in der Verwendbarkeit sind die oben erwähnten Unstetigkeitsstellen, wenn die Bilddimension in sich zusammenfällt. Daher dürfen nur Mengen, die den Kameraausschnitt nur teilweise überdecken, im Mengensystem berücksichtigt werden. Das Mengensystem $\mathcal{A}' \subseteq \mathcal{A}$, das dann noch übrigbleibt hat die Form

$$\mathcal{A}' = \left\{ A \in \bar{\mathcal{A}} \mid 0 < \sum_{(k,l) \in I} e_1(A)_{kl} < m \cdot n \right\}.$$

Eine weitere Einschränkung von \mathcal{A}' hin zu einer LH-kompakten Menge sollte je nach vorliegender Situation geschehen.

6.2 Die Grauwertmodelladaption und ihr Zusammenhang zur semilinearen Approximation

In Kapitel 2.4 wurde mit der Grauwertmodelladaption der in [Dre98] entwickelte Ansatz zur Objektlokalisierung unter Berücksichtigung eines subpixelgenauen Objektmodells vorgestellt. Die Modellbildung folgte bei diesem Verfahren im Wesentlichen den folgenden Punkten: Ein ideales stufenförmiges Signal wird durch eine *point spread function* (PSF) verschmiert und schließlich durch einen Scanner an diskreten Punkten abgetastet. Die PSF wird dabei durch eine normalverteilte Funktion mit Mittelwert 0 und Standardabweichung $\sigma > 0$ modelliert. σ ist ein freier Parameter des Modells und wird in der konkreten Anwendung mit optimiert.

Das mit der PSF gefaltete Signal für eine vertikale Stufe der Höhe h_1 über einem Grundniveau h_0 ist dann gegeben durch

$$M_{\text{step}}(x, y) = h_0 + h_1 \cdot \phi\left(\frac{x - x_0}{\sigma}\right)$$

(siehe dazu auch Kapitel 2.4). Freie Parameter dieses Modells sind die Standardabweichung σ der PSF, die (Subpixel-)position $x_0 \in \mathbb{R}$ der Stufe sowie die beiden Stufenniveaus h_0 und h_1 .

Am Beispiel von M_{step} kann man bereits erkennen, daß wohl ein Zusammenhang zur semilinearen Modellierung aus Kapitel 3 vorhanden sein müßte, denn hält man die Parameter x_0 und σ fest, ist M_{step} linear in h_0 und h_1 . Es bleibt daher, die Voraussetzungen von Definition 3.1.1 zu überprüfen und M_{step} gegebenenfalls anzupassen. Als Allererstes muß dazu die Frage der Mengenmodellierung geklärt werden. Es ist klar, daß die Modellfunktionen in der Grauwertmodelladaption auf Mengenmodellen basieren, auch wenn diese nur implizit über ihre Parametrisierung erwähnt werden. Für das Modell M_{step} einer vertikalen Stufe bietet sich

$$\mathcal{A} := \{(\sigma, x_0) \mid \sigma > 0, x_0 \in \mathbb{R}\}$$

an. \mathcal{A} repräsentiert damit durch $x_0 \in \mathbb{R}$ die in M_{step} verwendeten Halbräume. Zusätzlich wird der Grad der Verschmierung bei der Bildberechnung als freier Parameter benötigt und daher in \mathcal{A} mit aufgenommen. Man beachte, daß \mathcal{A} in der vorliegenden Form nicht kompakt ist, überhaupt werden Kompaktheitsfragen wie auch Terminierungsfragen des Abstiegsverfahrens in [Dre98] nicht behandelt (mehr dazu bei der Modellierung von Kreisscheiben weiter hinten in diesem Abschnitt). Als Vektorraum E bietet sich der Raum $L^2(M, \mathbb{R})$ der quadratintegrierbaren Funktionen über einem kompakten Grundraum $M \subseteq \mathbb{R}^2$ an. Dieser Grundraum M , der als Aufnahmebereich oder Kameraausschnitt angesehen werden kann, ist in [Dre98] ebenfalls nicht explizit erwähnt worden, eine Einschränkung auf M ist aber nötig um Quadratintegrierbarkeit für die Modellfunktionen zu erreichen³.

Unter diesen Bedingungen bildet die Modellfunktion

$$M_{\text{step}}^{(s)} : \mathbb{R}^2 \times \mathcal{A} \rightarrow L^2(M, \mathbb{R}), \quad M_{\text{step}}^{(s)}(h_0, h_1, (\sigma, x_0)) := h_0 \mathbb{1}_M + h_1 \cdot \phi\left(\frac{\cdot - x_0}{\sigma}\right)$$

eine semilineare Funktion nach Definition 3.1.1.

Um die so definierten Funktionen in ein semilineares Approximationsproblem nach Definition 3.2.1 einzubinden, wird auf $L^2(M, \mathbb{R})$ ein Skalarprodukt benötigt. Dieses Skalarprodukt muß zusätzlich so gewählt sein, daß sich schlußendlich die in der Grauwertmodelladaption verwendete Zielfunktion $\mathcal{E} : \mathbb{R}^2 \times \mathcal{A} \rightarrow \mathbb{R}$,

$$\mathcal{E}(h_0, h_1, A) := \sum_{(k,l)=(1,1)}^{(m,n)} \left[a_{kl} - M_{\text{step}}^{(s)}(h_0, h_1, (\sigma, x_0))(k, l) \right]^2 \quad (6.2)$$

ergibt, wobei mit $a = (a_{kl})_{\substack{k=1, \dots, m \\ l=1, \dots, n}}$ ein gegebenes diskretes Grauwertbild bezeichnet wird. In dieser Zielfunktion \mathcal{E} steckt die Auswertung der Modellfunktionen an bestimmten diskreten Bildpunkten $(k, l) \in \{1, \dots, m\} \times \{1, \dots, n\}$. Zum Einbinden in eine semilineare Approximation müßte daher das Skalarprodukt auf $L^2(M, \mathbb{R})$ durch

$$s : L^2(M, \mathbb{R}) \times L^2(M, \mathbb{R}) \rightarrow \mathbb{R}, \quad s(f, g) := \sum_{(k,l)=(1,1)}^{(m,n)} f(k, l) \cdot g(k, l)$$

definiert werden. Allerdings wird durch $\sqrt{s(\cdot, \cdot)}$ keine Norm auf $L^2(M, \mathbb{R})$ definiert, denn für jede Funktion $f \in L^2(M, \mathbb{R})$ mit

$$f(k, l) = 0 \quad \forall (k, l) \in \{(1, 1), \dots, (m, n)\}$$

ist $\sqrt{s(f, f)} = 0$, obwohl f nicht notwendigerweise die Nullfunktion ist.

Eine mögliche Lösung dieses Problems wäre die Einschränkung von $L^2(M, \mathbb{R})$ auf eine Teilmenge \mathcal{S} , so daß \mathcal{S} zusammen mit s einen euklidischen Vektorraum bildet. Im folgenden wird eine andere Möglichkeit vorgestellt, die auf einfache Weise diese Einschränkung umgeht und gleichzeitig einen Bezug zum Sensormodell aus Kapitel 4 herstellt:

³Natürlich ist die im Modell M_{step} verwendete Funktion $(x, y) \mapsto 1$ nicht quadratintegrierbar über \mathbb{R}^2 .

In [Dre98] und auch bisher in diesem Abschnitt haben die Modellfunktionen in den Funktionenraum $L^2(M, \mathbb{R})$ abgebildet und wurden erst in der Fehlerfunktion \mathcal{E} diskretisiert. Im Gegensatz dazu wurde in Kapitel 4.3 das Sensormodell aus Definition 4.1.1 verwendet, mit dessen Hilfe bereits die Erzeugendenfunktionen in diskrete Räume abbilden.

Diese Vorgehensweise soll hier ebenfalls angewendet werden. Dazu muß zuerst ein passendes Sensormodell gewählt werden. Das folgende Modell

$$(A_{kl}, \delta_{kl})_{\substack{k=1, \dots, m, \\ l=1, \dots, n}},$$

wobei δ_{kl} das Dirac-Maß im Punkt (k, l) bezeichnet, bietet sich dabei an, denn das diskrete Bild entsteht in dem in der Grauwertmodelladaptation verwendeten Bildaufnahme-prozeß durch Auswertung an diskreten Punkten.

Für die Aperturen A_{kl} muß

$$(k, l) \in A_{kl} \quad \forall (k, l) \in \{(1, 1), \dots, (m, n)\}$$

gelten, die weitere Form der Aperturen ist nicht von Bedeutung. Vom Vektorraum $L^2(M, \mathbb{R})$ wird nun auf den diskreten Raum $\mathbb{R}^{m \cdot n}$ zusammen mit dem Standardskalarprodukt gewechselt⁴. Statt der bisherigen Modellfunktion $M_{\text{step}}^{(s)}$ wird jetzt ihre diskrete Variante

$$M_{\text{step}}^{(d)} : \mathbb{R}^2 \times \mathcal{A} \rightarrow \mathbb{R}^{m \times n}, \quad M_{\text{step}}^{(d)}(h_0, h_1, (\sigma, x_0)) = \begin{pmatrix} \kappa(1) & \cdots & \kappa(m) \\ \vdots & \ddots & \vdots \\ \kappa(1) & \cdots & \kappa(m) \end{pmatrix}$$

betrachtet, mit $\kappa(k) := h_0 + h_1 \cdot \phi\left(\frac{k-x_0}{\sigma}\right)$.

Ist nun ein diskretes Grauwertbild $a = (a_{kl})_{\substack{k=1, \dots, m \\ l=1, \dots, n}}$ gegeben, bildet das so erhaltene Optimierungsproblem

$$\inf_{(\sigma, x_0) \in \mathcal{A}} \min_{x \in \mathbb{R}^2} \|a - M_{\text{step}}^{(d)}\|$$

ein semilineares Approximationsproblem nach Definition 3.2.1. Der Ausdruck $\|a - M_{\text{step}}^{(d)}\|$ entspricht dabei der Zielfunktion \mathcal{E} aus (6.2). Eine Optimierung von \mathcal{E} kann daher gemäß der Algorithmusskizze von Seite 39 erfolgen, wobei das konkrete Abstiegsverfahren natürlich noch zu wählen ist. Die wesentliche Neuerung dieses Algorithmus gegenüber der ursprünglichen Grauwertmodelladaptation ist, daß die beiden linearen Anteile h_0 und h_1 nicht mehr als zu optimierende Parameter in das lokale Abstiegsverfahren eingehen, da sie für jede Wahl von $(\sigma, x_0) \in \mathcal{A}$ bereits optimal berechnet werden können. Im folgenden wird diese neue semilineare Variante des Verfahrens als *modifizierte Grauwertmodelladaptation* bezeichnet.

Analoge Überlegungen zur Übertragung der Grauwertmodelladaptation in die semilineare Approximation lassen sich für alle weiteren in [Dre98] vorgestellten Modellfunktionen

⁴Statt $\mathbb{R}^{m \cdot n}$ wird wieder $\mathbb{R}^{m \times n}$ geschrieben, um die Bildstruktur besser zu veranschaulichen.

durchführen. Hier interessiert vor allem die Passung durch Kreisscheiben. In (2.6) auf Seite 28 wurde das Bildsignal einer Kreisscheibe um den Nullpunkt bereits auf ihre Werte entlang der x -Achse reduziert. Analog gilt für einen allgemeinen Kreis mit Radius R und Mittelpunkt (c_x, c_y) im Bildpunkt (x, y)

$$(D_{R,c_x,c_y} * G_\sigma)(x, y) = (D_R * G_\sigma)(\sqrt{(x - c_x)^2 + (y - c_y)^2}, 0).$$

D_{R,c_x,c_y} bezeichnet dabei die Indikatorfunktion über dem nicht-zentrierten Kreis mit Radius R und Mittelpunkt (c_x, c_y) (vgl. Seite 28).

An dieser Stelle ein paar Worte zu \mathcal{A} , das an die Kreisscheibenmodellierung angepaßt werden muß. Zusätzlich zu den Parametern des verwendeten Mengenmodells — hier repräsentiert durch die Kreisparameter Radius R und Zentrum (c_x, c_y) — muß wie beim Stufenmodell der Parameter σ der Gaußverschmierung mit aufgenommen werden. Die einfachste Wahl für \mathcal{A} ist daher

$$\mathcal{A} := \left\{ (\sigma, R, c_x, c_y) \in \mathbb{R}_+^2 \times \mathbb{R}^2 \mid \sigma, R \geq 0, c_x, c_y \in \mathbb{R} \right\}.$$

Da die Modellfunktion $(D_R * G_\sigma)$ nicht von einer Kreisscheibe an sich, sondern nur von ihren Parametern abhängt, wird auf eine Erweiterung von \mathcal{A} hin zu einem LH-kompakten System verzichtet. Allerdings dürfen nun bei einer Passung keine Fälle auftreten, in denen die Kreisscheibe während der Optimierung bzw. im Optimum zu einem Halbraum entarten könnte. Genau diese Fälle sind durch \mathcal{A} nicht abgedeckt. Eine Einschränkung von \mathcal{A} auf eine kompakte Teilmenge ist zur Vermeidung dieser Situation natürlich jederzeit möglich.

Soll zusätzlich zur Kreisscheibe auch der Hintergrund in das Modell mit einbezogen werden, muß dies zusätzlich formuliert werden. Man erhält dann insgesamt zwei Erzeugendenfunktionen $e_1, e_2 : \mathcal{A} \rightarrow \mathbb{R}^{m \times n}$, die durch

$$\begin{aligned} e_1(\sigma, R, c_x, c_y)(k, l) &:= \begin{cases} (D_R * G_\sigma)\left(\sqrt{(k - c_x)^2 + (l - c_y)^2}, 0\right), & R, \sigma \neq 0 \\ D'_R\left(\sqrt{(k - c_x)^2 + (l - c_y)^2}, 0\right), & \sigma = 0, R \neq 0 \\ 0, & \sigma = 0, R = 0 \end{cases} \\ e_2(\sigma, R, c_x, c_y)(k, l) &:= 1 \end{aligned}$$

gegeben sind. Mit D'_R wird dabei eine leicht modifizierte Version von D_R bezeichnet, die durch

$$D'_R(x, y) := \begin{cases} 1, & x^2 + y^2 < R^2 \\ \frac{1}{2}, & x^2 + y^2 = R^2 \\ 0 & \text{sonst} \end{cases}$$

gegeben ist. e_1 ist stetig falls $\sigma \neq 0$ und $R \neq 0$. Gilt $\sigma \neq 0$, ist e_1 im Punkt $R = 0$ ebenfalls stetig, denn aus $R \rightarrow 0$ folgt $e_1(\sigma, R, c_x, c_y) \rightarrow 0$. Setzt man dagegen $R \neq 0$ voraus, konvergiert e_1 für $\sigma \rightarrow 0$ gegen D'_R , denn für $\sigma \neq 0$ gilt zunächst

$$(D_R * G_\sigma)(x, 0) = \int_{-R}^R \int_{-\sqrt{R^2 - a^2}}^{\sqrt{R^2 - a^2}} G_\sigma(x - a, -b) db da$$

$$\begin{aligned}
 &= \int_{-R}^R \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \left(\int_{-\sqrt{R^2-a^2}}^{\sqrt{R^2-a^2}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{b^2}{2\sigma^2}} db \right) da \\
 &= \int_{-R}^R \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}} \left(\int_{-\frac{\sqrt{R^2-a^2}}{\sigma}}^{\frac{\sqrt{R^2-a^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) da \\
 &= \int_{\frac{x-R}{\sigma}}^{\frac{x+R}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \left(\int_{-\frac{\sqrt{R^2-(x-\sigma v)^2}}{\sigma}}^{\frac{\sqrt{R^2-(x-\sigma v)^2}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) dv.
 \end{aligned}$$

Für $\sigma \rightarrow 0$ ergeben sich dann die folgenden Fälle:

$$\lim_{\sigma \rightarrow 0} (D_R * G_\sigma)(x, 0) = \begin{cases} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) dv = 1, & |x| < R \\ \int_{\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) dv = 0, & x > R \\ \int_{-\infty}^{-\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) dv = 0, & x < -R \\ \int_0^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) dv = \frac{1}{2}, & x = R \\ \int_{-\infty}^0 \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} \left(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du \right) dv = \frac{1}{2}, & x = -R. \end{cases}$$

Dies bedeutet, daß e_1 für $\sigma = 0$ stetig durch D'_R fortgesetzt wird. Allerdings ist D'_R eine Stufenfunktion und damit selber nicht stetig. Für $\sigma = 0$ zeigt e_1 also ein unstetiges Verhalten in den restlichen Variablen. Eine Verwendung in der Approximation ist damit im Fall $\sigma = 0$ nicht sehr sinnvoll. Zudem gilt

$$\lim_{R \rightarrow 0} D'_R(x, y) = \begin{cases} \frac{1}{2}, & x = y = 0 \\ 0, & \text{sonst,} \end{cases}$$

d.h. D'_R konvergiert für $R \rightarrow 0$ nicht gegen die Nullfunktion und damit existiert kein stetiger Anschluß zum Fall $\sigma = R = 0$. Die Möglichkeit $\sigma = 0$ sollte daher ausgeschlossen werden. Im folgenden wird $\sigma \neq 0$ vorausgesetzt.

Nun zur Stetigkeit der Zielfunktion: Ist ein Bild $a \in E$ gegeben, hat das zugehörige Optimierungsproblem die Form

$$\inf_{(\sigma, R, c_x, c_y) \in \mathcal{A}'} \min_{x \in \mathbb{R}^2} \sum_{(k,l) \in I} \left(a_{kl} - x_1 \cdot e_1(\sigma, R, c_x, c_y)(k, l) - x_2 \right)^2$$

bzw.

$$\inf_{(\sigma, R, c_x, c_y) \in \mathcal{A}'} \psi_{g,a}(\sigma, R, c_x, c_y) = \inf_{(\sigma, R, c_x, c_y) \in \mathcal{A}'} \left\| a - \text{proj}_{\text{linH}\{e_1(\sigma, R, c_x, c_y), \mathbf{1}\}} \right\|,$$

wobei mit $\mathbf{1}$ die konstante Erzeugendenfunktion e_2 bezeichnet wird. Diese Zielfunktion $\psi_{g,a}$ ist stetig für $R \neq 0$, was sich aus $e_1(\sigma, R, c_x, c_y) \neq 0$ für alle $(\sigma, R, c_x, c_y) \in \mathcal{A}$ mit $R \neq 0$ wie folgt ergibt: Da e_1 (durch die Faltung mit der Gaußfunktion) nicht konstant ist und die Faltung stets alle Pixel beeinflusst, ist $e_1(\sigma, R, c_x, c_y)$ für jede Wahl

von $(\sigma, R, c_x, c_y) \in \mathcal{A}$ mit $R \neq 0$ eine nicht-konstante Funktion. Da e_2 konstant ist, folgt daraus $\dim(\text{linH}\{e_1(\sigma, R, c_x, c_y), 1\}) = 2$.

Bleibt zu prüfen, ob $\psi_{g,a}$ im Übergang zu $\dim(\text{linH}\{e_1(\sigma, R, c_x, c_y), 1\}) = 1$ stetig ist. Das ist die Frage nach der Stetigkeit von $R \rightarrow \psi_{g,a}(\sigma, R, c_x, c_y)$ im Punkt $R = 0$. Dazu betrachte man zunächst die normierte Version e'_1 von e_1 . Sie hat die Form

$$e'_1(\sigma, R, c_x, c_y)_{kl} = \frac{(D_R * G_\sigma)(\sqrt{(k - c_x)^2 + (l - c_y)^2}, 0)}{\left\| (D_R * G_\sigma)(\sqrt{(k - c_x)^2 + (l - c_y)^2}, 0) \right\|} \quad \text{für alle } (k, l) \in I$$

für $R \neq 0$ sowie für $R = 0$ die Form

$$e'_1(\sigma, 0, c_x, c_y)_{kl} = 0 \quad \text{für alle } (k, l) \in I.$$

Für $R \neq 0$ ist nun durch die Normierung die Gesamthelligkeit des durch e'_1 erzeugten Bildes immer gleich, nämlich gleich

$$\sum_{(k,l) \in I} e'^2_1(\sigma, R, c_x, c_y)_{kl} = 1.$$

Dieser Wert ändert sich für $R \rightarrow 0$, $R \neq 0$ nicht. Daraus folgt, daß sich e'_1 nicht der Nullfunktion annähern kann. e'_1 ist damit unstetig im Punkt $R = 0$.

Aus der Unstetigkeit von e'_1 folgt zwar nicht direkt die Unstetigkeit von $\psi_{g,a}$. Da dieses Szenario aber stark denen aus Kapitel 6.1 bzw. Beispiel 3.3.8 ähnelt, ist kaum zu erwarten daß sich die Zielfunktion $\psi_{g,a}$ gerade in diesem Fall stetig verhält. Der Fall $R = 0$ sollte daher — wie vorher schon der Fall $\sigma = 0$ — aus der Optimierung ausgeschlossen werden. Damit ist die ursprüngliche Menge \mathcal{A} von Seite 83 auf

$$\mathcal{A}' = \left\{ (\sigma, R, c_x, c_y) \in \mathbb{R}_+^2 \times \mathbb{R}^2 \mid \sigma, R > 0, c_x, c_y \in \mathbb{R} \right\}$$

zusammengeschmolzen. Eine weitere Einschränkung auf eine kompakte Teilmenge von \mathcal{A}' , um die Existenz eines lokalen Optimums zu garantieren, sollte — wie bei der Kreispassung im vorigen Abschnitt — abhängig von der konkreten Situation geschehen.

6.3 Test beider Verfahren

Im konkreten Anwendungsfall soll der Querschnitt von Blutgefäßen aus MR-Angiogrammen rekonstruiert werden. Die Rekonstruktion dieser Querschnitte ist wichtig zur Erkennung von Stenosen — krankhaften Gefäßverengungen — die unbehandelt zu Infarkten führen können. In [Kol95] und [Kus92] wurde diese Problematik für Gefäße im Kopf-Hals-Bereich bereits behandelt, allerdings beschränkte sich die Segmentation der Gefäße dort auf die Pixelauflösung, eine subpixelgenaue Rekonstruktion fand nicht statt. Es zeigte sich dabei, daß zur Beurteilung des größten Teils der Gefäße eine pixelbasierte

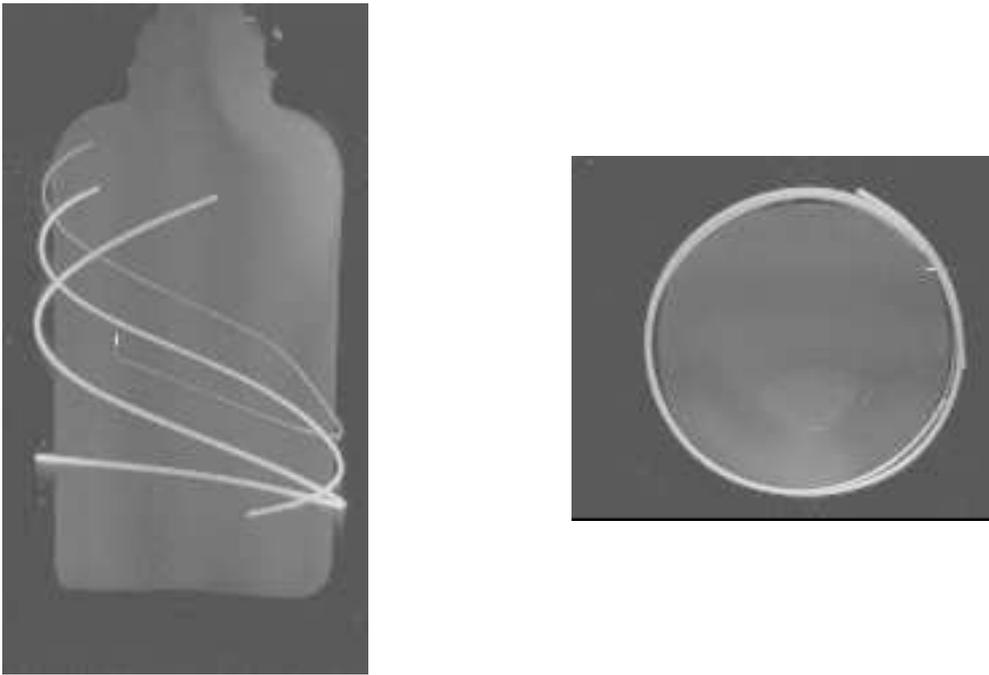


Abbildung 6.2: Maximum-Intensitäts-Projektion des MR-Datenwürfels. Links Seitenansicht, rechts Draufsicht.

Segmentierung ausreicht. Nur Gefäße mit sehr kleinem Querschnitt wurden unzureichend erfaßt. Da aber vor allem diese Gefäße potentielle Kandidaten für Stenosen sind, ist gerade ihre möglichst exakte Rekonstruktion von überdurchschnittlicher Wichtigkeit.

Beide vorgestellte Verfahren (Kreisscheibenpassung und modifizierte Grauwertmodelladaption) wurden an 307 Querschnittsbildern getestet⁵. Bei den Bildern handelt es sich um MR-Angiogramme von mit Kontrastmittel gefüllten Schläuchen, die um ein weiteres Objekt (Flasche) gewickelt wurden. Die Schläuche haben einen festen Durchmesser von 2 mm. Abbildung 6.2 zeigt zwei Ansichten des Datenwürfels in einer Maximum-Intensitäts-Projektion. Man kann deutlich den (hellen) Testschlauch sowie die Flasche, um die er gewickelt wurde, erkennen. Zwei einzelne Schnitte durch den Datenwürfel zeigt Abbildung 6.3.

In der in Abbildung 6.3 vorliegenden Form können diese Daten allerdings noch nicht für eine Passung mit zweidimensionalen Kreisscheiben verwendet werden, denn dazu müßte sichergestellt sein, daß das Bildmaterial stets senkrechte Schnitte durch die Schlauchzentrale zeigt. Dies ist zunächst nicht gegeben — wie vor allem der linke Schnitt in Abbildung 6.3 verdeutlicht, der aus dem Randbereich des Datenwürfels stammt. Aber auch im rechten Schnitt in Abbildung 6.3 aus dem Mittelteil des Datenwürfels ist nicht garantiert, daß die Testschläuche wirklich senkrecht zu den Schnittebenen liegen und somit der Schlauchquerschnitt wirklich kreisförmig ist. Als erste Vorverarbeitung muß

⁵Die ursprüngliche Grauwertmodelladaption wurde nicht mehr als Vergleich hinzugezogen, da ihre Ergebnisse höchstens genauso gut wie die der modifizierten Grauwertmodelladaption sein können.

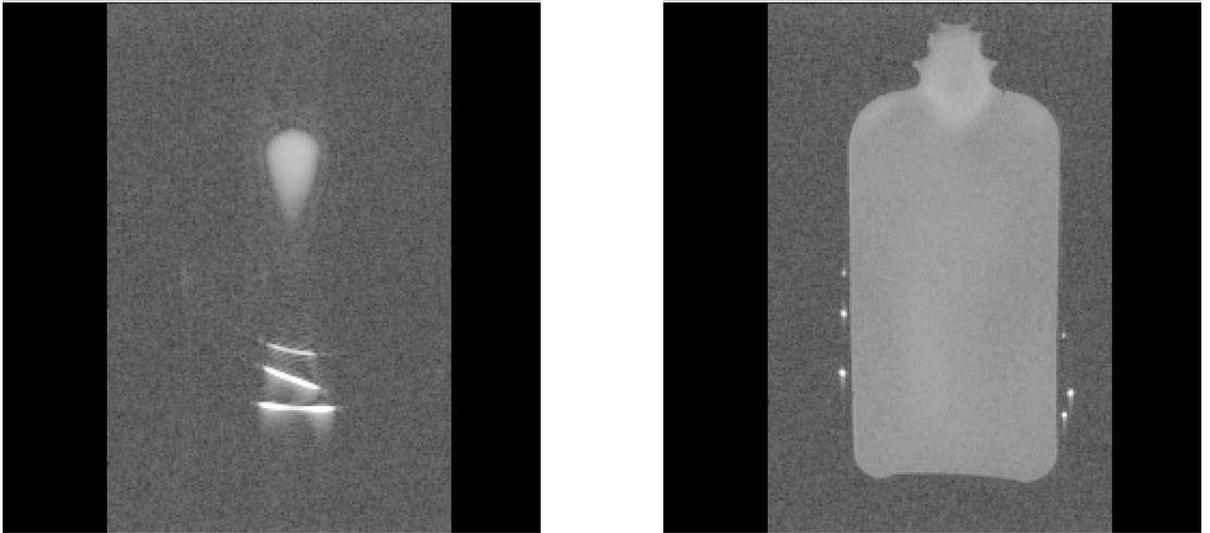


Abbildung 6.3: Zwei Beispielschnitte aus dem MR-Datenwürfel (Schnitt Nr. 12 und 43 von 140 Schnitten).

daher eine Redigitalisierung der Daten senkrecht zur Schlauchzentralen erfolgen. Eine solche Redigitalisierung ist in [Kol95] beschrieben. Dabei wird zunächst eine dreidimensionale Vorsegmentierung der Testschläuche durchgeführt. Anhand dieser Vorsegmentierung erfolgt dann eine Splinepassung der Schlauchzentralen, so daß für jeden Punkt im Schlauch seine aktuelle Richtung bekannt ist. Senkrecht zu dieser Richtung wird nun der Datenwürfel redigitalisiert. Ein derart redigitalisiertes Bild ist in Abbildung 6.4 dargestellt.

An dieser Stelle könnte bereits eine Kreispassung des mittleren Schlauches erfolgen. Der Redigitalisierungsausschnitt wurde jedoch so gewählt, daß auch der Flaschenrand, um den der Schlauch gewickelt wurde im redigitalisierten Bild sichtbar ist. Zusätzlich existieren zwei weitere Schläuche sowie ein schwarzer Bereich auf der rechten Seite, der das Ende des Datenwürfels kennzeichnet. Wollte man dieses Bild als Ausgangsbild für eine Kreispassung verwenden, müsste man diese zusätzlichen Strukturen in den Hintergrund mit einmodellieren, was eindeutig einen zu großen und zudem unnötigen Aufwand darstellt. Man ist also darauf angewiesen, einen entsprechend kleinen Ausschnitt rund um die (mittlere) Schlauchzentrale zu wählen, der einen homogenen Hintergrund bietet. Genau dies leistet das in [Kol95] vorgestellte Verfahren zur automatischen Segmentation einer *area of interest*, die in [Kol95] *Segmentationslupe* genannt wird. Eine nachfolgende Gefäßsegmentierung — egal ob pixelbasiert oder subpixelgenau — braucht dann nicht mehr das gesamte Bild, sondern nur noch die Pixel innerhalb dieser Lupe zu berücksichtigen.

Abbildung 6.5 zeigt den redigitalisierten Querschnitt aus Abbildung 6.4, links ohne und rechts mit zugehöriger Lupenmaskierung. In der nachfolgenden Passung werden dann nur noch die Pixel innerhalb dieser Lupe berücksichtigt. Da alle Pixel des Schlauchquerschnitts innerhalb der Lupe liegen, sollte auch der gepaßte Ergebniskreis komplett

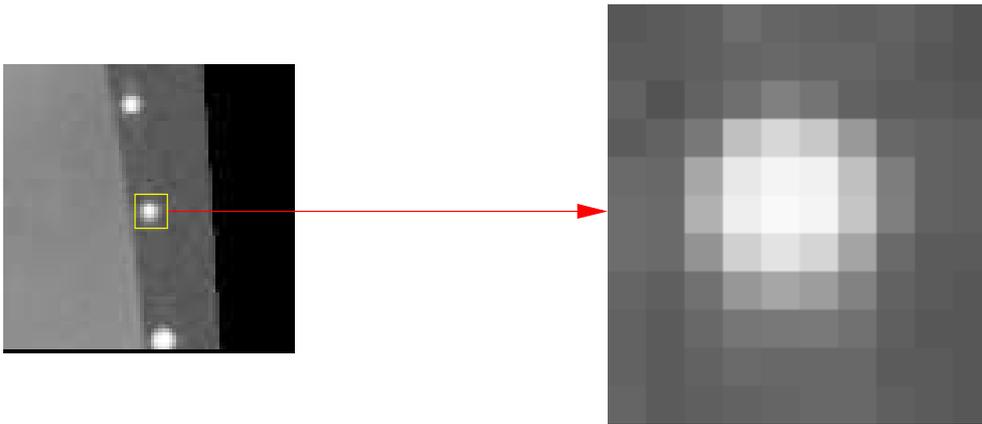


Abbildung 6.4: Links ein senkrecht zur (mittleren) Schlauchzentralen redigitalisiertes Bild, rechts eine Ausschnittsvergrößerung des mittleren Schlauches.

innerhalb der Lupe liegen und sie damit insbesondere weder völlig abdecken noch komplett aus ihr rausfallen. Damit können die Mengen \mathcal{A} aus Kreisscheibenpassung bzw. modifizierter Grauwertadaptation ohne Probleme passend gewählt werden.

Für den Test beider Verfahren standen insgesamt 307 redigitalisierte lupenmaskierte Querschnittsbilder zur Verfügung. Als Abstiegsverfahren in der Optimierung der Zielfunktionen (gemäß Algorithmusskizze auf Seite 39) wurde die *downhill simplex method* nach Nelder und Mead ([NM65]) verwendet. Dieses Abstiegsverfahren kommt ohne weitere Annahmen an die Zielfunktion aus, insbesondere benötigt es keine Gradientenfunktionen. Außerdem zeigte es im Vergleich mit anderen Abstiegsverfahren kaum Ausreißer durch ein „Hängenbleiben“ in lokalen Optima und erwies sich zudem am störunanfälligsten gegenüber Startwertänderungen. Die ausführlichen Ergebnisse der Testdurchläufe finden sich in Anhang B.

Im Vergleich beider Verfahren erwies sich die modifizierte Grauwertmodelladaptation als stabiler, wie der Radienvergleich in Abbildung B.11 auf Seite 120 zeigt. Diese Tatsache wird auch durch die deutlich geringere Standardabweichung von 0,024 im Gegensatz zu 0.065 bei der Kreisscheibenpassung belegt. Außerdem ist der Fehler im Optimum bei der modifizierten Grauwertmodelladaptation deutlich geringer als bei der Kreisscheibenpassung (Abbildung B.10). Diese Diskrepanz mag darin liegen, daß die Modellierung in der (modifizierten) Grauwertmodelladaptation dem tatsächlichen Bildentstehungsprozeß näher kommt als die in der Kreisscheibenpassung verwendete Modellierung. Beispielhaft kann das in den Abbildungen 6.6 und 6.7 nachvollzogen werden. Neben den optimalen Paßergebnissen werden dort auch die mit dem jeweiligen Modell (d.h. der jeweiligen semilinearen Funktion g) berechneten Bilder der optimalen Kreisscheibe gezeigt. Das auf Basis der Kreisscheibenpassung diskretisierte Bild wirkt in beiden Fällen gröber als das auf Basis der modifizierten Grauwertmodelladaptation diskretisierte Bild. Alles in allem steht die Kreisscheibenpassung hinten an, was die Genauigkeit angeht.

Andererseits existieren weitere Aspekte im Vergleich, die ebenfalls berücksichtigt werden müssen. So ist die Schwankung der Ergebnisradien bei der Kreisscheibenpassung nicht so

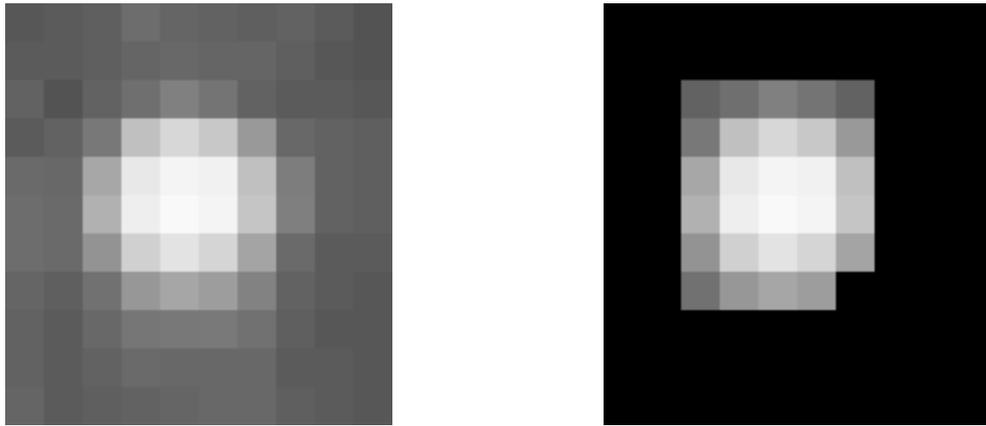


Abbildung 6.5: Der redigitalisierte Gefäßquerschnitt aus Abbildung 6.4, links ohne, rechts mit Lupenmaskierung. Die schwarzen Pixel im rechten Bild liegen außerhalb der Lupe. Bei einer Gefäßpassung brauchen nur noch die Pixel innerhalb der Lupe berücksichtigt werden.

hoch, daß die Ergebnisse nicht mehr verwendet werden können. Abbildung 6.7 zeigt das Bild innerhalb der untersuchten Testreihe, bei dem die beiden Paßergebnisse die größte Radiendifferenz aufwiesen. Trotz dieser großen Differenz unterscheiden sich die beiden Ergebnisse visuell nur wenig. Dazu kommt ein weiterer Vorteil der Kreisscheibenpassung, nämlich der der Schnelligkeit. Dieser Vorteil resultiert aus der unterschiedlichen Komplexität in der Berechnung der Fehlerfunktion. Für die Kreisscheibenpassung ist pro Pixel (k, l) die Berechnung der Schnittmenge von Kreisscheibe mit Pixelquadrat in konstanter Zeit möglich. Bei der modifizierten Grauwertmodelladaption muß der Teilterm (2.7) von Seite 29 jedoch numerisch approximiert werden und darin geht pro Pixel die Anzahl der Abtastschritte des Intervalls $[-R, R]$ linear in die Berechnung ein. Dieser Unterschied in der Komplexität schlägt sich sehr deutlich in der Rechenzeit nieder: Für die vorliegenden Testbilder betrug die durchschnittliche Zeit zur einfachen Berechnung der Fehlerfunktion bei der modifizierten Grauwertmodelladaption 13 ms, die gesamte Optimierung einer Kreisscheibe dauerte durchschnittlich 7700 ms. Bei der Kreisscheibenpassung dagegen war die gesamte Optimierung für ein Bild bereits in durchschnittlich 22 ms erledigt, die Zeit für die einzelne Berechnung der Fehlerfunktion war nicht mehr meßbar.

Es liegt also an der Ausgangssituation, welches der beiden Verfahren verwendet werden sollte: Steht die Schnelligkeit im Vordergrund muß der Kreispassung der Vorzug gegeben werden — vorausgesetzt die damit erzielte Genauigkeit ist ausreichend. Werden höhere Genauigkeiten verlangt, sollte wenn möglich die modifizierte Grauwertmodelladaption verwendet werden — wobei dann allerdings deutlich längere Rechenzeiten in Kauf genommen werden müssen.

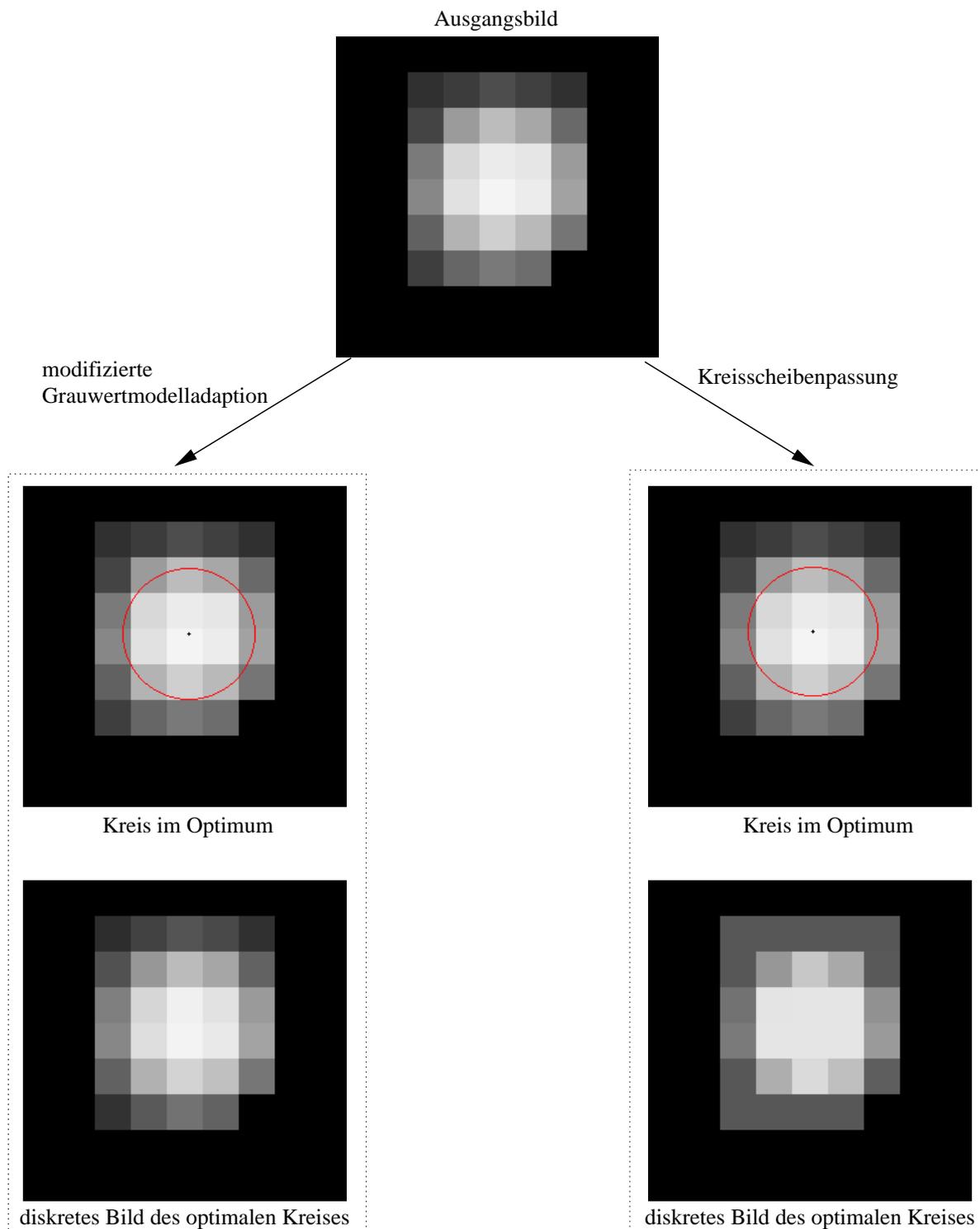


Abbildung 6.6: Das lupenmaskierte Querschnittsbild aus Abbildung 6.5 zusammen mit den Ergebnispassungen beider Verfahren. Die Ergebnisbilder zeigen für beide Verfahren die berechnete Kreisscheibe in das Originalbild eingeblendet und darunter das gemäß der jeweiligen Modellierung diskretisierte Bild der Kreisscheibe.

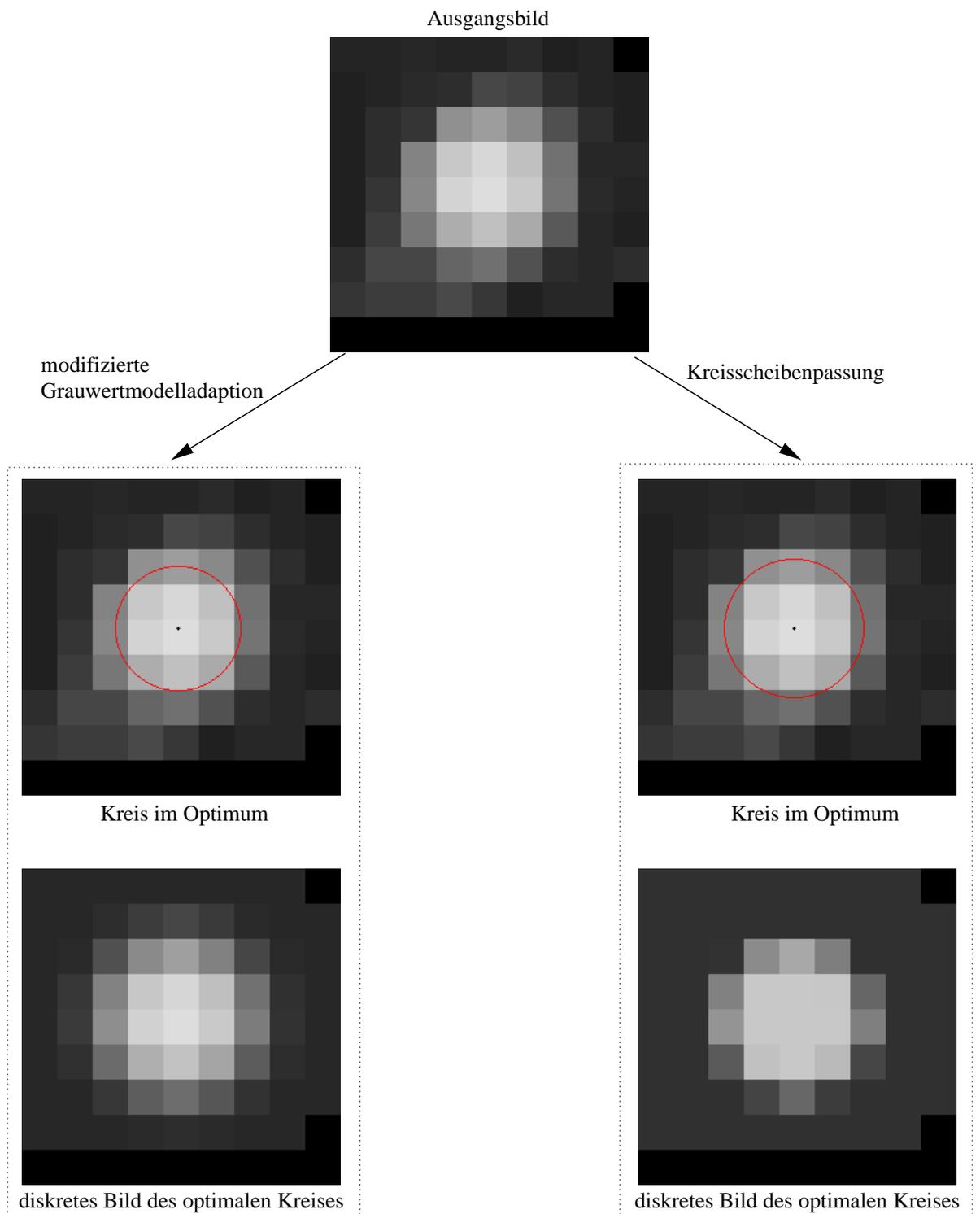


Abbildung 6.7: Die berechneten optimalen Ergebniskreise für Testbild Nr. 304. Dieses Testbild wies beim Vergleich beider Verfahren die größte Radiendifferenz im Ergebnis auf. Der berechnete optimale Radius links betrug 1.79 Pixel, rechts dagegen 1.98 Pixel, ein Unterschied von 0.19 Pixel.

7 Objektrekonstruktion in der Computer-Tomographie

Dieses Kapitel behandelt die bereits in der Einleitung beschriebene Einbrandtiefenbestimmung. Im Gegensatz zu der Ausgangssituation in Kapitel 6, wo die Subpixelgenauigkeit im Vordergrund stand, geht es hier darum, sichtbare Strukturen zu eliminieren um untergeordnete Strukturen, die bis dahin überdeckt wurden, sichtbar zu machen.

Dieses Problem war zwar der eigentliche Auslöser für die hier vorliegende Arbeit. Es zeigte sich aber schnell, daß bei der Anwendung der in den vorigen Kapiteln entwickelten Konzepte Kompromisse eingegangen werden mußten, um die Komplexität des Problems in Grenzen zu halten. Es wurden daher nur ein paar Aspekte aus dieser Arbeit verwendet und aus diesem Grund (und angesichts der geringen Zahl der Testobjekte) sind auch die Ergebnisse nur als Indiz für das Funktionieren der angewendeten Methode und weniger als exakte Lösung des gestellten Problems zu werten.

7.1 Problemstellung

Gegeben sind Aluminiumteile, die durch einen Schweißvorgang miteinander verbunden wurden. Die Festigkeit der Schweißnaht hängt dabei vor allem davon ab, wie weit sie beim Schweißvorgang in die Aluminiumteile eingedrungen ist. Ein guter Halt ist nur gegeben, wenn der Schweißvorgang die beiden Teile tief miteinander verbunden hat. Haften die Aluminiumteile dagegen nur oberflächlich zusammen, sind sie nicht belastbar und brechen relativ schnell wieder auseinander. Von außen ist allerdings nicht ersichtlich, wie tief — d.h. wie haltbar — eine vorliegende Schweißnaht ist. Die Prüfung einer Schweißnaht geht bisher nur zerstörend, ein vorliegendes Stück muß an der gewünschten Stelle zersägt werden. Nach einer entsprechenden Nachbehandlung (Ätzen und Polieren) kann die Schweißnaht dann gut von den ursprünglichen Aluminiumteilen unterschieden werden. Abbildung 7.1 zeigt ein derart behandeltes Teststück, das aus zwei miteinander verschweißten Aluminiumteilen besteht. Man sieht, wie sich die (mit dem Kreis markierte) Schweißnaht deutlich heller vom umgebenden Aluminium abgrenzt. In diesem Fall hat sie sich relativ tief mit den Aluminiumteilen verbunden.

Es soll nun die Güte dieser Schweißnaht bestimmt werden, ohne daß dabei der Testkörper zerstört wird. Ein Röntgen-CT-Verfahren war dabei die Methode der Wahl, weil dadurch die meisten inneren Strukturen eines Werkstückes sichtbar gemacht werden können. Um

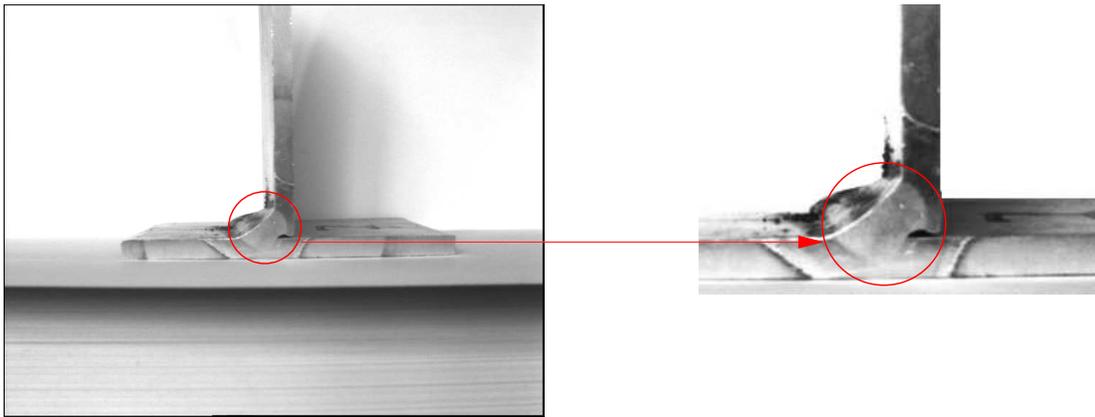


Abbildung 7.1: Ein aufgeschnittener Aluminium-Testkörper (Länge der Seitenteile ca. 6 cm). Die Tiefe der hier sichtbaren Verschweißung (heller Bereich innerhalb des Kreises) soll im CT-Bild ermittelt werden.

die Ergebnisse überprüfen zu können, wurden daher von sieben Teststücken an einer markierten Stelle CT-Aufnahmen gemacht und die Teststücke nachher an derselben Stelle zersägt. Abbildung 7.2 zeigt das über eine gefilterte Rückprojektion rekonstruierte CT-Bild des Teststückes aus Abbildung 7.1. Auffällig sind die starken Rekonstruktionsartefakte (dunklen Streifen) im Bereich der Schweißnaht. Der Schweißnahtbereich an sich ist dagegen nicht vom restlichen Aluminium zu unterscheiden.

In einem anderen Teststück (Abbildung 7.3) mit weitaus kürzeren Seitenteilen ist die Schweißnaht dagegen (wenn auch nur schemenhaft) auch im CT-Bild zu erkennen. Diese Beobachtung legt nahe, daß die Schweißnaht auch in Abbildung 7.2 zu erkennen sein müßte, die relativ geringe Änderung im Röntgensignal aber durch das starke Signal der beiden langen Seitenteile verdeckt wird.

Der Ausgangspunkt für die folgenden Überlegungen ist also, daß in den CT-Rohdaten die Schweißnaht durchaus vorhanden ist, in der Rekonstruktion aber unter anderem durch Artefaktbildung und Überlagerung verdeckt wird. Könnten also aus den Rohdaten die störenden Signale der beiden Seitenteile eliminiert werden, würde theoretisch nur noch das Signal der Schweißnaht übrigbleiben und sollte in einer anschließenden Rekonstruktion sichtbar werden. Die folgenden Abschnitte zeigen, wie diese Überlegungen durch eine semilineare Modellierung realisiert werden kann.

Das vorliegende Problem ist klar unterscheidbar von einem *incomplete data problem*. Bei Problemen dieser Art liegt kein vollständiger (diskreter) CT-Datensatz vor. Das bekannteste Problem dieser Art ist das *limited angle problem*, bei dem die CT-Daten nur in einem eingeschränkten Winkelbereich vorhanden sind (siehe z.B. [Nat86] für eine ausführlichere Beschreibung). Für diese Art von Problemen existieren je nach konkreter Fragestellung viele Algorithmen (siehe z.B. [LR84] oder [Qui93]). Im vorliegenden Fall liegen aber vollständige CT-Daten vor, jedoch sind manche Strukturen so schwach ausgeprägt, daß sie in der gefilterten Rückprojektion unsichtbar sind.

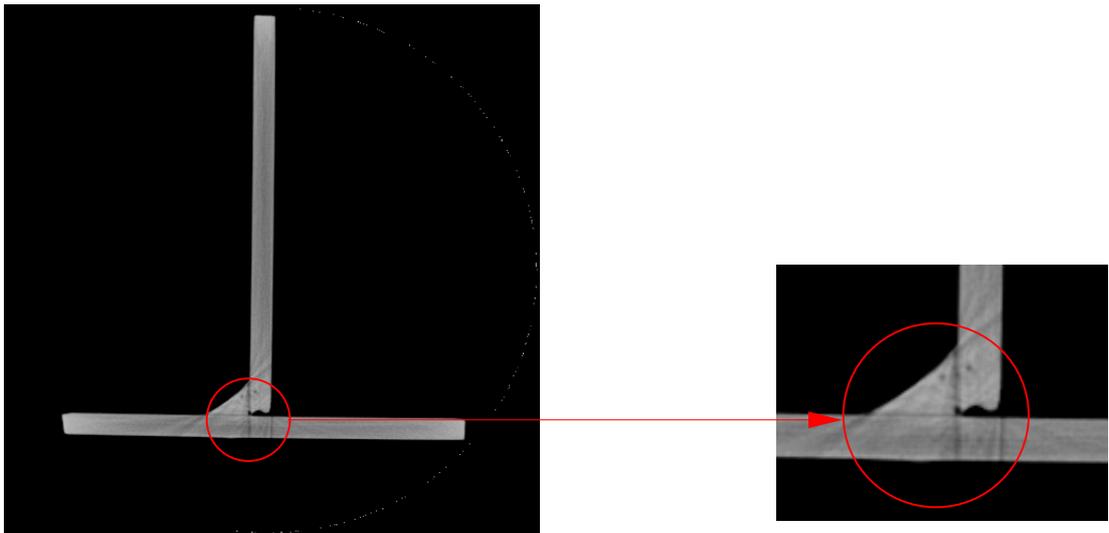


Abbildung 7.2: Das Teststück von Abbildung 7.1 als CT-Bild. Die schwarzen Streifen im Schweißnahtbereich sind Rekonstruktionsartefakte, die Schweißnaht selber ist nicht sichtbar.

7.2 Absorptionsverhalten von Röntgenstrahlung

Abbildung 7.4 zeigt den prinzipiellen Aufbau einer Röntgen-CT-Anlage: Ausgehend von der Quelle passiert der Röntgenstrahl den Rekonstruktionsbereich, in dem sich das zu untersuchende Objekt befindet und trifft schließlich auf den Detektor. Bei seinem Weg durch das zu untersuchende Objekt wird die Strahlintensität entsprechend der Art des Objektes geschwächt, der Detektor registriert die verbliebene Intensität. In der Regel existiert ein zusätzlicher Referenzdetektor, der die ungeschwächte Intensität des Röntgenstrahls mißt, um auftretende Energieschwankungen in der Quelle zu erkennen.

Für einen vollständigen (diskreten) Datensatz aus einer Richtung müssen Quelle und Detektor senkrecht zur Strahlrichtung transliert werden, bis der gesamte Rekonstruktionsbereich abgedeckt ist (Parallelstrahlgeometrie). Moderne Anlagen besitzen meist eine davon abweichende Anlagengeometrie: Die Röntgenquelle sendet einen fächerförmigen Strahl aus, der dann von einer Detektorzeile registriert wird (Fächerstrahlgeometrie). Um Daten aus verschiedenen Richtungen zu erhalten, werden Quelle und Detektor zusätzlich um den Rekonstruktionsbereich rotiert.

Zwischen der Startintensität I_0 an der Quelle und der Restintensität I am Detektor wird normalerweise der folgende Zusammenhang angenommen:

$$I = I_0 e^{-\int_L \mu(x) dx} \quad (7.1)$$

L bezeichnet dabei die Strecke, die der Röntgenstrahl von der Quelle bis zum Detektor zurücklegt (siehe Abb. 7.4) und μ die Schwächung des Strahls am Punkt x . Für den Schwächungskoeffizienten μ gelten die folgenden Annahmen:

- μ ist nur abhängig von der Art des durchstrahlten Materials und

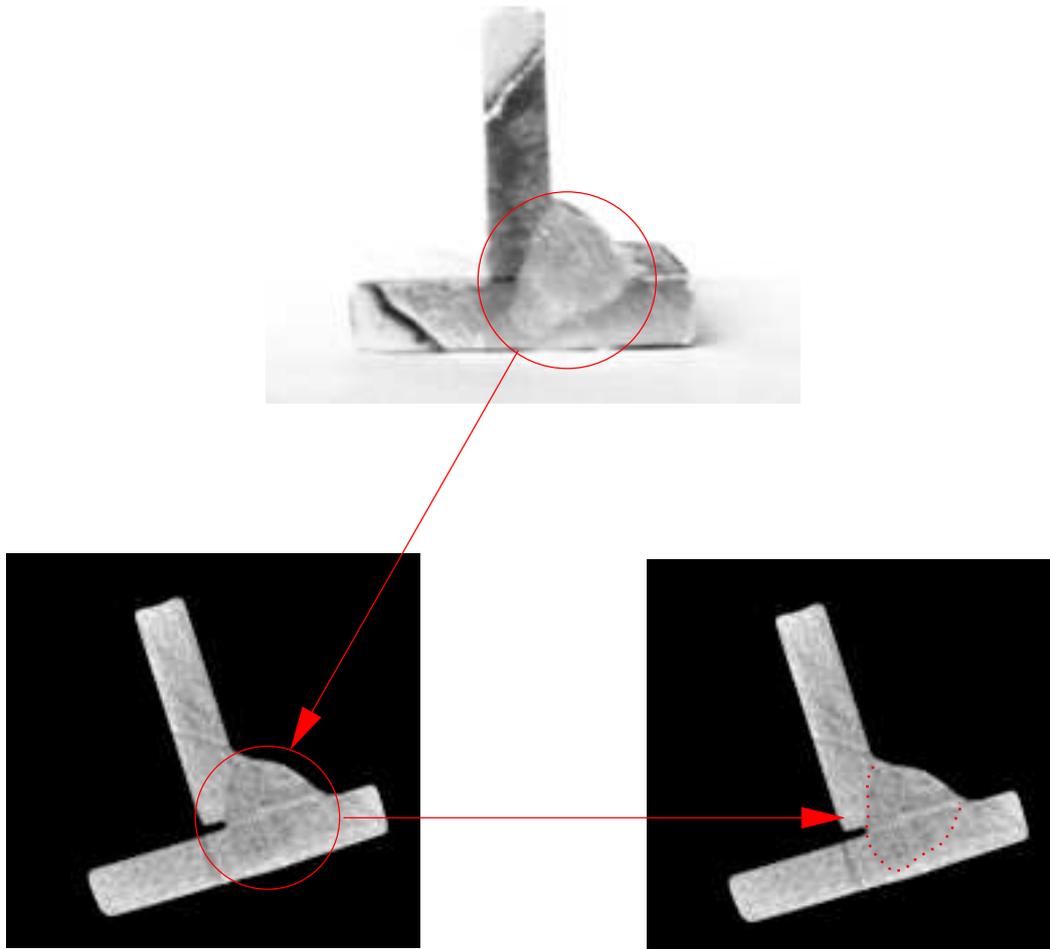


Abbildung 7.3: Ein weiteres Teststück mit kürzeren Seitenteilen (Länge der Seitenteile ca. 1,5 – 2 cm). Oben der aufgesägte Querschnitt mit deutlich sichtbarer Schweißnaht, unten links der CT-Schnitt. Der Schweißnahtbereich ist hier nur schemenhaft sichtbar und wird teilweise von Rekonstruktionsartefakten überlagert, korreliert in seinen Abmessungen aber dennoch deutlich mit dem Schweißnahtbereich im Foto. Zur Verdeutlichung ist er im CT-Schnitt unten rechts markiert worden.

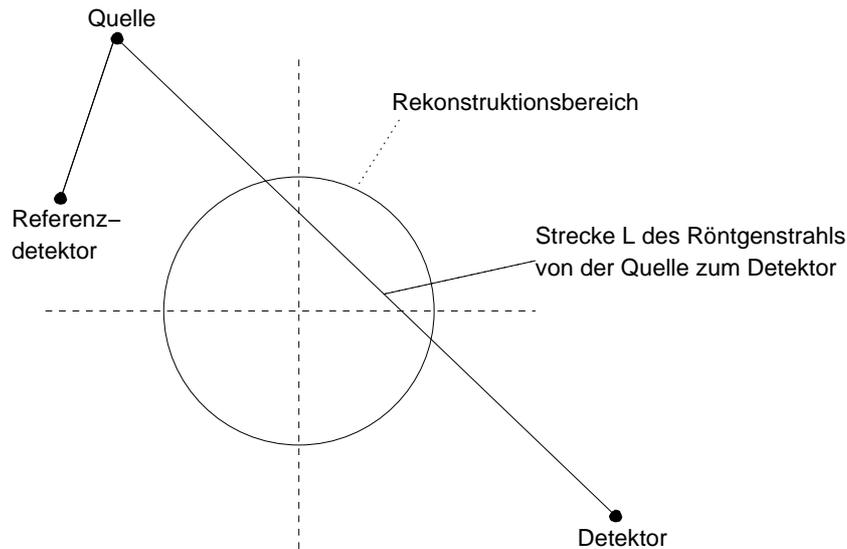


Abbildung 7.4: Skizze einer Röntgen-CT-Anlage (ohne Rotation und Translation)

- der Röntgenstrahl wird in der Luft (und damit auch außerhalb des Rekonstruktionsbereiches) nicht geschwächt, d.h. μ ist in diesen Bereichen gleich 0.

Befindet sich im Rekonstruktionsbereich also ein Objekt, das nur aus einem einzigen Material mit Schwächungskoeffizient μ_0 besteht — wie es für die Aluminiumteile der Fall ist — ist μ konstant gleich μ_0 und die Schwächung reduziert sich zu

$$I = I_0 e^{-\int_L \mu(x) dx} = I_0 e^{-\mu_0 \cdot d}$$

wobei d die Länge des Weges angibt, den der Röntgenstrahl im Objekt zurücklegt. Wegen des dadurch gegebenen linearen Zusammenhangs

$$\ln \left(\frac{I_0}{I} \right) = \mu \cdot d \quad (7.2)$$

heißt μ auch der *lineare Massenschwächungskoeffizient*. Durch (7.1) wird eine semilineare Modellierung erst möglich: d kann über eine Erzeugendenfunktion berechnet werden und zu jedem d der jeweils passende Schwächungskoeffizient μ optimal bestimmt werden.

Der in (7.1) zugrundeliegende Zusammenhang und auch die Annahmen an den Massenschwächungskoeffizienten μ sind allerdings nur eine Annäherung an die tatsächlichen Verhältnisse. In Wahrheit strahlt eine Röntgenquelle normalerweise nicht ein einzelnes Energieniveau sondern ein ganzes Energiespektrum aus und auch die Detektoren reagieren nicht auf ein einzelnes Energieniveau sondern wiederum auf ein ganzes Energiespektrum. Zusätzlich ist der Massenschwächungskoeffizient nicht (wie laut Annahme) unabhängig von der Energieintensität, was zu wesentlich komplizierteren Zusammenhängen führt (siehe dazu auch [Nat86]). Die Vernachlässigung dieser Zusammenhänge kann zu Rekonstruktionsartefakten, sogenannten *Strahlaufhärtungsartefakten* führen (siehe dazu

auch [Her80]). Diese Art der Artefaktbildung in den Rekonstruktionen wird normalerweise hingegenommen, da sie in der Regel wenig störend ist und vor allem in medizinischen Anwendungen nur in verschwindend geringem Maße auftritt. Gerade in der vorliegenden Anwendung aber treten durch die — im Vergleich zu Wasser — höhere Dichte von Aluminium starke Artefakte vor allem auch im interessanten Bereich der Schweißnaht auf, durch die die Qualität des Rekonstruktionsergebnisses zusätzlich vermindert wird.

Eine weitere Voraussetzung, die normalerweise nicht explizit formuliert wird, in sämtlichen Rekonstruktionsalgorithmen aber eine wichtige Rolle spielt, ist die Annahme der Punktförmigkeit von Quelle und Detektor. Dies hat den Vorteil, daß entlang der Strecke Quelle–Detektor nur ein Linienintegral berücksichtigt werden muß. Würde man für die Detektoren eine flächige Ausdehnung voraussetzen (wie es in der Praxis der Fall ist), wäre die Voraussetzung eines Linienintegrals nicht mehr gegeben und die Rekonstruktionsalgorithmen würden ungleich komplizierter werden. Die Annahme der Punktförmigkeit wird daher auch in den folgenden Abschnitten vorausgesetzt. Gemäß der Sensormodellierung nach Kapitel 4 müssen dann die Sensorcharakteristiken durch Punktmaße repräsentiert werden.

7.3 Modellierung der Objektform

In der hier beschriebenen Modellierung wird die CT-Anlage unabhängig von ihrer konkreten Geometrie (Fächer- bzw. Parallelstrahl) durch einen Sensor $(A_i, \mu_i)_{i \in I}$ im Sinne von Definition 4.1.1 beschrieben. Da die Punktförmigkeit von Quelle und Detektor eine Grundannahme ist, werden für die Sensorelemente μ_i Dirac-Maße

$$\mu_i = \delta_{x_i} \quad \text{für } x_i \in A_i \quad \text{und } i \in I$$

gewählt. Unter dieser Voraussetzung kann auch die Form der Aperturen A_i durch

$$A_i := \{x_i\} \quad \text{für } i \in I$$

als einelementig festgelegt werden. Durch diese Punktförmigkeit der Sensorelemente kann im folgenden der Digitalisierungsschritt vernachlässigt werden, denn für die Digitalisierung $l(f)$ gemäß Kapitel 4 einer Sensorinputfunktion $f : \mathcal{A} \rightarrow L^1_{\delta_{x_i}}(\{x_i\}, \mathbb{R})$ gilt

$$l(f) = \int f d\delta_{x_i} = f(x_i).$$

Eine zweidimensionale CT-Anlage wird somit im folgenden durch einen Sensor

$$(\{x_i\}, \delta_{x_i})_{i \in I}$$

repräsentiert. Die Aufgabe besteht nun darin, aus einer gegebenen Aufnahme $y \in \mathbb{R}^{\#I}$ eines Aluminiumobjektes den Schweißnahtbereich zu extrahieren. Man beachte dabei,

daß y hier nicht das bereits rekonstruierte Bild bezeichnet, sondern die von der Anlage gelieferten *Rohdaten* (bzw. ihre logarithmierte Version gemäß (7.2)). Ein solches Vorgehen scheint ziemlich umständlich zu sein, ist aber schon allein deshalb notwendig, weil in den gefilterten Rückprojektionen fast aller Objekte der Schweißnahtbereich vollständig verschwunden ist. Ein erfolgversprechendes Verfahren kann daher nur auf den Rohdaten aufsetzen.

Das prinzipielle Vorgehen des hier verwendeten Algorithmus wurde bereits auf Seite 94 beschrieben: Zunächst wird angenommen, daß der Bereich der Schweißnaht dieselbe Dichte wie das restliche Objekt besitzt. Dann wird das Signal des gesamten Aluminiumobjektes aus den Rohdaten eliminiert. Das verbliebene Signal enthält nun außer dem Hintergrundrauschen nur noch die Dichteänderung Aluminiumteil–Schweißnaht, die in der Modellierung des Aluminiumobjektes vernachlässigt wurde. Da diese Dichteänderung in der Rekonstruktion nicht mehr von dem starken Absorptionssignal der Aluminiumteile überdeckt wird, besteht die Möglichkeit, daß sie sichtbar wird, wenn das verbliebene Signal zu einem Bild rekonstruiert wird.

Sei $\mathcal{A} \subseteq \mathfrak{A}(\mathbb{R}^2)$ die Menge der möglichen Aluminiumobjekte. Jede Apertur $\{x_i\}$ „sieht“ einen Strahl ausgehend von der Röntgenquelle $Q \in \mathbb{R}^2$ hin zu x_i , der dabei eventuell durch das zu untersuchende Objekt $A \in \mathcal{A}$ hindurchgeht. Die Länge dieses Strahls entspricht

$$\int_{Qx_i} \mathbb{1}_A d\lambda.$$

Da in diesem ersten Schritt der Schweißnahtbereich vernachlässigt wird, genügt eine einzelne Erzeugendenfunktion, um das Objekt vollständig zu beschreiben:

$$e^{(1)} : \mathcal{A} \rightarrow \mathbb{R}^{\#I}, \quad e_i^{(1)}(A) := \int_{Qx_i} \mathbb{1}_A d\lambda \quad \text{für alle } i \in I.$$

Laut Annahme auf Seite 97 wird der Röntgenstrahl durch die Luft nicht geschwächt. Da in der Praxis dennoch eine leichte Schwächung auftritt, wird eine weitere Erzeugendenfunktion

$$e^{(2)} : \mathcal{A} \rightarrow \mathbb{R}^{\#I}, \quad e_i^{(2)}(A) := L_i - \int_{Qx_i} \mathbb{1}_A d\lambda \quad \text{für alle } i \in I$$

verwendet, um dies abzufangen. $L_i \in \mathbb{R}$ bezeichnet die Länge der Strecke $\overline{Qx_i}$. Durch diese zwei Erzeugendenfunktionen sind Objekt und Hintergrund ausreichend modelliert.

Bemerkung 7.3.1

Dieses Vorgehen stellt einen Kompromiss zwischen einer möglichst exakten Modellierung der Problematik durch entsprechende Mengensysteme und der Rechengeschwindigkeit dar. Für eine — im Sinne dieser Arbeit — exakte Beschreibung des Schweißnahtrekonstruktionsproblems müßte neben dem Objekt selber auch der Schweißnahtbereich gesondert durch eine Erzeugendenfunktion modelliert werden. Das macht eine Änderung von \mathcal{A} nötig, denn es werden nun nicht mehr nur ein sondern zwei Mengen —

für das Aluminiumobjekt sowie für seine Schweißnaht — benötigt. Zudem gilt für jeden Schweißnahtbereich $S \subseteq \mathbb{R}^2$ von $A \in \mathcal{A}$ notwendigerweise $S \subseteq A$. Statt \mathcal{A} wird dann die Menge

$$\mathcal{A}_S := \{ (A, S) \in \mathcal{A} \times \mathfrak{A}(\mathbb{R}^2) \mid S \subseteq A \}$$

benötigt. Die Erzeugendenfunktionen von Objekt und Hintergrund übertragen sich:

$$e^{(1)} : \mathcal{A}_S \rightarrow \mathbb{R}^{\#I}, \quad e_i^{(1)}(A, S) := \int_{Qx_i} \mathbb{1}_A d\lambda$$

und

$$e^{(2)} : \mathcal{A}_S \rightarrow \mathbb{R}^{\#I}, \quad e_i^{(2)}(A, S) := L_i - \int_{Qx_i} \mathbb{1}_A d\lambda.$$

Zusätzlich wird noch eine dritte Erzeugendenfunktion

$$e^{(3)} : \mathcal{A}_S \rightarrow \mathbb{R}^{\#I}, \quad e_i^{(3)}(A, S) := \int_{Qx_i} \mathbb{1}_S d\lambda$$

für den Schweißnahtbereich benötigt. Diese drei Erzeugendenfunktionen stellen eine hinreichende Basis für die Formulierung eines semilinearen Approximationsproblems dar.

Problematisch an einer derartigen Modellierung ist allerdings die Komplexität der Menge \mathcal{A}_S . Ein derartig variabler Bereich wie eine Schweißnaht läßt sich algorithmisch kaum adäquat beschreiben, so daß dieses Vorgehen nicht realisierbar ist.

Die Überlegungen zur Komplexität von \mathcal{A}_S aus Bemerkung 7.3.1 gelten leider auch für die Menge \mathcal{A} der Aluminiumobjekte. Zwar scheint ein solches Objekt auf den ersten Blick einfach beschreibbar zu sein, das Problem liegt aber wiederum im Schweißnahtbereich, an dem Ritzen oder gar Löcher auftreten können. Daher wurde die Objektform A anhand seiner gefilterten Rückprojektion über eine Kantenextraktion ermittelt und \mathcal{A} nur einelementig als

$$\mathcal{A} := \{A\}$$

gewählt. Dadurch reduziert sich die semilineare Approximation zu einem einfachen linearen Ausgleichsproblem:

$$\min_{x \in \mathbb{R}^2} \|y - x_1 e^{(1)}(A) - x_2 e^{(2)}(A)\|, \quad (7.3)$$

wobei $y \in \mathbb{R}^{\#I}$ die logarithmierte Version der von der CT-Anlage gelieferten Rohdaten bezeichnet. Abbildung 7.5 zeigt beispielhaft das Ergebnis dieses Optimierungsproblems für eine Teilmenge von I .

Mit (7.3) ist bereits der wichtigste Schritt in der Schweißnahtrekonstruktion geschehen. Bezeichnet nun $(\alpha_0, \beta_0)' \in \mathbb{R}^2$ den in (7.3) gefundenen optimalen Vektor, können jetzt die Differenzdaten

$$y_d := y - \alpha_0 e^{(1)}(A) - \beta_0 e^{(2)}(A) \in \mathbb{R}^{\#I}$$

gebildet werden. Diese Differenzdaten, die laut Voraussetzung nur noch die Dichteänderung zwischen Aluminiumobjekt und Schweißnaht enthalten, können schließlich zu einem neuen Bild rekonstruiert werden (Abbildungen 7.6 bis 7.14).

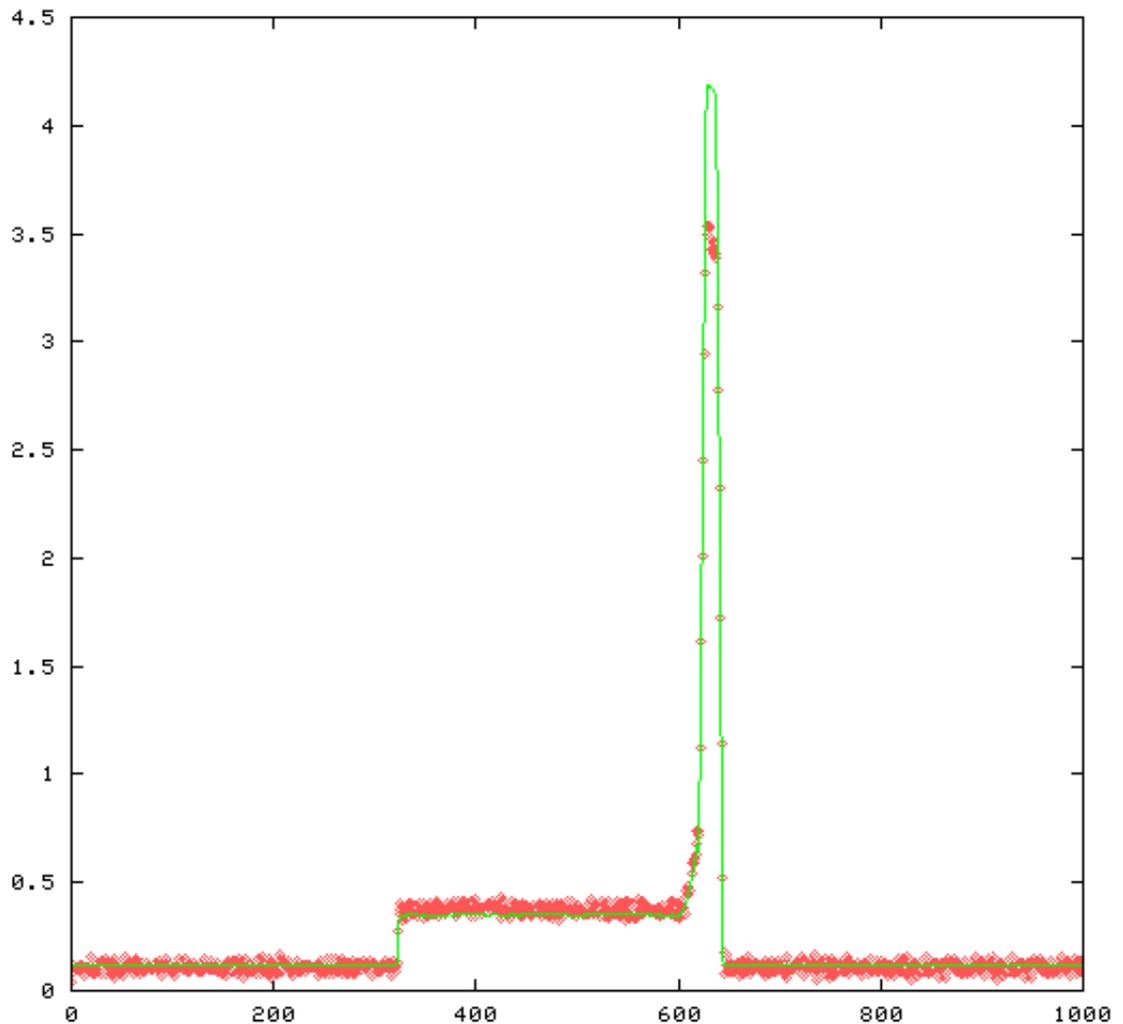


Abbildung 7.5: Logarithmierte Röntgen-CT-Daten gemäß Gleichung (7.2) eines Aluminiumobjektes (rot) sowie die berechnete Passung $x_1^{(0)}e^{(1)}(A) + x_2^{(0)}e^{(2)}(A)$ (grün) gemäß (7.3). $x_1^{(0)}$ und $x_2^{(0)}$ bezeichnen die in (7.3) gefundenen optimalen Schwächungskoeffizienten von Aluminiumobjekt bzw. Hintergrund. Die Form des Objektes wurde vorher durch eine Kantenextraktion ermittelt. Diese Abbildung zeigt nur eine Durchstrahlungsrichtung, d.h. nur eine Teilmenge der gesamten Daten.

7.4 Ergebnisse und Ausblick

Die Abbildungen 7.6 bis 7.14 zeigen die aufgeschnittenen Testkörper zusammen mit den Rekonstruktionen ihrer Differenzdaten. Der Hintergrund wurde in den Ergebnisrekonstruktionen ausgeblendet. In allen Ergebnisbildern kann ein dunkler Bereich ausgemacht werden, der in etwa dem Schweißnahtbereich entspricht. Dies kann als deutlicher Hinweis gewertet werden, daß das verwendete Verfahren erfolgreich ist. Allerdings zeigen die Ergebnisse auch, daß die Position dieses dunklen Bereiches nicht notwendigerweise mit der tatsächlichen Lage des Schweißnahtbereiches übereinstimmt. Insbesondere fällt die Ergebnisrekonstruktion von Objekt Nr. 4 (Abbildung 7.11) negativ auf, denn in deren ursprünglicher Objektrekonstruktion (Abbildung 7.3) war der Schweißnahtbereich bereits erkennbar. Derartige Unkorrektheiten lassen sich am ehesten durch eine fehlerhafte Modellierung der Objektform erklären. Einen deutlichen Hinweis darauf bietet auch Abbildung 7.6 links, die die Rekonstruktion der Differenzdaten von Objekt Nr. 5 vor dem Ausblenden des Hintergrundes zeigt: Obwohl laut Annahme das Objekt aus den CT-Daten eliminiert wurde, sind die Objektränder noch deutlich sichtbar.

Eine Weiterentwicklung dieses Verfahrens bedeutet zuallererst die Bereitstellung eines passenden Mengenmodells für die Objektform, damit die Artefaktbildung in den Ergebnisrekonstruktionen wie sie Abbildung 7.6 zeigt, minimiert werden kann. Eine möglichst exakte Bestimmung von Position und Form des Aluminiumobjektes ist für eine genaue Bestimmung der Einbrandtiefe sowieso unerlässlich. Erst wenn dann in den Ergebnisrekonstruktionen die dunklen Bereiche weitestgehend mit den tatsächlichen Einbrandtiefen übereinstimmen, ist das Verfahren an sich erfolgversprechend.

Eventuell ist auch die Einbeziehung weiterer physikalischer Effekte in den Modellierungsprozeß notwendig. Hier ist vor allem die Strahlauvhärtung zu nennen, die hauptverantwortlich für die Artefaktbildung in den ursprünglichen Objektrekonstruktionen (Abbildung 7.2) sein dürfte.

Abschließend läßt sich sagen, daß die im folgenden gezeigten Ergebnisse (Abbildungen 7.6 bis 7.14) keine gesicherten Rückschlüsse auf die tatsächliche Leistungsfähigkeit des Verfahrens zulassen. Es sind zum einen zu wenig Testobjekte vorhanden und zum anderen lassen sich Größe und Lage des Schweißnahtbereiches in den Ergebnisbildern nicht vermessen. Dieses Kapitel ist daher eher als vorgezogener Ausblick auf noch vorhandene Möglichkeiten der semilinearen Approximation als als konkrete Bereitstellung eines Verfahrens aufzufassen.

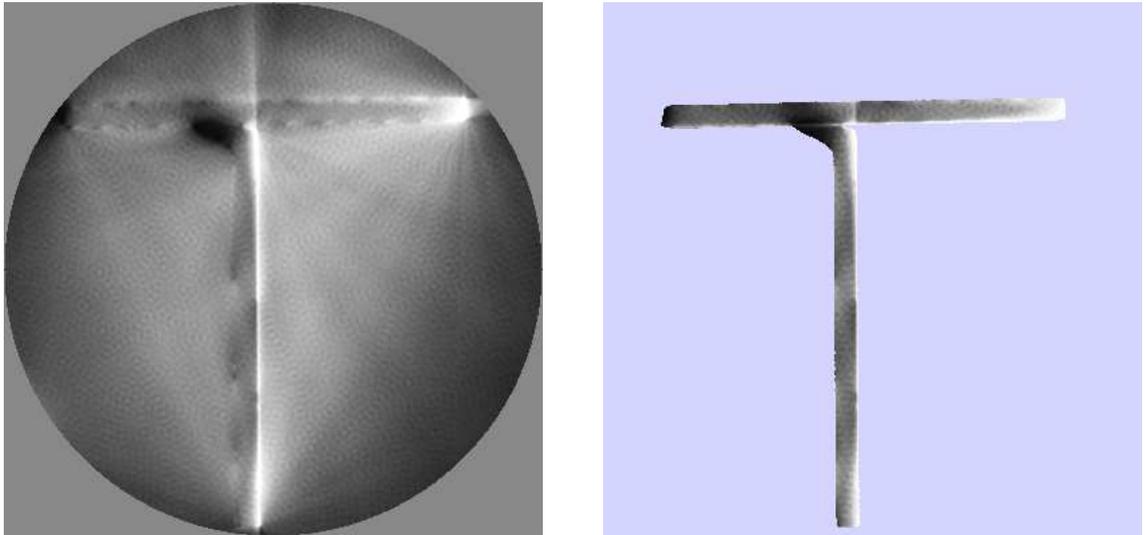


Abbildung 7.6: Ergebnisrekonstruktion von Objekt Nr. 5, links der Gesamtbereich, rechts nur der Bereich innerhalb des eigentlichen Objektes. Die Schweißnaht ist im gedruckten Bild deutlich schlechter sichtbar als auf dem Bildschirm. Um dies zu kompensieren, wurden diese (und alle folgenden) Ergebnisbilder stark kontrastverstärkt.

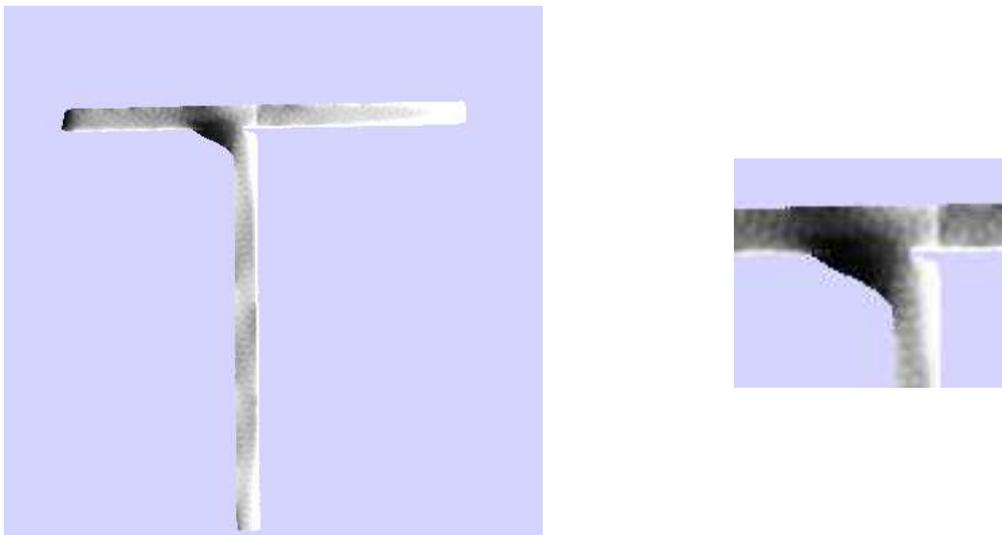


Abbildung 7.7: Ergebnisrekonstruktion von Objekt Nr. 5 mit ausgeblendetem Hintergrund. Im Gegensatz zu Abbildung 7.6 wurde die Strahlaufhärtung im Schweißnahtbereich retuschiert. Die rechte Abbildung zeigt den eigentlich interessanten Ausschnitt.



Abbildung 7.8: Der aufgeschnittene Testkörper Nr. 1 (links) und seine Ergebnisrekonstruktion (rechts). Die Verwischungen im gekennzeichneten Bereich des Ergebnisbildes wurden wohl durch eine stark fehlerhafte Modellierung der Objektform verursacht: das Loch im Objekt (heller Punkt im Ergebnisbild) wurde nicht mit berücksichtigt.



Abbildung 7.9: Der aufgeschnittene Testkörper Nr. 2 (links) und seine Ergebnisrekonstruktion (rechts).



Abbildung 7.10: Der aufgeschnittene Testkörper Nr. 3 (links) und seine Ergebnisrekonstruktion (rechts).



Abbildung 7.11: Der aufgeschnittene Testkörper Nr. 4 (links) und seine Ergebnisrekonstruktion (rechts).



Abbildung 7.12: Der aufgeschnittene Testkörper Nr. 5 (links) und seine Ergebnisrekonstruktion (rechts).



Abbildung 7.13: Der aufgeschnittene Testkörper Nr. 6 (links) und seine Ergebnisrekonstruktion (rechts).



Abbildung 7.14: Der aufgeschnittene Testkörper Nr. 7 (links) und seine Ergebnisrekonstruktion (rechts).

8 Zusammenfassung und Ausblick

In dieser Arbeit wurde ein Verfahren zur objektbasierten Bildsegmentierung vorgestellt. Kapitel 2 beschrieb Vor- und Nachteile bereits bestehender Verfahren im Hinblick auf ihre Möglichkeiten nach einer Vorauswahl der zu segmentierenden Mengen. In Kapitel 3 wurden die theoretischen Grundlagen der semilinearen Approximation entwickelt. Wie dieses Verfahren geeignet diskretisiert und so in eine vorhandene Sensorik integriert werden kann, wurde in Kapitel 4 dargelegt. Kapitel 5 stellte dann ein Mengenmodell vor, das in der Kreisscheibenpassung in Kapitel 6 Verwendung fand. In Kapitel 7 schließlich wurde eine weitere Anwendung vorgestellt, die aber — bedingt durch ihre Komplexität — nicht vollständig umgesetzt werden konnte.

Der Schwerpunkt bei der Entwicklung der in dieser Arbeit vorgestellten semilinearen Approximation lag auf einer vorherigen Festlegung der Art der zu segmentierenden Objekte. Im Gegensatz zu anderen Segmentierungsverfahren kann dadurch Vorwissen über die Art der zu segmentierenden Mengen eingebracht werden. Dadurch kann die Anzahl der möglichen Segmentierungsergebnisse bereits bei der Problemformulierung eingeschränkt werden. Auf diese Weise können auch in schlechten Ausgangssituationen — wie der Subpixelpassung aus Kapitel 6 — stabile Segmentierungsergebnisse erzielt werden.

Eine derartige Vorauswahl über ein Mengenmodell kann sich allerdings auch als nachteilig erweisen, nämlich genau dann wenn noch nicht mal die ungefähre Form des zu segmentierenden Objektes bekannt ist. Soll in solch einem Fall trotzdem eine semilineare Approximation zur Objektfindung verwendet werden, muß das verwendete Mengenmodell allgemein genug gefaßt sein, um sämtliche eventuell vorkommenden Objektformen zu beinhalten. In dieser Arbeit geschah dies beispielhaft bei der Berechnung des Segmentierungsergebnisses in Abbildung 2.5. Solche allgemeinen Ansätze sind prinzipiell möglich, führen aber im allgemeinen schnell zu algorithmisch schlecht handhabbaren Mengensystemen und dadurch zu einer langwierigen Optimierung mit unter Umständen zu vielen lokalen Optima. Die Stärke der hier vorgestellten semilinearen Approximation liegt daher hauptsächlich in der Nachiteration, wenn besonders genaue Einpassungen von speziellen Objektformen gewünscht werden.

Den Schluß dieser Arbeit bildet ein Ausblick auf zwei mögliche Erweiterungen. Die erste Erweiterung bezieht sich auf die Verarbeitung mehrkanaliger Bilder und die zweite Erweiterung ist eine Fortführung der Überlegungen von Bemerkung 7.3.1 aus Kapitel 7.

8.1 Mehrkanalige Bilder

Die Anwendungsbeispiele in den Kapiteln 6 und 7 basierten auf einkanaligen Bilddaten. Eine semilineare Modellierung mehrkanaliger Daten ist aber ebenfalls möglich. Ist zum Beispiel ein diskretes Farbbild

$$(B_r, B_g, B_b), \quad B_r, B_g, B_b \in \mathbb{R}^{m \times n}$$

gegeben, lassen sich die drei Farbkanäle B_r, B_g und B_b zunächst zu einem neuen einkanaligen Bild

$$B := \begin{pmatrix} B_r \\ B_g \\ B_b \end{pmatrix} \in \mathbb{R}^{m \times 3n}$$

zusammenfassen. Der Vektorraum E aus Definition 3.1.1 ist nun durch den Raum $\mathbb{R}^{m \times 3n}$ aller *Dreifach-Bilder* gegeben¹.

Jede semilineare Funktion g muß dann so gewählt werden, daß sie der speziellen Struktur dieses Vektorraumes Rechnung trägt. Eine der einfachsten Formen, die nur den ersten Kanal berücksichtigt, ist durch

$$g : \mathbb{R} \times \mathcal{A} \rightarrow \mathbb{R}^{m \times 3n}, \quad g(x, A) = x \begin{pmatrix} e(A) \\ 0 \\ 0 \end{pmatrix}$$

für eine *einkanalige Erzeugendenfunktion* $e : \mathcal{A} \rightarrow \mathbb{R}^{m \times n}$ gegeben. Will man e ebenfalls für die beiden restlichen Kanäle verwenden, erhält man durch

$$g : \mathbb{R}^3 \times \mathcal{A} \rightarrow \mathbb{R}^{m \times 3n}, \quad g(x, A) = x_1 \begin{pmatrix} e(A) \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ e(A) \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ e(A) \end{pmatrix}$$

eine gleichmäßige Gewichtung aller Kanäle. In diesem Kontext kann $e(A)$ als ein dreifach angewendetes Grauwertbild angesehen werden, das durch x_1, x_2 und x_3 Werte für die jeweiligen Farbkanäle zugewiesen bekommt.

Natürlich muß nicht für jeden Kanal dieselbe Funktion e verwendet werden. Überlegungen in diese Richtung weisen den Weg zur Modellierung von Multisensoranlagen, denn wie bereits in Kapitel 4 erwähnt können auch die Bilder einer Multisensoranlage als ein einzelnes Mehrkanalbild aufgefaßt und dementsprechend behandelt werden. Durch die auf diese Weise mögliche gleichzeitige Betrachtung und Optimierung über alle vorliegenden Daten kann theoretisch eine optimale Datennutzung erreicht werden.

¹Eigentlich natürlich wieder der Raum $\mathbb{R}^{m \cdot 3n}$

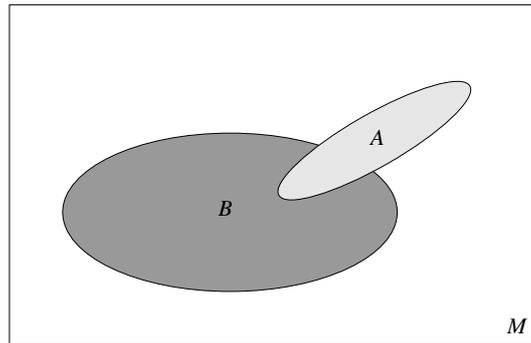


Abbildung 8.1: Beispiel für zwei sich überlappende Mengen A und B in einem Aufnahmebereich M . $g(A, B, (x_1, x_2)) = x_1 \mathbb{1}_A + x_2 \mathbb{1}_{B \setminus A}$ beschreibt diese Situation.

8.2 Erweiterte Mengensysteme und überlappende Mengen

Ein weiterer Punkt betrifft das Mengenmodell. In dieser Arbeit wurde — vor allem bedingt durch die Anwendung in Kapitel 6 — als Mengenmodell \mathcal{A} eine LH-abgeschlossene Teilmenge von $\mathfrak{A}(\mathbb{R}^d)$ gewählt. Das ist natürlich bei weitem nicht die einzige mögliche — und sinnvolle — Wahl eines Mengenmodells. Denkbar ist zum Beispiel auch, \mathcal{A} als Segmentierung im Sinne von [Han01] zu wählen, was vor allem in einer pixelbasierten Segmentierung von Vorteil sein kann. Eine weitere Möglichkeit könnte die gleichzeitige Betrachtung eines dreidimensionalen Objektes von mehreren verschiedenen Seiten sein. Das Mengenmodell \mathcal{A} enthielte in diesem Fall die möglichen dreidimensionalen Objekte. Auf der Basis dieses Modells wird dann für festgelegte Richtungen die jeweilige zweidimensionale Ansicht ermittelt wie sie schließlich von den Sensoren gesehen wird. Die so gewonnenen Erzeugendenfunktionen beinhalten dann die Berechnung der jeweiligen zweidimensionalen Ansicht aus einem 3D-Objekt $A \in \mathcal{A}$, zum Beispiel über eine Projektion in eine vorgegebene Ebene.

Ausgehend von LH-abgeschlossenen Teilmengen von $\mathfrak{A}(\mathbb{R}^d)$ existiert aber eine durchaus ins Auge springende Verallgemeinerung die hier kurz skizziert werden soll: Seien dazu \mathcal{A}_1 und \mathcal{A}_2 zwei LH-abgeschlossene Teilmengen von $\mathfrak{A}(\mathbb{R}^d)$. Nun wählt man Erzeugendenfunktionen mit dem Kreuzprodukt $\mathcal{A}_1 \times \mathcal{A}_2$ als Definitionsbereich, also beispielsweise als $e_1, e_2 : \mathcal{A}_1 \times \mathcal{A}_2 \rightarrow E$,

$$\begin{aligned} e_1(A, B) &:= \mathbb{1}_A, \\ e_2(A, B) &:= \mathbb{1}_{B \setminus A}. \end{aligned}$$

Mit der daraus resultierenden semilinearen Funktion

$$g : \mathcal{A}_1 \times \mathcal{A}_2 \times \mathbb{R}^2 \rightarrow E, \quad g(A, B, x) = x_1 e_1(A, B) + x_2 e_2(A, B) = x_1 \mathbb{1}_A + x_2 \mathbb{1}_{B \setminus A}$$

hat man auf einfache Weise eine Überlappungssituation modelliert. Abbildung 8.1 zeigt eine derartige Situation.

A Algebraische Grundlagen

Definition A.0.1

Ein \mathbb{R} -Vektorraum V zusammen mit einem Skalarprodukt $\langle \cdot | \cdot \rangle$ heißt ein euklidischer Vektorraum, wenn die Norm in V durch das Skalarprodukt induziert ist $\| \cdot \| := \sqrt{\langle \cdot | \cdot \rangle}$. Ist V vollständig bezüglich $\| \cdot \|$, heißt V Hilbertraum.

Für den folgenden Satz wird die *Distanzfunktion* benötigt, die im allgemeinen über einem metrischen Raum (X, δ) durch $\text{dist} : \mathfrak{P}(X) \times \mathfrak{P}(X) \rightarrow \mathbb{R} \cup \{\infty, -\infty\}$ mit

$$\text{dist}(A, B) := \begin{cases} \inf_{a \in A} \inf_{b \in B} d(a, b), & \text{falls } A \neq \emptyset \text{ und } B \neq \emptyset \\ \infty, & \text{sonst} \end{cases}$$

definiert ist. Für einen euklidischen Vektorraum V ist jedoch die Abbildung $d(x, y) := \|x - y\|$ eine Metrik und somit ist (V, d) auch ein metrischer Raum.

Man schreibt $\text{dist}(x, A)$ bzw. $\text{dist}(A, x)$ statt $\text{dist}(\{x\}, A)$ bzw. $\text{dist}(A, \{x\})$.

Satz A.0.2

Sei V ein euklidischer Vektorraum, W_n ein n -dimensionaler Unterraum von V sowie $b \in V$. Dann existiert ein eindeutig bestimmter Punkt $x_0 \in W_n$ kürzesten Abstandes von b gegen W_n , d.h.

$$\|b - x_0\| = \text{dist}(b, W_n).$$

Ist $\{w_1, \dots, w_n\}$ eine Basis von W_n , so existieren $c_1, \dots, c_n \in \mathbb{R}$ mit

$$x_0 = \sum_{i=1}^n c_i w_i, \quad \sum_{i=1}^n c_i \langle w_i | w_j \rangle = \langle b | w_j \rangle \quad \text{für alle } j = 1, \dots, n$$

und

$$\|b - x_0\|^2 = \langle b - x_0 | b - x_0 \rangle = \|b\|^2 - \sum_{i=1}^n c_i \langle b | w_i \rangle.$$

Ist $\{w_1, \dots, w_n\}$ sogar eine Orthonormalbasis von W_n , so gilt $c_i = \langle b | w_i \rangle$ für alle $i = 1, \dots, n$. x_0 wird auch Lotfußpunkt von b in W_n genannt.

Satz A.0.3 (Schmidtsches Orthonormalisierungsverfahren)

Ist V ein euklidischer Vektorraum, W_n ein n -dimensionaler Unterraum von V und $\{w_1, \dots, w_n\}$ eine Basis von W_n , so ist $\{w'_1, \dots, w'_n\}$ mit

$$w'_1 := \frac{w_1}{\|w_1\|}$$

und

$$\widetilde{w}_k' := w_k - \sum_{i=1}^{k-1} \langle w_k | w_i' \rangle w_i', \quad w_k' := \frac{\widetilde{w}_k'}{\|\widetilde{w}_k'\|} \quad \text{für } k = 2, \dots, n$$

eine Orthonormalbasis von W_n .

B Ergebnisse der Paßverfahren aus Kapitel 6

Dieser Anhang beinhaltet die Ergebnisse der Testbildsegmentierungen mit den beiden in Kapitel 6 vorgestellten Verfahren. Als Ausgangsdaten wurden 307 redigitalisierte Querschnittsbilder aus einem MR-Testdatenwürfel verwendet (eine genauere Beschreibung der Daten findet sich in Kapitel 6.3). Die folgenden Abbildungen zeigen die Ergebnisse aller Testbilder nach Optimierung der jeweiligen Zielfunktion über die *downhill simplex method* ([NM65]). Aufgrund der Abbildungsgeometrie beträgt der Querschnitt der redigitalisierten Schläuche 2 Pixel.

Als Startwert für den Kreismittelpunkt wurde der jeweilige Grauwertschwerpunkt des Ausgangsbildes gewählt. Da die Lage des Kreiscentrums im Optimum nicht sonderlich interessant ist, werden im folgenden nur die restlichen Parameter dargestellt.

B.1 Modifizierte Grauwertmodelladaptation

Bei der modifizierten Grauwertmodelladaptation muß neben den Kreisparametern Radius R und Mittelpunkt auch der Parameter σ der Gaußfunktion mitoptimiert werden. Abbildung B.1 zeigt die so bestimmten optimalen Radien aller 307 Testbilder bei Startwerten von $R = 1$ und $\sigma = 0.7$. Abbildung B.2 enthält die zugehörigen optimalen Werte von σ . Die diesen Werten entsprechenden Quadratmittelfehler sind in Abbildung B.3 dargestellt.

Die Durchschnittswerte aller Testbilder im Optimum sind:

	Durchschnitt	Standardabweichung
Radius	1.813360	0.024359
σ	0.577689	0.058879
Fehler	46.115735	14.193110

Das Verfahren ist relativ stabil gegenüber Änderungen des Startradius r . Berechnungen mit $R = 0,5$; $R = 2$ und $R = 3$ bei $\sigma = 0,7$ als Startwert führten zu denselben Ergebnissen. Erst bei einem Startwert von $R = 0.1$ kam es vereinzelt zu Ausreißern im Optimum. Ähnliche Ergebnisse erhält man auch bei Änderungen der Startwerte von σ

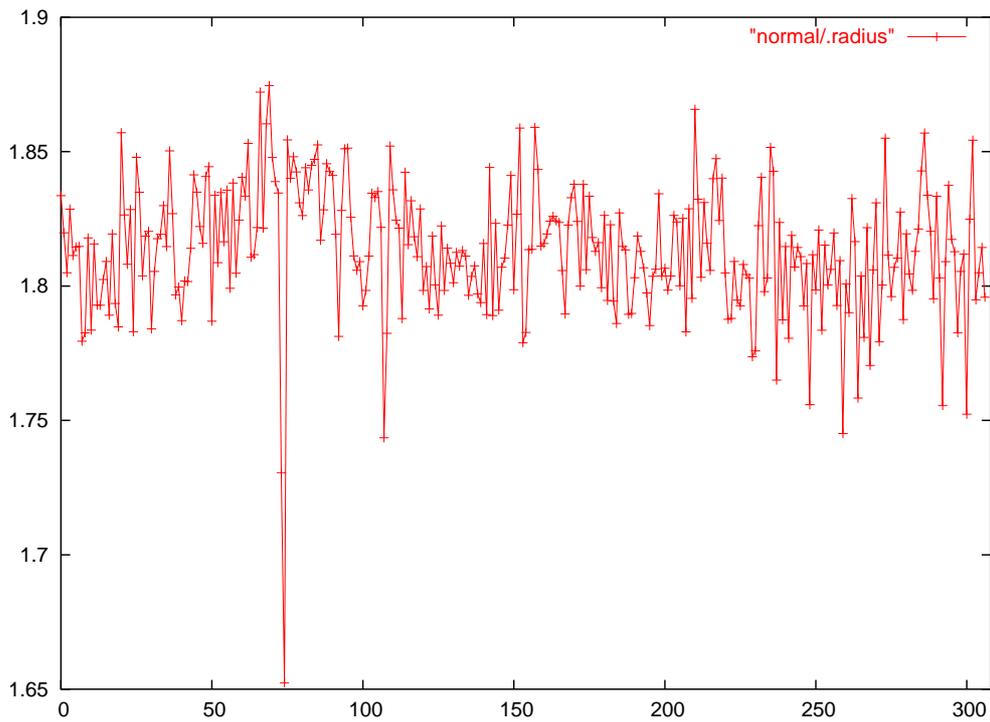


Abbildung B.1: Die berechneten Radien der 307 Querschnittsbilder bei der modifizierten Grauwertmodelladaptation. Startwerte waren Radius $R = 1$ und $\sigma = 0,7$.

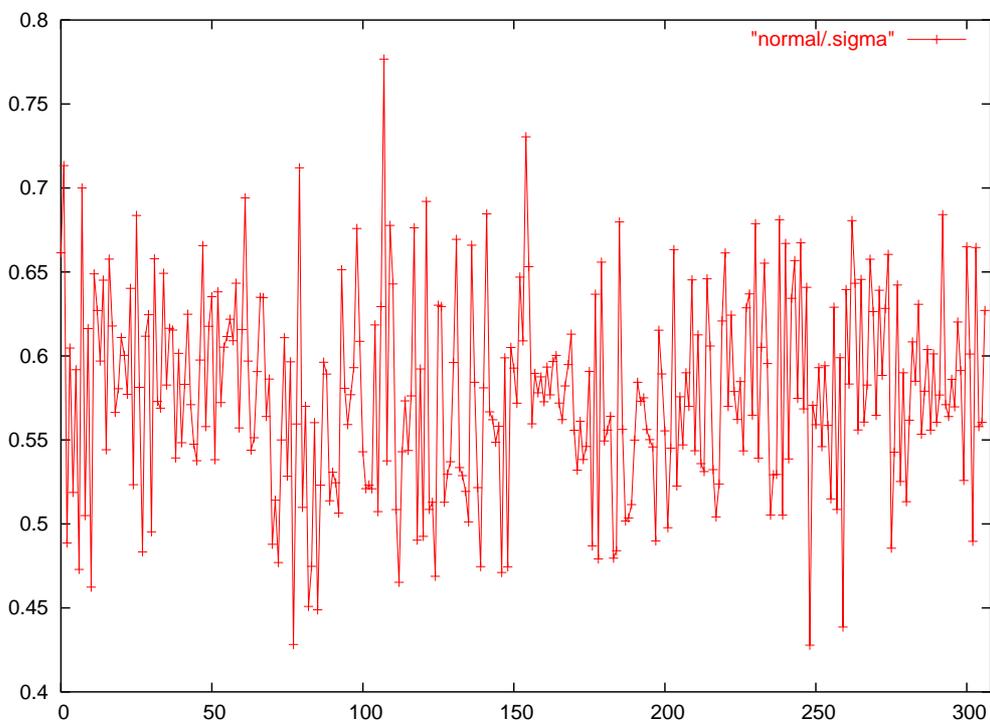


Abbildung B.2: Die mit den Radien mitoptimierten σ -Werte zu Abbildung B.1.

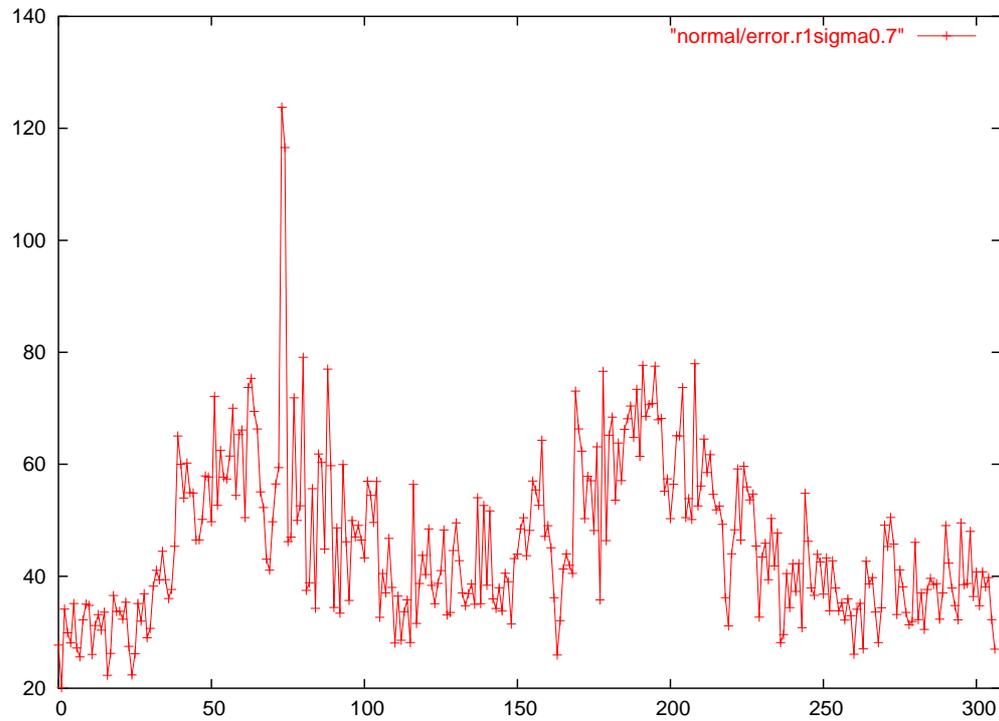


Abbildung B.3: Die zu den Abbildungen B.1 und B.2 zugehörigen Fehler im Optimum.

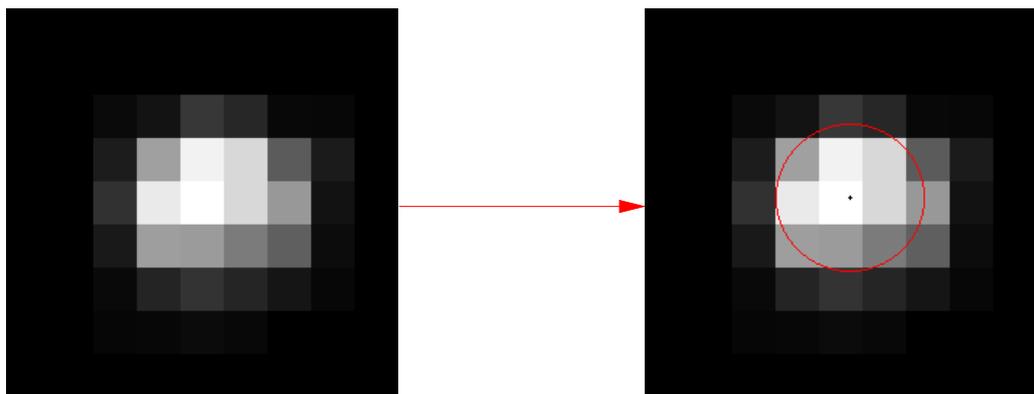


Abbildung B.4: Testbild Nr. 74. Die Werte im Optimum sind Radius $R = 1.73$, $\sigma = 0.55$ und Fehler = 123.77.

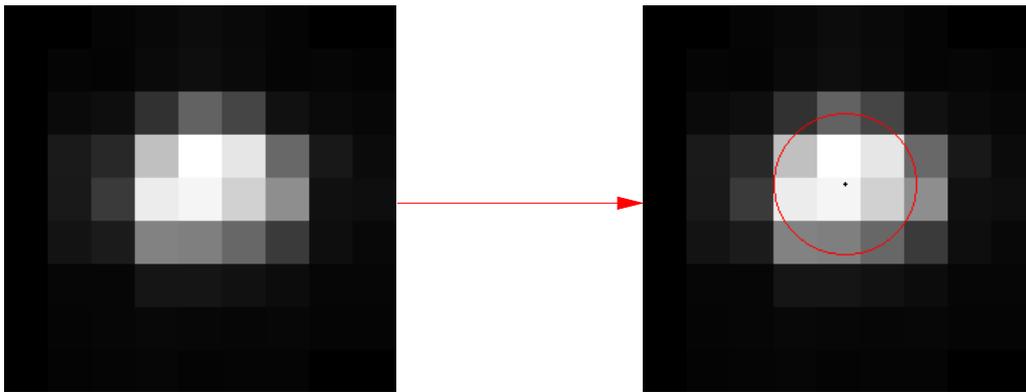


Abbildung B.5: Testbild Nr. 75 mit den folgenden Werten im Optimum: $R = 1.65$, $\sigma = 0.61$ und Fehler = 116.55.

(Ausreißer erst bei einer Startwertänderung von σ um mehr als 0.5) bzw. des Kreisentrums (Ausreißer erst bei einer Verschiebung des Startwertes um mehr als ein Pixel vom Grauwertschwerpunkt weg).

Die Testbilder Nr. 74 und Nr. 75 deuten zunächst auf eine Fehlpassung hin, denn sowohl Radius als auch Fehler im Optimum weichen bei diesen beiden Bildern deutlich vom Rest ab. Betrachtet man allerdings die Ergebnispassungen (Abbildungen B.4 und B.5), ist die Sachlage nicht mehr so eindeutig, denn die eingepaßten Kreisscheiben sind durchaus realistisch. Der relativ kleine Kreisradius wurde wahrscheinlich durch den stärkeren Abfall der Grauwerte in y -Richtung im Gegensatz zu anderen Bildern dieser Testreihe verursacht (vgl. dazu auch Abbildung 6.6 von Seite 90 und Abbildung 6.7 von Seite 91). Allerdings ist der Fehler im Optimum mit 116.55 bei diesen beiden Bildern doch relativ groß.

Die Abbildungen B.6 und B.7 zeigen die Ergebnisse, wenn σ auf den obigen Durchschnittswert von 0.577 fixiert ist. Da die Standardabweichung von σ sowieso recht gering war, weichen diese Ergebnisse nur wenig von den vorigen Ergebnissen aus den Abbildungen B.1 und B.3 ab.

Für die Durchschnittswerte aus den Abbildungen B.6 und B.7 gilt:

	Durchschnitt	Standardabweichung
Radius	1.815680	0.028453
σ	0.577 fest	0.0
Fehler	47.966906	14.145489

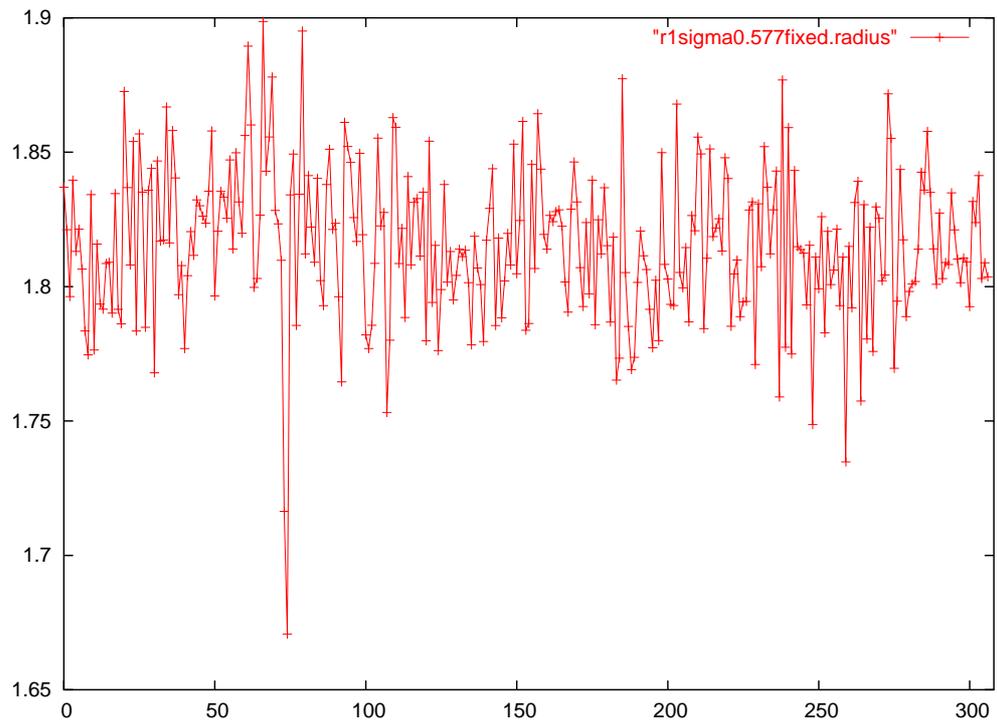


Abbildung B.6: Berechnete Radien im Optimum für $\sigma = 0.577$.

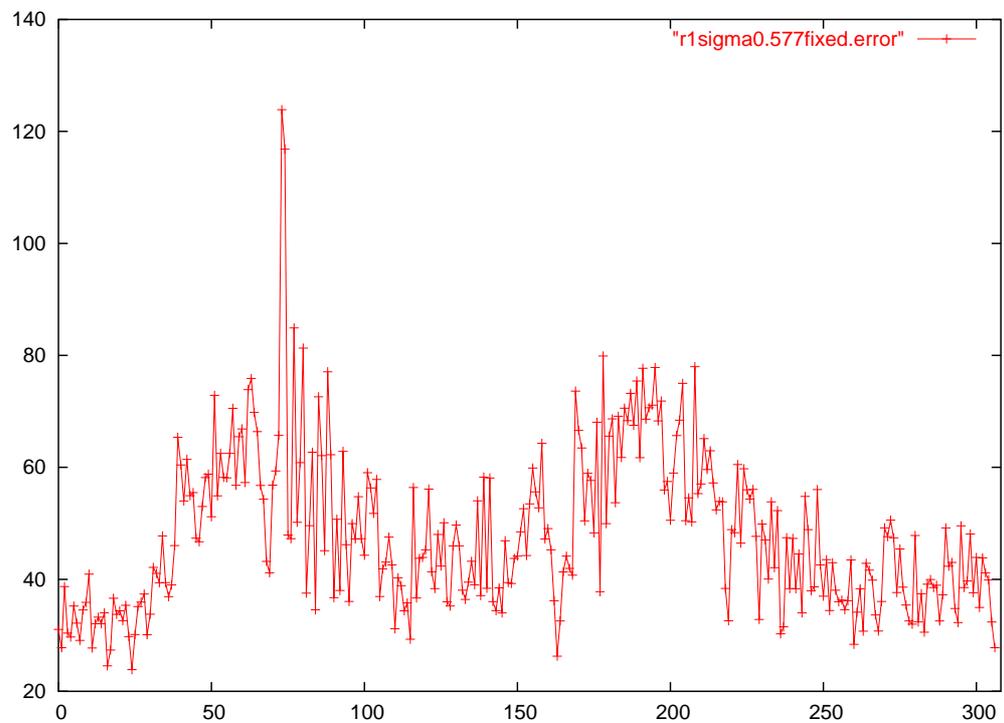


Abbildung B.7: Entsprechender Fehler im Optimum zu Abbildung B.6.

B.2 Kreisscheibenpassung und Vergleich beider Verfahren

Die Abbildungen B.8 und B.9 zeigen die Ergebnisse der Testbildsegmentierungen für die Kreisscheibenpassung. Bei Start mit dem Grauwertschwerpunkt als Kreiszentrum und einem Radius von 1 ergeben sich die folgenden Durchschnittswerte im Optimum:

	Durchschnitt	Standardabweichung
Radius	1.864846	0.065176
Fehler	90.115384	18.212526

Im Vergleich zur modifizierten Grauwertmodelladaption ist hier der durchschnittliche Fehler im Optimum höher (Abbildung B.10). Zudem schwanken die Resultate von Testbild zu Testbild bei der Kreisscheibenpassung deutlich stärker als bei der modifizierten Grauwertmodelladaption (Abbildung B.11), was sich auch in der höheren Standardabweichung zeigt. Der durchschnittlich berechnete Radius liegt dagegen bei beiden Verfahren um die 1.8 Pixel. Das entspricht den Erwartungen, daß der berechnete Radius der Testbilder bei etwa 2 Pixeln liegen sollte. Dieses Verfahren ist ähnlich unempfindlich gegenüber Startwertverschiebungen wie die modifizierte Grauwertmodelladaption.

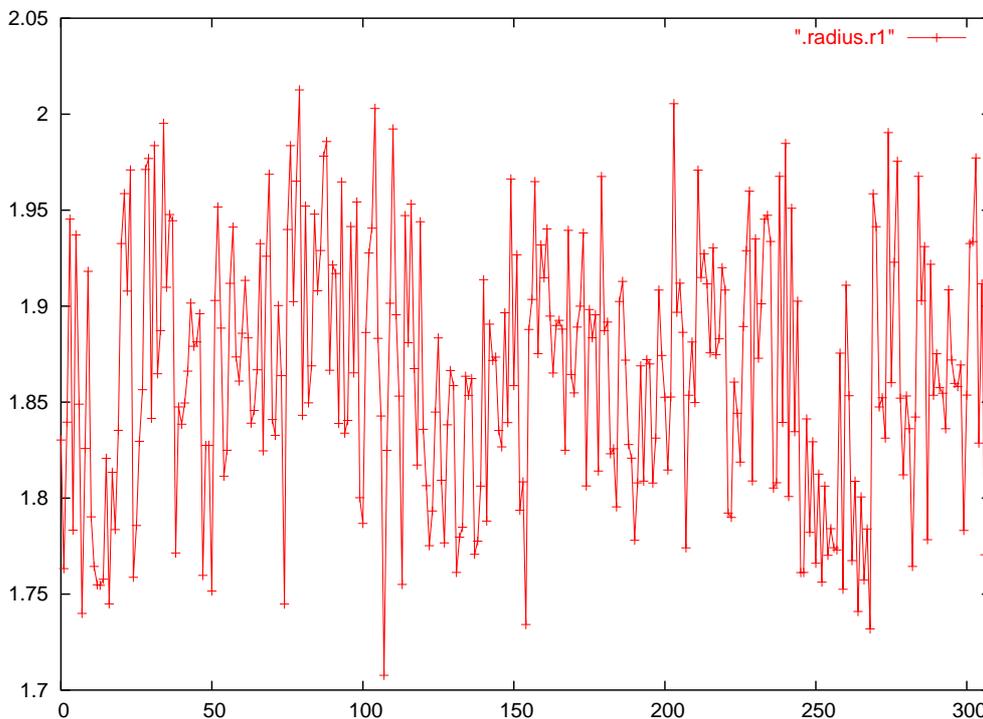


Abbildung B.8: Ergebnisradien mit Grauwertschwerpunkt als Startzentrum und einem Startradius von $R = 1$ Pixel.

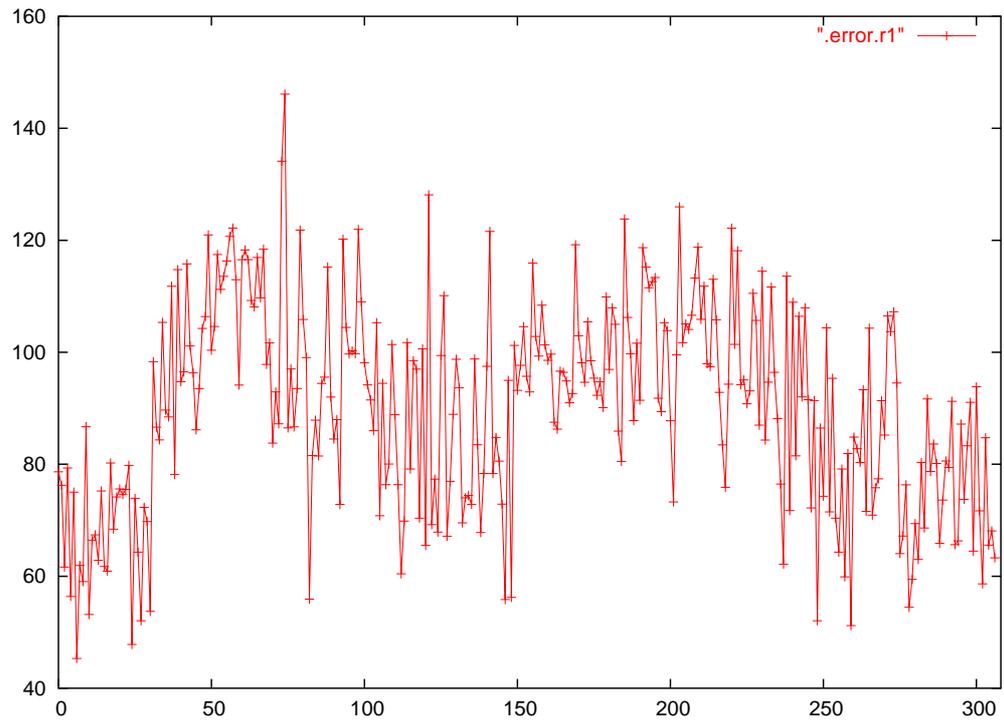


Abbildung B.9: Entsprechende Quadratmittelfehler zu Abbildung B.8.

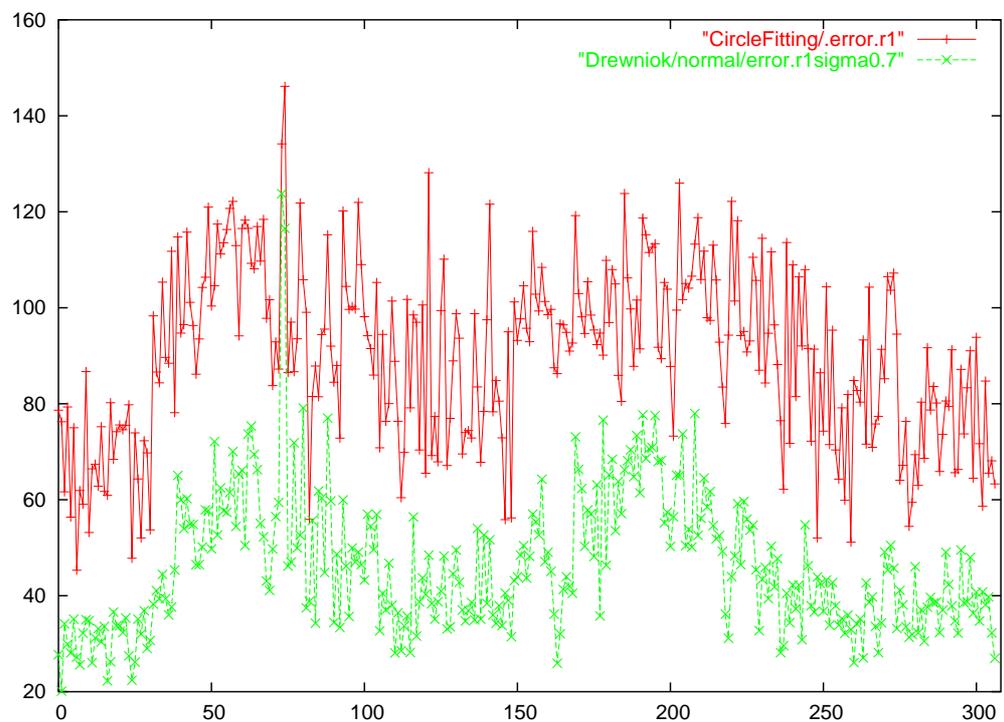


Abbildung B.10: Vergleich der Quadratmittelfehler beider Verfahren (Kreisscheibenpassung rot, modifizierte Grauwertmodelladaption grün).

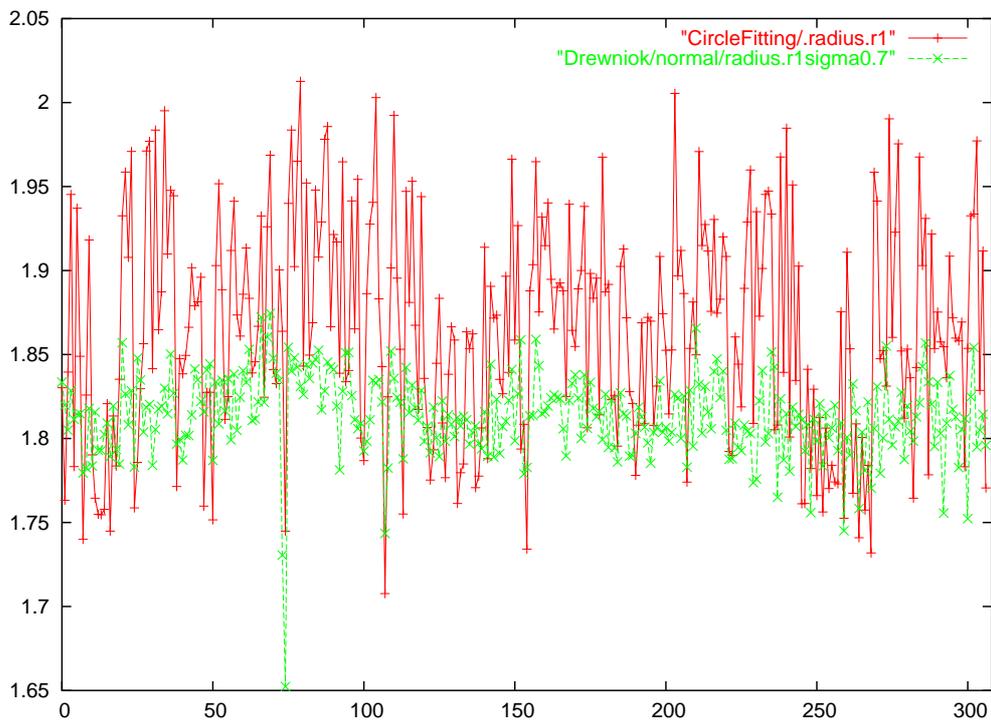


Abbildung B.11: Vergleich der optimalen Radien beider Verfahren (Kreisscheibenpassung rot, modifizierte Grauwertmodelladaption grün).

Literaturverzeichnis

- [AR99] AHN, S.; RAUH, W.: *Geometric Least Squares Fitting of Circle and Ellipse*; Int Journal of Pattern Recognition and Artivicial Intelligence; Bd. 13 (7), S. 987 – 996, November 1999
- [ARR99] AHN, S.; RAUH, W.; RECKNAGEL, M.: *Geometric Fitting of Line, Plane, Circle, Sphere, and Ellipse*; in: *Proc. 6. ABW Workshop Optische 3D-Formerfassung*; Technische Akademie Esslingen, 25.–26. Januar 1999
- [BS91] BRONSTEIN, I.; SEMENDJAJEW, K.: *Taschenbuch der Mathematik*; 25. Aufl.; B.G.Teubner Verlagsgesellschaft, Stuttgart, Leipzig, 1991
- [BS96] BRAY, D. E.; STANLEY, R. K.: *Nondestructive Evaluation*; CRC Press, 1996
- [Bur96] BURKE, M. W.: *Image Acquisition – Handbook of Machine Vision Engineering, Volume 1*; Chapman & Hall, London, 1996
- [BZ87] BLAKE, A.; ZISSERMAN, A.: *Visual Reconstruction*; MIT Press, Cambridge MA, 1987
- [Die85] DIEUDONNÉ, J.: *Grundzüge der modernen Analysis, Band 1*; 3. Aufl.; Friedr. Vieweg & Sohn, Braunschweig, 1985
- [Don91] DONNER, K.: *Hierarchical Approximation in Pattern Recognition*; in: *Proceedings Applications of Artificial Intelligence*; S. 213–226; Prague, 1991
- [Don97] DONNER, K.: *Image Interpretation Based on Local Transform Characterization*; Pattern Recognition and Image Analysis; Bd. 7 (4), S. 431 – 447, 1997
- [Don98] DONNER, K.: *Persönliche Notizen*, 1998
- [Dre98] DREWNIOK, C.: *Objektlokalisierung durch Adaption parametrischer Grauwertmodelle und ihre Anwendung in der Luftbildauswertung*; Dissertation; Universität Hamburg, Fachbereich Informatik, 1998

- [FDF⁺94] FOLEY, J. D.; VAN DAM, A.; FEINER, S. K.; HUGHES, J. F.; PHILLIPS, R. L.: *Introduction to Computer Graphics*; Addison-Wesley Publishing Company, 1994
- [FPF99] FITZGIBBON, A.; PLIU, M.; FISHER, R.: *Direct Least Square Fitting of Ellipses*; IEEE Transactions on Pattern Analysis and Machine Intelligence; Bd. 21 (5), S. 476 ff, 1999
- [Haa99] HAAS, J.: *Echtzeit-Korrespondenzprobleme in Bildsequenzen und Subpixelgenauigkeit*; Dissertation; Universität Passau, 1999
- [Hab91] HABERÄCKER, P.: *Digitale Bildverarbeitung, Grundlagen und Anwendungen*; Carl Hanser Verlag, München, 1991
- [Han00] HANNING, T.: *Nonparametric Segmentation of Grayvalue Images by n-level-set-fitting*; in: *World Multiconference on Systemics, Cybernetics and Informatics 2000*; Bd. V; S. 90–95; Orlando, July 2000
- [Han01] HANNING, T.: *Vektorielle Mehrniveaupassung — Anwendungen in der Bildsegmentierung*; Dissertation; Universität Passau, 2001
- [Her80] HERMAN, G. T.: *Image Reconstruction from Projections*; Academic Press New York, 1980
- [Hue71] HUECKEL, M. H.: *An Operator which Locates Edges in Digitized Pictures*; Journal of the Assoc for Computing Machinery; Bd. 18 (1), S. 113 – 125, Januar 1971
- [Käm00] KÄMPKE, T.: *Optimal and Near Optimal Quantization of Integrable Functions*; Computers and Mathematics with Applications; Bd. 40, 2000
- [KK98] KÄMPKE, T.; KOBER, R.: *Nonparametric Optimal Binarization*; in: *Proceedings of the 14th International Conference on Pattern Recognition ICPR*; Brisbane, 1998
- [Kol95] KOLLER, W.: *Segmentierung und Diagnose von Gefäßen mit Atlasunterstützung*; Dissertation; Universität Passau, 1995
- [Kus92] KUSTERER, S.: *Querschnittspassung von Blutgefäßen in MR-Bilddatensätzen*; Diplomarbeit; Universität Passau, 1992
- [Kus98] KUSTERER, S.: *Prototypische Passung endlicher Vereinigungen von Hyperebenen*; Dissertation; Universität Passau, 1998
- [LH79] LAWSON, C. L.; HANSON, R. J.: *Solving Least Squares Problems*; Prentice Hall, New Jersey, 1979

-
- [LP04] LAUREN, V.; PISINGER, G.: *Automated Analysis of Vessel Diameters in MR-Images*; in: *Proceedings of the Fourth IASTED International Conference on Visualization, Imaging, and Image Processing*; Marbella, Spain, September 2004
- [LR84] LINDGREN, A.; RATTEY, P.: *Tomographic Image Reconstruction from a Limited Number of Projections*; *Signal Processing*; Bd. 7, S. 371 – 387, 1984
- [MS85] MUMFORD, D.; SHAH, J.: *Boundary Detection by Minimizing Functionals*; in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*; S. 22 – 26; San Francisco, 1985
- [Nat86] NATTERER, F.: *The Mathematics of Computerized Tomography*; Teubner, Stuttgart, 1986
- [NB86] NALWA, V. S.; BINFORD, T. O.: *On Detecting Edges*; *IEEE Trans on Pattern Analysis and Machine Intelligence*; Bd. 8 (6), S. 699 – 714, November 1986
- [NM65] NELDER, J.; MEAD, R.: *A Simplex Method for Function Minimization*; *Computer Journal*; Bd. 7, S. 308 – 313, 1965
- [PTVF96] PRESS, W.; TEUKOLSKY, S.; VETTERLING, W.; FLANNERY, B.: *Numerical Recipes in C*; 2. Aufl.; Cambridge University Press, Cambridge, 1996
- [Qui93] QUINTO, E. T.: *Singularities of the X-ray Transform and Limited Data Tomography in \mathbb{R}^2 and \mathbb{R}^3* ; *Siam J Math Anal*; Bd. 24 (5), S. 1215 – 1225, September 1993
- [Roh92] ROHR, K.: *Recognizing Corners by Fitting Parametric Models*; *International Journal of Computer Vision*; Bd. 9 (3), S. 213 – 230, 1992
- [SA96] SHAH, S.; AGGARWAL, J. K.: *Intrinsic Parameter Calibration Procedure for a (High-Distortion) Fish-eye Lens Camera with Distortion Model and Accuracy Estimation*; *Pattern Recognition*; Bd. 29 (11), S. 1775–1788, 1996
- [Sch86] SCHREIBER, W. F.: *Fundamentals of Electronic Imaging Systems*; Bd. 15 von *Springer Series in Information Sciences*; Springer Verlag, Berlin, 1986
- [Spe93] SPELLUCCI, P.: *Numerische Verfahren der nichtlinearen Optimierung*; Birkhäuser Verlag, Berlin, 1993
- [TM82] THURGOOD, J. D.; MIKHAIL, E. M.: *Photogrammetric Analysis of Digital Images*; in: *Proc. of ISPRS Commission III, Helsinki*; S. 576 – 590, 1982
- [Tsa87] TSAI, R.: *A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses*; *IEEE Journal of Robotics and Automation*; Bd. 3 (4), S. 323–344, 1987
-